

République algérienne démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Kasdi Merbeh Ouargla



Faculté des Nouvelles Technologies de L'informatique
Et de la communication
Département d'informatique et Technologies de l'information

Mémoire présenté par :

Gouarah Bouthaina

Gouarah Abir

Pour l'obtention du diplôme de Master

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Industrielle

Sujet :

Identification de la misogynie sur les réseaux sociaux

Superviseure : Dr. Chahrazad Toumi

Président : Dr. Oussama Zerdoumi

Examineur : Dr. Fouad Bekkari

Année universitaire : 2023 /2024

Remerciements

Nous remercions Dieu tout-puissant qui nous a donné le don de la raison pour nous guider sur le chemin et nous a permis d'achever ce travail, et nous remercions également nos parents qui nous ont encouragés et motivés pour atteindre ce niveau d'étude. Nous tenons également à remercier sincèrement tous ceux qui nous ont aidés et ont contribué à notre réussite, que ce soit par un mot, un geste ou une opinion, en particulier le professeur **Chahrazad Toumi**, qui a supervisé cette recherche et nous a guidés en permanence pour résoudre les problèmes qu'elle soulevait. De même, nous adressons nos sincères remerciements aux membres du jury. Le président, le **Dr. Oussama Zerdoumi**, et l'examineur, le **Dr. Fouad Bekkari**, pour l'honneur qu'ils nous ont fait en participant au jury et en examinant ce travail. Nous nous en voudrions de ne pas mentionner tous nos professeurs pour leurs conseils au cours des cinq dernières années au département des technologies de l'information de l'université Kasdi Merbah de Ouargla. Nous remercions tout particulièrement les membres de notre famille et nos amis qui ont toujours été là pour nous. Enfin, nous tenons à remercier toutes les personnes qui nous ont aidés de près ou de loin durant nos études à toutes les étapes de nos études et dans la préparation de cette thèse avec notre profonde gratitude et notre respect.

Résumé

Les réseaux sociaux font partie intégrante de notre vie quotidienne, car ils ont contribué à faciliter la communication entre les personnes à l'échelle mondiale et sont devenus une plateforme de communication et de partage d'informations, d'idées, d'opinions, ...ect. Malgré que ces sites présentent des avantages et des inconvénients qui affectent grandement la société. La misogynie qui est **un sentiment de mépris ou d'hostilité à l'égard des femmes, motivé par leur sexe**. La misogynie est un phénomène social très répandu ces dernières années, qui est devenu une source d'inquiétude croissante pour les femmes sur l'internet et qui a acquis de nouvelles. L'objectif de notre travail est de créer un modèle qui peut détecter efficacement la misogynie dans les sites de réseaux sociaux. Nous allons présentons deux modèles d'apprentissage d'ensemble qui utilise CNN, LSTM, pour l'identification de la misogynie. Pour ce faire, l'ensemble de données Let-Mi collectées par les chercheurs à partir des commentaires Twitter est utilisé, et un examen des réseaux sociaux et du phénomène de la misogynie dans les réseaux sociaux est présenté. Des algorithmes d'apprentissage automatique sont utilisés pour catégoriser les textes et identifier les commentaires misogynes et non misogynes.

Mots clés : Réseaux sociaux, Misogynie, CNN, LSTM, Let-Mi, Twitter, Apprentissage automatique, Deep Learning.

Abstract

Social networks are an integral part of our daily lives, as they have helped to facilitate communication between people on a global scale and have become a platform for communicating and sharing information, ideas, opinions, ...ect. However, these sites have both advantages and disadvantages that greatly affect society. Misogyny is a feeling of contempt or hostility towards women, motivated by their gender. Misogyny is a widespread social phenomenon in recent years, which has become a growing source of concern for women on the internet and has acquired new ones. The aim of our work is to create a model that can effectively detect misogyny in social networking sites. We will present two ensemble learning models that use CNN, LSTM, for misogyny identification. To this end, the Let-Mi dataset collected by researchers from Twitter comments is used, and an examination of social networks and the phenomenon of misogyny in social networks is presented. Machine learning algorithms are used to categorize texts and identify misogynistic and non-misogynistic comments.

Keywords: Social networks, Misogyny, CNN, LSTM, Let-Mi, Twitter, Machine learning, Deep Learning.

ملخص

تعد شبكات التواصل الاجتماعي جزءاً لا يتجزأ من حياتنا اليومية، حيث ساعدت على تسهيل التواصل بين الناس على نطاق عالمي وأصبحت منصة للتواصل وتبادل المعلومات والأفكار والآراء ... إلخ. على الرغم من أن هذه المواقع لها مزايا وعيوب تؤثر بشكل كبير على المجتمع. كره النساء، وهو شعور بالازدراء أو العداء تجاه المرأة بدافع جنسها. وتعد كراهية النساء ظاهرة اجتماعية منتشرة على نطاق واسع في السنوات الأخيرة، والتي أصبحت مصدر قلق متزايد للنساء على الإنترنت. الهدف من عملنا هو إنشاء نموذج يمكنه الكشف بفعالية عن كراهية النساء في مواقع التواصل الاجتماعي. سنقدم نموذجين للتعلم التجميعي يستخدمان شبكة CNN، LSTM، لتحديد كراهية النساء. وللقيام بذلك، يتم استخدام مجموعة بيانات Let-Mi التي جمعها الباحثون من تعليقات تويتر، ويتم تقديم فحص للشبكات الاجتماعية وظاهرة كراهية النساء في الشبكات الاجتماعية. تُستخدم خوارزميات التعلم الآلي لتصنيف النصوص وتحديد التعليقات الكارهة للنساء وغير الكارهة للنساء.

كلمات مفتاحية: مواقع التواصل الاجتماعي، كراهية النساء، CNN، LSTM، Let-Mi، تويتر، التعلم الآلي، التعلم العميق.

Table des matières

Introduction générale	1
Chapitre I : Généralité sur les réseaux sociaux et la misogynie	
I.1 Introduction	3
I.2 Définition des réseaux sociaux.....	3
I.3 Comment et quand sont nés les réseaux sociaux ?.....	4
I.4 L'impact positifs et négatifs des réseaux sociaux sur la société.....	5
I.4.1 L'impact positifs	5
I.4.2 L'impact négatifs.....	6
I.5 Qu'est-ce qu'un comportement antisocial ?.....	8
I.6 Types de comportements antisociaux	9
I.6.1 La cyberintimidation.....	9
I.6.2 Le cyberharcèlement.....	9
I.6.3 La misogynie	9
I.7 Définition de la misogynie	9
I.8 Étymologie et utilisation historique	10
I.9 La misogynie dans la société.....	11
I.10 La misogynie dans les réseaux sociaux.....	12
I.11 Causes de la misogynie	13
I.12 Les méfaits de la misogynie en réseaux sociaux.....	13
I.13 Moyens populaires de lutter contre la misogynie sur les réseaux sociaux (Twitter)	14
I.14 Conclusion.....	14
Chapitre II : Machine Learning et détection misogynie	
II.1 Introduction.....	15
II.2 Qu'est-ce que l'intelligence artificielle ?	15

II.3 Exemples d'intelligence artificielle :.....	16
II.4 L'intelligence artificielle dans le monde du travail	16
II.5 Qu'est-ce que l'intelligence générale artificielle (AGI) ?.....	17
II.6 Types d'IA	17
II.6.1 Machines interactives	18
II.6.2 Les machines à mémoire limitée	18
II.6.3 Machines à théorie de l'esprit	18
II.6.4 Machines conscientes d'elles-mêmes	18
II.7 Avantages et risques de l'intelligence artificielle.....	19
II.8 Qu'est-ce que l'apprentissage automatique (Machine Learning) ?.....	20
II.9 Types d'apprentissage automatique (Machine Learning)	20
II.9.1 Apprentissage automatique supervisé	21
II.9.2 Apprentissage automatique non supervisé	21
II.9.3 Apprentissage semi-supervisé	21
II.9.4 Apprentissage par renforcement.....	22
II.10 Les 3 principales applications de l'apprentissage automatique (Machine Learning)...	22
II.10.1 Secteur du commerce de détail.....	22
II.10.2 L'industrie du voyage	22
II.10.3 Réseaux sociaux	23
II.11 Les 10 Principes tendances de l'apprentissage automatique en 2022	23
II.12 Techniques d'apprentissage automatique	24
II.12.1 Natural language processing (NLP)	24
II.12.2 L'apprentissage e profond (Deep Learning)	26
II.13 Détection de la misogynie en ligne	31
II.13.1 Méthodes et outils de détection automatique du langage misogyne	31
II.13.2 Identification automatique de la misogynie	34
II.14 Conclusion	40

Chapitre III : Conception, Implémentation et résultats

III.1 Introduction	48
III.2 Motivation	48
III.3 Dataset	48
III.4 Les outils et bibliothèques utilisé	49
III.5 Configuration expérimentale	53
III.6 Implémentation	53
III.6.1 Prétraitement	53
III.6.2 Intégration des données.....	54
III.6.3 Diviser les données	55
III.7 Résultats expérimentaux et discussion	55
III.8 Résultats de l'expérimentation	56
III.9 Conclusion	Erreur ! Signet non défini.
Conclusion générale.....	60
Les références.....	61

Liste des figures

Figure 1: Exemples de réseaux sociaux.	4
Figure 2: Types d'apprentissage automatique.	21
Figure 3: Tendances de l'apprentissage automatique à attendre en 2022.....	24
Figure 4: Déroulement de réseau neuronal récurrent.	27
Figure 5: Convolutional Neural Network.	27
Figure 6: Long Short-Term Memory Network and Block.	28
Figure 7: Un neurone unique et son équivalent mathématique.....	29
Figure 8: Un réseau neuronal multicouche (3 couches).....	29
Figure 9: Nuage de mots du dataset après prétraitement.	54
Figure 10: les statistiques relatives à la division de l'ensemble de données.	56

Liste des tableaux

Tableau 1: Avantages et risques de l'intelligence artificielle.....	19
Tableau 2: Résumé des approches d'identification automatique de la misogynie.....	40
Tableau 3: Caractéristiques de l'ensemble de données utilisé.	48
Tableau 4: Tweetez des exemples d'étiquettes d'annotation.	49
Tableau 5: Mesures d'évaluation des performances sur l'ensemble de données Let-Mi.	57
Tableau 6: Résultats de la tâche d'identification de la misogynie.[15].....	58

Liste d'abréviation

PDG -Président Directeur Général

AI-Artificial Intelligence (Intelligence Artificielle)

NLP-Natural Language Processing (Traitement du Langage Naturel)

IAG-Intelligence Artificielle Générale

ChatGPT-Chat Generative Pre-trained Transformer

LLM- Large Language Model (Grand Modèle de Langage)

SDI - Système de Détection d'Intrusion

RNN - Recurrent Neural Network (Réseau de Neurones Récurrents)

AR - Augmented Reality (Réalité Augmentée)

GAN – Generative Adversarial Network (Réseau Adversarial Génératif)

CNN - Convolutional Neural Network (Réseau de Neurones Convolutifs)

ML - Machine Learning (Apprentissage Automatique)

DL- Deep Learning (Apprentissage Profond)

TALN -Traitement Automatique du Langage Naturel

NLU- Natural Language Understanding (Compréhension du Langage Naturel)

NLG - Natural Language Generation (Génération de Langage Naturel)

SVM - Support Vector Machine (Machine à Vecteurs de Support)

ACM - Association for Computing Machinery

DNN - Deep Neural Network (Réseau Neuronal Profond)

NN - Neural Network (Réseau Neuronal)

NLTK - Natural Language Toolkit

NUMPY - Numerical Python

CSV -Comma-Separated Values

Re- Regular Expressions

Let-Mi-Levantine Twitter dataset for Misogynistic

Pad-sequences -Padding sequences

Introduction générale

Aujourd'hui, les plateformes de réseaux sociaux sont devenues de puissants outils d'expression et de communication, offrant un moyen efficace de communiquer et d'interagir avec les autres. Cependant, malgré les nombreux avantages qu'elles offrent, ces plateformes sont également devenues un terrain propice à la propagation de certains phénomènes négatifs et positifs dans la société. Ces dernières années, plusieurs plateformes sont apparues à des fins de communication, de transfert d'informations et d'expression d'opinions, telles que Facebook et Twitter. La misogynie est l'un des phénomènes les plus marquants. Il s'agit d'un problème urgent qui touche de nombreuses femmes dans le monde, car les manifestations de haine et de préjugés à leur égard se manifestent de multiples façons, notamment par des commentaires insultants, des blagues blessantes, la diffusion de stéréotypes sexistes et la remise en question de leurs droits et de leurs valeurs.

Un misogyne sur les réseaux sociaux est une personne qui exprime et promeut la haine ou le mépris à l'égard des femmes par le biais de ces plateformes, et cela peut se manifester de diverses manières, notamment par des commentaires désobligeants, des tactiques haineuses, des blagues offensantes, la diffusion de stéréotypes de genre, voire la remise en question des droits et des valeurs des femmes.

Dans la littérature, il existe différentes façons de détecter la misogynie sur les plateformes sociales, comme la sensibilisation à l'importance des femmes et à leur rôle actif dans la société, la réalisation de l'égalité entre les hommes et les femmes, etc. Dans le cadre de notre projet, nous avons proposé de créer un programme de détection de la misogynie. Dans le cadre de notre projet, nous avons proposé de créer un programme de détection de la misogynie. Notre thèse est donc divisée en trois chapitres :

Dans le premier chapitre, nous présenterons les réseaux sociaux en les définissant, en mentionnant certains de leurs avantages et inconvénients et en donnant un aperçu de leurs aspects négatifs, y compris la misogynie dont nous allons parler.

Dans le deuxième chapitre, nous donnerons un aperçu de l'intelligence artificielle et de ses utilisations dans la vie quotidienne, en mentionnant l'intelligence artificielle générale, puis nous aborderons le concept d'apprentissage automatique et ses techniques, nous expliquerons le

concept d'apprentissage profond et de NLP et, à la fin du chapitre, nous parlerons des moyens de lutter contre la misogynie.

Dans le troisième chapitre, nous donnerons un aperçu des étapes de la mémoire et de la méthodologie nécessaire à la mise en œuvre du programme, nous expliquerons comment il a été mis en œuvre et nous discuterons des résultats obtenus.

Chapitre

I Généralité sur les réseaux sociaux et la misogynie

I.1 Introduction

De nos jours, l'internet est devenu une routine pour presque tout le monde et une source majeure d'information pour de nombreuses personnes, et ces dernières années, plusieurs plateformes sont apparues à des fins de communication, de transfert d'informations et d'expression d'opinions, organisées selon le même principe que les magazines, les publications, les articles, etc. dans plusieurs sections telles que : (nouvelles, santé, mode et sections de voyage...), et les utilisateurs de ces plateformes peuvent ne pas être scientifiquement compétents mais des personnes ordinaires qui possèdent des connaissances approfondies dans certains domaines. Les utilisateurs de ces plateformes peuvent ne pas être scientifiquement compétents, mais être des personnes ordinaires ayant des connaissances approfondies dans certains domaines, et les participants à la discussion peuvent se permettre d'utiliser un langage abusif et des remarques insultantes à l'égard d'autres personnes, ce qui est considéré comme un problème - langage abusif ou discours de haine sur les sites de réseaux sociaux - et l'un de ses types est le discours misogyne qui s'est considérablement répandu ces dernières années et est devenu une limitation de la liberté des femmes sur les sites de réseaux sociaux, et il est nécessaire d'aborder ce problème, d'autant plus qu'il est devenu une source de préoccupation et d'intérêt, et pour le résoudre nous étudierons ce problème dans ce chapitre.

I.2 Définition des réseaux sociaux

Les réseaux sociaux sont des sites internet (ou des applications mobile) qui permettent aux usagers et aux internautes de partager du contenu personnel, de créer une page et d'échanger des informations, des photos et des vidéos avec une communauté d'amis et de connaissances. Cette communauté est virtuelle, même si elle peut être rencontrée dans la réalité. Parmi les réseaux sociaux les plus connus, on peut citer Facebook, Twitter, Instagram, etc.[1]



Figure 1: Exemples de réseaux sociaux.

I.3 Comment et quand sont nés les réseaux sociaux ?

Mark Zuckerberg, **fondateur de Facebook** dont la fortune s'élève à 112 milliards est souvent désigné comme le père fondateur du concept de réseau social. Pourtant, ce n'est pas lui qui l'a inventé !

- 1997 : **lancement de Sixdegrees.com**, considéré comme LE 1er réseau social. Sur ce site, on peut créer une page profil, se constituer un réseau et envoyer des messages à ses contacts.
- 1999 : **naissance de MSN** (outil de discussion entre internautes rebaptisé Windows Live Messenger en 2005) et de Blogger (une plateforme de blog sur laquelle chacun peut poster ce qu'il veut : textes, articles, photos, vidéos...).
- 2000 : **Hot or not Hot** était un site où les internautes pouvaient noter les photos d'autres utilisateurs (il semblerait que ce site ait influencé les fondateurs de Facebook).
- 2002 : l'année où tout bascule... Les réseaux sociaux se multiplient comme des petits pains sur la toile. On peut citer notamment **Friendster** (qui n'a pas survécu à un afflux trop important d'utilisateurs) et **Myspace** (l'ancêtre de YouTube plébiscités par les ados qui y postent leurs créations artistiques).
- 2003 : Mark Zuckerberg crée **Facemash** qui deviendra ensuite **Facebook** en 2005 pendant que 2 autres sites de partage de photos Flickr et Photobucket sont lancés.

- 2005 : **YouTube** publie [sa première vidéo](#)... au Zoo !
- 2006 : **Twitter**, plateforme de microblogging sur laquelle on publie des micro-messages appelés « tweets » est lancé... La même année, bien que fondée en 2002, la plateforme de mise en relation entre professionnels LinkedIn connaît enfin le succès.
- 2010 : fondé par un américain et un brésilien, **Instagram**, réseau social dédié au partage de vidéos et de photos, est en ligne.
- 2011 : **Snapchat**, application créée par 2 jeunes étudiants de Stanford, apparaît dans le catalogue de l'Apple Store (puis en 2012 sur Android).
- 2017 : **TikTok** arrive sur les marchés situés en dehors de Chine. Cette app dédiée au partage de vidéo est déjà utilisée par les chinois depuis 2016 sous le nom de Douyin.[1]

I.4 L'impact positifs et négatifs des réseaux sociaux sur la société

La révolution des technologies de la communication est la principale raison de l'émergence de divers sites de réseaux sociaux, car les réseaux sociaux influencent grandement la culture, l'économie et la vision du monde d'une société et permettent de soulever et de discuter de nombreuses questions sociétales, telles que les questions de santé, les différences culturelles et les relations publiques, les effets de ces moyens sur les sociétés variant entre le négatif et le positif[2] :

I.4.1 L'impact positifs

- **Amélioration de la productivité scientifique**

Les réseaux sociaux améliorent la productivité scientifique des sociétés, car ils permettent à des groupes d'utilisateurs partageant les mêmes intérêts scientifiques dans divers domaines tels que la santé, la politique, l'économie, etc. de communiquer entre eux pour partager leurs connaissances et leur expertise.

- **Une plateforme médiatique gratuite**

Les réseaux sociaux sont une plateforme médiatique qui permet aux gens d'exprimer librement leurs opinions sans être liés par des lois répressives qui empêchent la liberté d'expression.

- **Faciliter la communication avec les autres**

Les réseaux sociaux sont d'une grande importance lorsqu'il s'agit de communiquer les uns avec les autres, car ils constituent un moyen facile de communiquer avec d'autres utilisateurs

ou d'apprendre à les connaître où qu'ils se trouvent, sans être limités par des frontières géographiques, ce qui peut briser les barrières culturelles entre différentes sociétés.

- **Sensibilisation aux problèmes de la communauté**

Les réseaux sociaux permettent aux membres des communautés de discuter de questions importantes pour ces communautés, qu'il s'agisse de questions environnementales, éthiques ou autres, ce qui sensibilise davantage les gens à ces questions, sans compter que le fait d'aborder ces questions et d'y sensibiliser les gens a fait pencher la balance du pouvoir médiatique en faveur du public, ce qui est difficile à réaliser pour les plateformes médiatiques traditionnelles.

- **Renforcement du pouvoir économique**

De nombreuses entreprises commerciales utilisent les réseaux sociaux pour communiquer avec leurs clients et commercialiser leurs produits, et les réseaux sociaux sont économiquement avantageux en raison du faible coût financier de la recherche économique sur un marché spécifique.

- **Réduire le taux de chômage**

Des études indiquent que 60 % des employeurs utilisent les réseaux sociaux pour rechercher des employés, et 19 % d'entre eux choisissent une personne pour occuper un poste spécifique en fonction des informations disponibles sur cette personne via ses comptes de réseaux sociaux, selon une enquête menée par Career Builder sur le recrutement dans les réseaux sociaux en 2016. Les réseaux sociaux sont définis comme un ensemble d'applications et de sites web qui sont utilisés pour communiquer avec d'autres personnes, diffuser des informations, partager du contenu, etc. Cela a un impact négatif sur le bien-être mental des jeunes. L'exposition à des images idéalisées sur ces réseaux peut influencer de manière défavorable leur perception de soi et les normes de beauté, altérant ainsi leur image corporelle.[2]

I.4.2 L'impact négatifs

- **Réduction de la communication en face à face**

Les réseaux sociaux limitent la communication en face à face entre les membres de la société, car de nombreuses communications se font par écrit sur un clavier d'ordinateur ou un téléphone portable, ce qui peut avoir un effet négatif et limiter les capacités réelles de communication avec les autres.

- **Difficulté à communiquer des émotions**

Les réseaux sociaux limitent la capacité de communication émotionnelle entre les membres de la société, car ils se limitent à l'envoi d'émoticônes pour exprimer la tristesse ou la joie, mais ces symboles utilisés n'interprètent pas nécessairement les véritables sentiments de leur propriétaire, ce qui affecte les relations entre les individus.

- **La paresse**

Les réseaux sociaux sont parfois à l'origine d'un sentiment de paresse, car il est plus facile d'utiliser les réseaux sociaux pour communiquer avec les autres que d'aller les rencontrer en personne, par exemple, ce qui crée une sorte de léthargie et de paresse chez la personne.

- **Impact sur les relations familiales**

L'impact des réseaux sociaux sur la désintégration de la famille et de la société, car les réseaux sociaux jouent un rôle dans la distanciation des membres de la famille les uns par rapport aux autres ; en effet, chacun d'entre eux passe de longues heures à utiliser ces sites au lieu de communiquer avec le reste des membres de sa famille et de renforcer leurs relations avec eux.

- **La diffusion de rumeurs et de fausses nouvelles**

Les réseaux sociaux peuvent être utilisés par certaines personnes pour répandre des rumeurs parmi les membres de la communauté et les induire en erreur avec de fausses nouvelles.

- **Atteinte à la vie privée**

Les réseaux sociaux sont parfois à l'origine de nombreux problèmes auxquels les utilisateurs sont confrontés en raison de la violation de leur vie privée par le biais de ces sites, que ce soit en accédant à leur contenu personnel par ce biais ou en connaissant leur position géographique actuelle.

- **L'impact négatif sur la santé des membres de la communauté**

Les réseaux sociaux jouent un rôle majeur dans l'impact négatif sur la santé de leurs utilisateurs, notamment en ce qui concerne le manque de sommeil.

- **La cyberintimidation**

Les réseaux sociaux sont parfois utilisés à mauvais escient contre les enfants dans la société, les exposant à la cyberintimidation, ce qui peut affecter les enfants et leur santé mentale, car cela peut provoquer la dépression et l'anxiété.

- **Développer de mauvaises habitudes**

Les réseaux sociaux peuvent parfois amener les gens à prendre de mauvaises habitudes, comme se négliger les uns les autres en parlant ou être distraits au volant, et l'utilisation constante de ces sites peut amener les gens à remettre leurs tâches à plus tard.

- **Elle peut entraîner une augmentation de la phobie sociale**

L'utilisation excessive des réseaux sociaux accroît le sentiment de solitude et d'isolement, et donc la phobie sociale, car elle peut amener les gens à perdre confiance en eux, en leur apparence et en ce qu'ils possèdent, et les empêcher de participer à des événements sociaux et de traiter directement avec leur entourage, ce qui leur cause de nombreux problèmes psychologiques et de santé lorsqu'ils sont exposés à de telles situations.[2]

I.5 Qu'est-ce qu'un comportement antisocial ?

Le comportement antisocial peut être caché ou manifeste. Souvent, le comportement antisocial caché implique des manifestations évidentes de préjudice intentionnel, comme le vandalisme, l'incendie criminel, l'effraction et le vol. Les comportements antisociaux manifestes sont ouverts et agressifs, tels que les insultes, les brimades et les bagarres. Les comportements cachés ou manifestes comprennent également la consommation de drogues et d'alcool, ainsi que d'autres tendances antisociales à haut risque.

Ceux qui commettent ces actes manquent généralement de considération pour les autres et leurs expériences. Les comportements antisociaux violent souvent les droits des autres et peuvent les mettre en danger, ainsi que d'autres personnes. Les personnes qui présentent des tendances antisociales n'éprouvent souvent aucun remords pour leurs erreurs, ce qui fait qu'il est difficile de les tenir pour responsables. Ce type de comportement peut être observé dans le cadre de divers troubles mentaux. Il est donc important de comprendre quels autres symptômes peuvent être en cause.[3]

I.6 Types de comportements antisociaux

Les comportements antisociaux peuvent varier en fonction du type de comportement et de l'impact qu'il a sur les autres.

Il existe trois catégories de comportements antisociaux :

I.6.1 La cyberintimidation

Selon le dictionnaire Oxford 13, l'intimidation est « l'utilisation de la force ou du pouvoir pour effrayer ou blesser des personnes plus faibles ». Il s'agit d'un comportement d'un individu ou d'un groupe qui blesse intentionnellement les autres sur le plan physique ou émotionnel, généralement sur une longue période. Il peut s'agir de taquineries, de rumeurs ou de blessures physiques. Les intimidations peuvent également être décrites comme un problème psychosocial, un comportement intentionnel et agressif à l'égard d'une victime plus faible, de manière répétée et sur la durée, avec des conséquences négatives pour les deux parties.[4]

I.6.2 Le cyberharcèlement

Le harcèlement est un comportement hostile qui comprend l'utilisation d'insultes, d'attaques mentales ou physiques et d'agressions, également décrit comme le harcèlement est une intimidation excessive. Dans la littérature relationnelle, le harcèlement est souvent utilisé comme synonyme de victimisation, alors que le harcèlement se réfère à un éventail plus large de comportements hostiles. Il est également utilisé pour décrire les hostilités entre adultes dans lesquelles il y a généralement un déséquilibre de pouvoir.[4]

I.6.3 La misogynie

La misogynie est considérée comme un comportement antisocial. La misogynie désigne la haine, le mépris ou les préjugés à l'égard des femmes, souvent ancrés dans des croyances stéréotypées et des perceptions négatives des femmes. Ce comportement va à l'encontre des valeurs sociétales qui prônent l'égalité et le respect mutuel entre les individus. Ce type de comportement hostile est le titre de notre mémo et sera abordé dans la section suivante.

I.7 Définition de la misogynie

La misogynie est la haine, le mépris ou les préjugés à l'égard des femmes ou des jeunes filles. Il s'agit d'une forme de sexisme qui peut maintenir les femmes dans un statut social inférieur à celui des hommes, maintenant ainsi les rôles sociaux du patriarcat. La misogynie est

largement répandue depuis des milliers d'années. Elle se reflète dans l'art, la littérature, la structure sociétale humaine, les événements historiques, la mythologie, la philosophie et la religion dans le monde entier.

Un exemple de misogynie est la violence à l'égard des femmes, qui comprend la violence domestique et, dans ses formes les plus extrêmes, le terrorisme misogyne et le féminicide. La misogynie se manifeste aussi souvent par le harcèlement sexuel, la coercition et les tactiques psychologiques visant à contrôler les femmes, en les excluant légalement ou socialement d'une citoyenneté à part entière. Dans certains cas, la misogynie récompense les femmes qui acceptent un statut inférieur.

La misogynie peut être comprise comme une attitude adoptée par des individus, en particulier des hommes, et comme une norme ou un système culturel largement répandu. Parfois, la misogynie se manifeste de manière évidente et audacieuse ; d'autres fois, elle est plus subtile ou déguisée de manière à offrir la possibilité d'un déni plausible.

Dans la pensée féministe, la misogynie comprend également le rejet des qualités féminines. Elle méprise les institutions, le travail, les loisirs ou les coutumes associés aux femmes. Le racisme et d'autres préjugés peuvent renforcer la misogynie et s'y superposer.[5]

I.8 Étymologie et utilisation historique

La pensée misogyne remonterait à l'Antiquité ; Aristote, par exemple, était célèbre pour avoir affirmé que les femmes étaient des versions inférieures et déformées des hommes. Le terme a été inventé au XVIIe siècle et est dérivé du grec *misos*, qui signifie « haine », et *gunē*, qui signifie « femme ».

L'utilisation du terme misogynie pour désigner les actes de mépris envers les femmes a été popularisée dans les années 1970 par les féministes de la deuxième vague. La misogynie était généralement distinguée du sexisme à l'égard des femmes : le premier se caractérisait par la violence, telle que l'agression sexuelle ou le meurtre, tandis que le second se manifestait plus subtilement. Toutefois, au cours de ce que l'on appelle la quatrième vague du féminisme, qui a débuté au début du XXIe siècle, la misogynie est devenue presque interchangeable avec le sexisme et peut être utilisée pour désigner les préjugés à l'égard des femmes, en plus des actes de violence ou de haine dont elles sont la cible. La misogynie a donc acquis de multiples significations qui occupent différents niveaux d'intensité. Certains dictionnaires ont adapté leurs entrées pour refléter ce changement sémantique. En 2002, l'Oxford English Dictionary a

modifié sa définition, passant de la « haine des femmes » à la « haine, aversion ou préjugé à l'égard des femmes ». Le dictionnaire Merriam-Webster lui a emboîté le pas.[6]

I.9 La misogynie dans la société

Selon la pensée féministe, la misogynie existe à la fois en tant que résultat des rôles de genre et en tant que moyen de les imposer. Dans les sociétés historiquement patriarcales, les femmes ont été encouragées à occuper des rôles qui soutiennent les hommes. Bien qu'il soit généralement admis que les droits des femmes se sont considérablement améliorés au cours du 20e siècle, de nombreuses femmes au début du 21e siècle ont continué à faire face à des réactions négatives lorsqu'elles ont assumé des rôles qui défiaient les hommes au lieu de les soutenir. Les femmes sont depuis longtemps confrontées à la misogynie et au sexisme dans le monde du sport, dominé par les hommes. En 1967, la coureuse Kathrine Switzer a été la première femme à participer officiellement au marathon de Boston ; elle a été accostée par le directeur de course qui tentait de la retirer à mi-course. Dans les années 2020, la plupart des athlètes féminines professionnelles étaient toujours moins bien payées que leurs homologues masculins. La misogynie reste présente dans d'autres domaines de la vie. En politique, la misogynie est souvent citée pour expliquer le traitement désobligeant que les femmes politiques subissent parfois aux États-Unis et dans d'autres pays de la part de leurs adversaires, de leurs collègues, des électeurs et des médias. On pense qu'elle se manifeste dans les établissements d'enseignement, sur le lieu de travail et dans d'autres domaines par l'écart de rémunération entre les hommes et les femmes, les disparités en matière de congé parental et l'inégalité de traitement en général. Les femmes continuent d'être moins susceptibles d'occuper des postes de direction dans les entreprises ; la liste des 500 entreprises les plus importantes du magazine Fortune pour 2021 ne comptait que 41 femmes PDG.

La misogynie doit être considérée sous l'angle de l'intersectionnalité. Les femmes blanches cisgenres (dont l'identité de genre correspond au sexe biologique identifié à la naissance) sont confrontées à des obstacles différents de ceux rencontrés par les femmes transgenres (dont l'identité de genre diffère du sexe biologique apparent à la naissance) ou les femmes de couleur. Le terme « transmisogynie » est apparu pour distinguer l'intersection unique de la transphobie et de la misogynie à laquelle les femmes trans ou d'autres personnes trans qui se présentent comme des femmes peuvent être confrontées. En 2008, l'universitaire féministe Moya Bailey a inventé le terme misogy noir pour décrire les préjugés spécifiques auxquels sont confrontées les femmes noires.

À la fin des années 1990, un nouveau terme lié à la misogynie est apparu sur un forum en ligne : incel. Bien que ce terme - abréviation de l'expression « célibataire involontaire » - ait été inventé à l'origine par une femme à la recherche d'une communauté en ligne, il a été adopté par des hommes sexuellement frustrés qui pensaient que les femmes étaient la source de leur malheur. Depuis les années 2000, le terme désigne essentiellement un homme qui « déteste les femmes ». La communauté était historiquement basée en ligne, mais depuis 2014, un certain nombre d'attaques ont eu lieu dans plusieurs pays, y compris les États-Unis, le Canada et le Royaume-Uni, par des hommes qui étaient soit autoproclamés incels, soit liés à la communauté incel. Ces attaques, qui visaient des femmes, ont été qualifiées d'actes de « terrorisme incel ». La première de ces attaques a eu lieu à Isla Vista, en Californie, lorsqu'Elliot Rodger a tué six personnes, prétendument pour se venger des femmes qu'il avait rejetées tout au long de sa vie. Ces cas de violence de masse contre les femmes ont été considérés comme de l'extrémisme misogyne ; à l'autre extrémité du spectre se trouve la misogynie intériorisée, un préjugé subconscient que toute femme ou personne présentant un visage féminin peut nourrir contre elle-même ou contre d'autres femmes. Ces préjugés sont le résultat d'une société imprégnée de rôles et de stéréotypes liés au genre. Par exemple, les femmes qui subissent une misogynie intériorisée peuvent être enclines à associer la féminité à la faiblesse.[6]

I.10 La misogynie dans les réseaux sociaux

Le discours misogyne sur les réseaux sociaux est très répandu et est devenu plus agressif au fil du temps. La misogynie sur les réseaux sociaux comprend à la fois des tentatives individuelles d'intimider et de dénigrer les femmes, en niant l'inégalité entre les sexes (le nouveau sexisme), ainsi que des tentatives collectives coordonnées telles que la campagne de harcèlement misogyne lors du vote et la campagne de harcèlement antiféministe de Gamergate. Dans un article publié dans le Journal of International Affairs, Kim Parker et Olga Jurasch expliquent comment la misogynie sur les réseaux sociaux peut conduire les femmes à rencontrer des obstacles lorsqu'elles tentent de s'engager dans les sphères publiques et politiques sur les réseaux sociaux en raison de la nature offensive de ces espaces. Elles suggèrent également de réglementer et de mettre fin à la misogynie sur les réseaux sociaux par des moyens gouvernementaux et non gouvernementaux.[5]

Basé sur le comportement misogyne, le langage misogyne peut être classé en plusieurs catégories :

- Diffamation.

- Domination.
- Harcèlement sexuel.
- Stéréotypes.
- Objectivation.
- Menaces de violence.

I.11 Causes de la misogynie

- L'exposition à de mauvaises expériences dans l'enfance, comme le fait d'être constamment puni par sa mère dans l'enfance, ou le fait que la mère soit mauvaise avec l'enfant, ce qui génère des sentiments de haine et d'envie.
- Avoir été humilié, rabaissé et puni par une femme ou à cause d'une femme dans le passé.
- Avoir été trahi par des personnes en qui il avait confiance, comme sa femme ou sa petite amie.
- L'exposition à de nombreux rejets de la part de femmes avec lesquelles il souhaitait avoir une relation amoureuse.[7]

I.12 Les méfaits de la misogynie en réseaux sociaux

Une enquête IPSOS MORI menée en 2017 par Amnesty International indique que plus des trois quarts (76 %) des femmes qui ont déclaré avoir été victimes ou harcelées sur une plateforme de réseaux sociaux ont modifié la façon dont elles utilisaient ces plateformes. Elles ont notamment restreint le contenu de leurs publications :

- 32 % des femmes ont cessé de publier des contenus exprimant leur opinion sur certaines questions, dans tous les pays.

-61 % ont déclaré avoir été victimes d'abus ou de harcèlement en ligne et avoir souffert d'une baisse d'estime de soi ou d'une perte de confiance en soi.

- Plus de la moitié (55 %) ont souffert de stress, d'anxiété ou de crises de panique après avoir été victimes d'abus ou de harcèlement en ligne.

-En outre, 63 % des personnes interrogées n'ont pas pu dormir correctement à la suite d'un abus ou d'un harcèlement en ligne.

-Les trois quarts (75 %) des Néo-Zélandais ont fait état de cet impact. Plus de la moitié (56 %) ont déclaré que les abus ou le harcèlement en ligne les empêchaient de se concentrer pendant

de longues périodes. Environ un quart (24 %) des personnes ayant déclaré avoir été victimes d'abus ont dit que cela leur faisait craindre pour la sécurité de leur famille.[8]

I.13 Moyens populaires de lutter contre la misogynie sur les réseaux sociaux (Twitter)

- Un hashtag est une étiquette descriptive couramment utilisée qui a largement circulé sur les réseaux sociaux à la fin de l'année 2017 pour encourager les femmes à témoigner des agressions sexuelles et du harcèlement qu'elles ont subis et qu'elles subissent encore.

- Outre les hashtags, il existe de nombreuses organisations très actives dans le domaine de la lutte contre les violences faites aux femmes sur les réseaux sociaux. Il s'agit de groupes féministes qui se mobilisent sur les réseaux sociaux pour lutter contre le harcèlement de rue, le sexisme en ligne ou la publicité sexiste : Ils créent des bulletins d'information et des podcasts et organisent des réunions et des festivals.

Dans notre recherche, nous proposons de résoudre ce problème en créant un logiciel qui identifie et catégorise les commentaires misogynes sur les réseaux sociaux (Twitter), que nous étudierons en détail au chapitre 3.

I.14 Conclusion

Dans ce chapitre, nous avons également abordé l'étude et l'analyse de la question de la misogynie dans la société en général et dans les réseaux sociaux, en nous concentrant sur sa définition, l'origine du mot et l'histoire de son utilisation, ainsi que sur ses causes et ses types en général et ses effets néfastes, et sur la connaissance des comportements hostiles et de leurs types et de leur relation avec la misogynie, et enfin nous avons proposé un ensemble de moyens communs pour la combattre et l'éliminer.

Chapitre



**Machine Learning et détection
misogynie**

II.1 Introduction

Dans le premier chapitre, nous avons étudié le concept de réseaux sociaux ainsi que le concept de misogynie sur les réseaux sociaux, qui est à son tour un type de problème auquel les femmes sont confrontées sur les réseaux sociaux et qui est défini comme un discours misogyne qui se propage largement sur les réseaux sociaux et devient plus agressif au fil du temps, Nous expliquerons et décrirons donc quelques méthodes intéressantes de détection des messages misogynes et quelques défis visant à créer des modèles automatiques de détection des messages misogynes dans les réseaux sociaux. Ce chapitre est organisé comme suit :

Dans la sous-section 1, nous expliquons ce que sont l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond en général, dans la sous-section 2, nous décrivons les méthodes et outils existants pour détecter et catégoriser la misogynie dans les réseaux sociaux, et enfin, la sous-section 3 présente une analyse et une discussion de certains résultats pertinents obtenus par d'autres chercheurs.

II.2 Qu'est-ce que l'intelligence artificielle ?

L'intelligence artificielle (IA) est la théorie et le développement de systèmes informatiques capables d'effectuer des tâches qui, historiquement, ont nécessité l'intelligence humaine, telles que la reconnaissance vocale, la prise de décision et la reconnaissance des formes. L'IA est un terme générique qui englobe un large éventail de technologies, notamment l'apprentissage automatique, l'apprentissage profond et le traitement du langage naturel (NLP).

Bien que ce terme soit couramment utilisé pour décrire une série de technologies différentes utilisées aujourd'hui, nombreux sont ceux qui ne s'accordent pas sur la question de savoir si ces technologies constituent réellement de l'IA. Certains soutiennent au contraire qu'une grande partie de la technologie utilisée dans le monde réel aujourd'hui constitue en fait un apprentissage automatique très avancé et n'est qu'une première étape vers la véritable IA, ou "intelligence artificielle générale" (IAG).

Cependant, malgré les nombreuses controverses philosophiques sur l'existence de "vraies" machines intelligentes, la plupart des gens utilisent aujourd'hui le terme "IA" pour désigner une série de technologies basées sur l'apprentissage automatique, telles que le Chat GPT ou la vision par ordinateur, qui permettent aux machines d'effectuer des tâches que seuls les humains pouvaient auparavant réaliser, telles que la création de contenu écrit, la conduite d'une voiture ou l'analyse de données.[9]

II.3 Exemples d'intelligence artificielle :

Bien qu'aucun robot humanoïde ne soit souvent associé à l'IA (pensez à Data de Star Trek : The Next Generation ou au T-800 de Terminator), vous avez probablement déjà interagi à de nombreuses reprises avec des services ou des appareils alimentés par l'apprentissage automatique.

Au niveau le plus élémentaire, l'apprentissage automatique utilise des algorithmes formés sur des ensembles de données pour créer des modèles d'apprentissage automatique qui permettent aux systèmes informatiques d'effectuer des tâches telles que recommander des chansons, déterminer le chemin le plus rapide pour se rendre à une destination ou traduire un texte d'une langue à l'autre. Voici quelques-uns des exemples les plus courants d'IA utilisés aujourd'hui :

- **ChatGPT** : Utilise de grands modèles de langage (LLM) pour créer des textes en réponse à des questions ou à des commentaires.
- **Google Translate** : utilise des algorithmes d'apprentissage profond pour traduire un texte d'une langue à une autre.
- **Netflix** : Utilise des algorithmes d'apprentissage automatique pour créer des moteurs de recommandation personnalisés pour les utilisateurs en fonction de leur historique de visionnage.
- **Tesla** : Utilise la vision par ordinateur pour alimenter les fonctions d'auto-conduite de ses voitures.[9]

II.4 L'intelligence artificielle dans le monde du travail

L'intelligence artificielle (IA) est omniprésente dans de nombreux secteurs. L'automatisation des tâches qui ne nécessitent pas d'intervention humaine permet d'économiser de l'argent et du temps, et peut minimiser le risque d'erreur humaine. Voici quelques exemples d'utilisation de l'IA dans différents secteurs :

- **L'industrie financière** : La détection des fraudes est un cas d'utilisation notable de l'IA dans le secteur financier. La capacité de l'IA à analyser de grandes quantités de données lui permet de détecter des anomalies ou des schémas indiquant un comportement frauduleux.

- **Le secteur de la santé :** Les robots dotés d'IA peuvent faciliter les opérations chirurgicales à proximité d'organes ou de tissus très sensibles afin de limiter les pertes de sang ou les risques d'infection.[9]

II.5 Qu'est-ce que l'intelligence générale artificielle (AGI) ?

L'intelligence artificielle générale (AGI) désigne un état théorique dans lequel les systèmes informatiques sont capables d'atteindre ou de dépasser l'intelligence humaine. En d'autres termes, l'AGI est la "véritable" intelligence artificielle décrite dans d'innombrables romans de science-fiction, émissions de télévision, films et bandes dessinées.

En ce qui concerne la signification exacte de l'AGI elle-même, les chercheurs ne sont pas tout à fait d'accord sur la manière de reconnaître la "véritable" AGI lorsqu'elle apparaît. Cependant, l'approche la plus connue pour déterminer si une machine est intelligente ou non est connue sous le nom de test de Turing ou jeu d'imitation, une expérience décrite pour la première fois par l'influent mathématicien, informaticien et cryptanalyste Alan Turing dans un article de 1950 sur l'intelligence informatique. Turing y décrit un jeu à trois joueurs dans lequel un "interrogateur" humain est invité à communiquer par texte avec un autre humain et une machine et à juger qui a composé chaque réponse. Si l'interrogateur ne peut pas reconnaître l'humain de manière fiable, Turing affirme que la machine peut être considérée comme intelligente.

Pour compliquer encore les choses, les chercheurs et les philosophes ne parviennent pas non plus à se mettre d'accord sur la question de savoir si nous avons commencé à réaliser une IA durable, si nous en sommes encore loin ou si c'est tout à fait impossible. Par exemple, alors qu'un document récent de Microsoft Research et OpenAI affirme que le chat GPT-4 est une forme précoce d'IA SDI, de nombreux autres chercheurs sont sceptiques quant à ces affirmations et soutiennent qu'il s'agit simplement de propagande.

Quelle que soit la distance qui nous sépare de l'AGI, vous pouvez supposer que lorsque quelqu'un utilise le terme AGI, il fait référence au type de programmes informatiques et de machines sensibles que l'on trouve couramment dans la science-fiction populaire.[9]

II.6 Types d'IA

Alors que les chercheurs tentent de construire des formes plus avancées d'IA, ils doivent également commencer à formuler une compréhension plus nuancée de ce que signifie

exactement être intelligent ou même conscient. Pour tenter de clarifier ces concepts, les chercheurs ont identifié quatre types d'IA.

Voici un résumé de chaque type d'IA, selon le professeur Arend Hintze de l'université du Michigan :

II.6.1 Machines interactives

Les machines interactives sont le type d'IA le plus basique. Les machines ainsi conçues n'ont aucune connaissance des événements passés, mais ne font que "réagir" à ce qui se trouve devant elles à un moment donné. Par conséquent, elles ne peuvent effectuer que certaines tâches avancées dans un cadre très étroit, comme jouer aux échecs, et sont incapables d'effectuer des tâches en dehors de ce contexte limité.

II.6.2 Les machines à mémoire limitée

Les machines à mémoire limitée ont une compréhension limitée des événements passés. Elles peuvent interagir avec le monde qui les entoure davantage que les machines à mémoire réactive. Par exemple, les voitures autonomes utilisent une forme de mémoire limitée pour prendre des virages, surveiller les véhicules qui s'approchent et ajuster leur vitesse. Toutefois, les machines dotées d'une mémoire limitée ne peuvent pas se faire une idée complète du monde, car leur récupération des événements passés est limitée et n'est utilisée que dans un laps de temps restreint.

II.6.3 Machines à théorie de l'esprit

Les machines à théorie de l'esprit représentent une forme précoce d'intelligence artificielle générale. En plus d'être capables de créer des représentations du monde, les machines de ce type auront également une compréhension des autres entités qui existent dans le monde. À l'heure actuelle, cette réalité n'a pas encore été concrétisée.

II.6.4 Machines conscientes d'elles-mêmes

Les machines conscientes d'elles-mêmes constituent le type d'IA le plus avancé sur le plan théorique, et elles auront une compréhension du monde, des autres et d'elles-mêmes. C'est ce que la plupart des gens veulent dire lorsqu'ils parlent d'une IA consciente d'elle-même. Pour l'instant, il s'agit d'une réalité lointaine.[9]

II.7 Avantages et risques de l'intelligence artificielle

L'intelligence artificielle (IA) a un éventail d'applications qui ont le potentiel de changer notre façon de travailler et de vivre au quotidien. Si nombre de ces transformations sont passionnantes, comme les voitures autonomes, les assistants virtuels ou les dispositifs portables dans le domaine de la santé, elles posent également de nombreux défis.

Il s'agit d'un tableau complexe qui évoque souvent des images contradictoires : Utopie pour les uns, dystopie pour les autres. La réalité sera probablement bien plus complexe. Voici quelques-uns des avantages et des risques potentiels de l'IA :

Avantages potentiels	Dangers potentiels
Une plus grande précision pour certaines tâches répétitives, telles que l'assemblage de véhicules ou d'ordinateurs.	La perte d'emplois due à l'automatisation croissante.
Diminution des coûts opérationnels grâce à une plus grande efficacité des machines.	Risque de partialité ou de discrimination en raison de l'ensemble de données sur lequel l'IA est entraînée.
Personnalisation accrue des services et produits numériques.	Problèmes éventuels de cybersécurité.
Amélioration de la prise de décision dans certaines situations.	Manque de transparence sur la manière dont les décisions sont prises, ce qui entraîne des solutions moins qu'optimales.
Capacité à générer rapidement de nouveaux contenus, tels que du texte ou des images.	Possibilité de créer des informations erronées et de violer par inadvertance des lois et des règlements.

Tableau 1: Avantages et risques de l'intelligence artificielle.

Ce ne sont là que quelques-unes des façons dont l'IA présente à la fois des avantages et des risques pour la société. Lorsque l'on utilise de nouvelles technologies comme l'IA, il est préférable d'avoir une vision claire de ce qui est inclus et de ce qui est exclu. Après tout, un grand pouvoir implique de grandes responsabilités.[9]

II.8 Qu'est-ce que l'apprentissage automatique (Machine Learning) ?

L'apprentissage automatique est une spécialité de l'intelligence artificielle (IA) qui donne aux machines la capacité d'apprendre automatiquement à partir de données et d'expériences antérieures.

Les méthodes d'apprentissage automatique permettent aux ordinateurs de travailler de manière autonome sans programmation explicite. Les applications d'apprentissage automatique sont alimentées par de nouvelles données et peuvent apprendre, évoluer, se développer et s'adapter de manière autonome.

L'apprentissage automatique permet d'apprendre à partir de grandes quantités de données en utilisant des algorithmes pour identifier des modèles et apprendre dans un processus itératif.

Les performances des algorithmes d'apprentissage automatique s'améliorent de manière adaptative à mesure que le nombre d'échantillons disponibles au cours du processus d'apprentissage augmente, par exemple l'apprentissage profond.[10]

II.9 Types d'apprentissage automatique (Machine Learning)

Les algorithmes d'apprentissage automatique peuvent être formés de différentes manières, et chaque méthode a ses propres avantages et inconvénients. Sur la base de ces méthodes, l'apprentissage automatique est généralement classé en quatre catégories principales :[10]

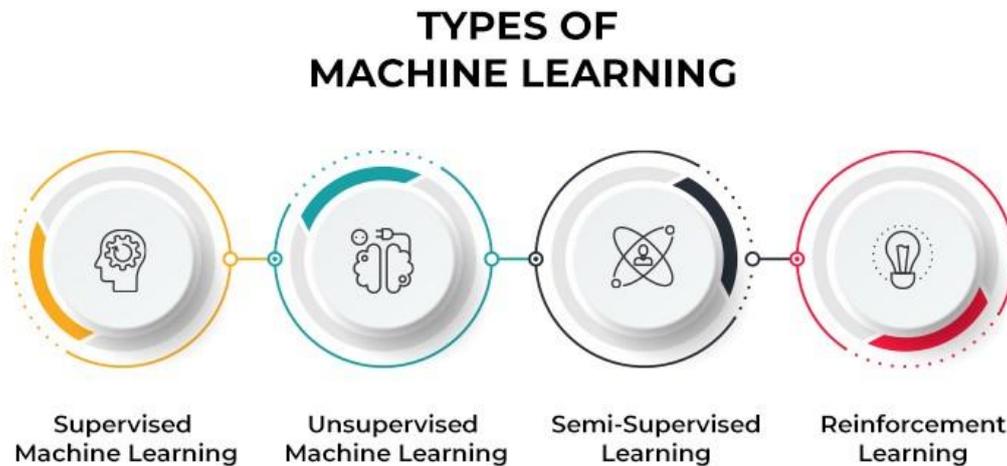


Figure 2: Types d'apprentissage automatique.

II.9.1 Apprentissage automatique supervisé

Ce type d'apprentissage automatique implique une supervision, où les machines sont formées sur des ensembles de données étiquetées et habilitées à prédire les résultats sur la base de la formation fournie.[10]

II.9.2 Apprentissage automatique non supervisé

L'apprentissage non supervisé fait référence à une technique d'apprentissage non supervisée. Dans ce cas, la machine est formée à l'aide d'un ensemble de données non étiquetées et est capable de prédire la sortie sans aucune supervision. Un algorithme d'apprentissage non supervisé vise à regrouper l'ensemble de données non étiquetées sur la base des similitudes, des différences et des modèles dans les entrées.[10]

II.9.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé englobe les caractéristiques de l'apprentissage automatique supervisé et non supervisé. Il utilise une combinaison d'ensembles de données étiquetés et non étiquetés pour former ses algorithmes. En utilisant les deux types d'ensembles de données, l'apprentissage semi-supervisé surmonte les inconvénients des options ci-dessus.[10]

II.9.4 Apprentissage par renforcement

L'apprentissage par renforcement est un processus axé sur le retour d'information. Dans ce cas, le composant d'IA évalue automatiquement son environnement par le biais d'une méthode de type "hit-and-run", prend des mesures, tire des enseignements de ses expériences et optimise ses performances. Le composant est récompensé pour chaque bonne action et pénalisé pour chaque mauvaise action. Ainsi, l'apprentissage par renforcement vise à maximiser les récompenses en réalisant de bonnes actions.[10]

II.10 Les 3 principales applications de l'apprentissage automatique (Machine Learning)

Les secteurs industriels qui traitent de grandes quantités de données ont reconnu l'importance et la valeur de la technologie de l'apprentissage automatique. Comme l'apprentissage automatique tire des enseignements des données en temps réel, les organisations qui l'utilisent peuvent fonctionner efficacement et prendre une longueur d'avance sur leurs concurrents.

Dans ce monde numérisé qui évolue rapidement, tous les secteurs d'activité bénéficient grandement de la technologie de l'apprentissage automatique. Nous examinons ici les cinq secteurs les plus importants pour les applications d'apprentissage automatique.[10]

II.10.1 Secteur du commerce de détail

Les sites web de vente au détail utilisent largement l'apprentissage automatique pour recommander des articles en fonction de l'historique d'achat des utilisateurs. Les détaillants utilisent l'apprentissage automatique pour capturer et analyser des données et offrir des expériences d'achat personnalisées à leurs clients. Ils utilisent également l'apprentissage automatique pour les campagnes de marketing, la connaissance des clients, la planification des marchandises et l'optimisation des prix.[10]

II.10.2 L'industrie du voyage

L'apprentissage automatique joue un rôle essentiel dans l'expansion de l'industrie du voyage. Les trajets proposés par Uber, Ola et même les voitures autonomes s'appuient fortement sur l'apprentissage automatique.

Par ailleurs, l'industrie du voyage utilise l'apprentissage automatique pour analyser les commentaires des utilisateurs. Les commentaires des utilisateurs sont classés par analyse des

sentiments sur la base de notes positives ou négatives. Les entreprises du secteur des voyages s'en servent pour le suivi des campagnes, la surveillance de la marque, le contrôle de la conformité, etc.[10]

II.10.3 Réseaux sociaux

Grâce à l'apprentissage automatique, des milliards d'utilisateurs peut interagir efficacement sur les réseaux sociaux. L'apprentissage automatique joue un rôle essentiel dans le fonctionnement des plateformes de réseaux sociaux, qu'il s'agisse de personnaliser les fils d'actualité ou de diffuser des publicités spécifiques à l'utilisateur. Par exemple, la fonction d'étiquetage automatique de Facebook utilise la reconnaissance d'images pour reconnaître et étiqueter automatiquement le visage d'un ami. Le réseau social utilise ANN pour reconnaître les visages familiers dans les listes de contacts des utilisateurs et faciliter le processus de marquage automatique.

De même, LinkedIn sait quand vous devez postuler à votre prochain emploi, avec qui vous devez entrer en contact et comment vos compétences se situent par rapport à celles de vos pairs. Toutes ces fonctionnalités sont rendues possibles par l'apprentissage automatique.[10]

II.11 Les 10 Principales tendances de l'apprentissage automatique en 2022

L'apprentissage automatique a eu un impact significatif sur toutes les verticales industrielles à travers le monde, des startups aux entreprises Fortune 500. Selon un rapport de Fortune Business Insights pour 2021, la taille du marché mondial de l'apprentissage automatique était de 15,50 milliards de dollars en 2021 et devrait atteindre 152,24 milliards de dollars d'ici 2028 à un taux de croissance annuel composé de 38,6 %.

Compte tenu de l'adoption croissante de l'apprentissage automatique, l'année 2022 devrait connaître une trajectoire similaire. Voici les 10 principales tendances de l'apprentissage automatique pour 2022 :

1. La blockchain rencontre l'apprentissage automatique.
2. Outils libre-service basés sur l'IA.
3. Assistants IA et moteurs de recherche personnalisés.
4. Assistance intelligente tout compris.
5. Dispositifs médicaux personnels.

6. Réalité augmentée (RA) améliorée.
7. Avancées dans l'industrie automobile.
8. Apprentissage profond complet.
9. Réseau contradictoire génératif (GAN).
10. Minuscule ML.[10]

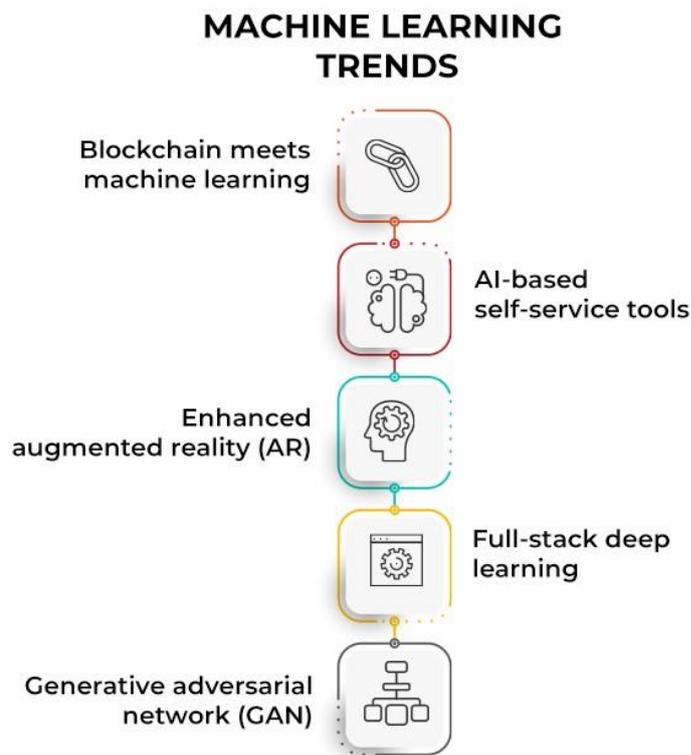


Figure 3: Tendances de l'apprentissage automatique à attendre en 2022.

II.12 Techniques d'apprentissage automatique

Dans l'apprentissage automatique, un large éventail d'outils et de techniques sont utilisés pour permettre aux systèmes informatiques de comprendre et de s'adapter à de nouvelles situations sur la base des données et des expériences qu'ils collectent. Voici quelques-unes des techniques que nous avons utilisées dans le cadre de nos recherches :

II.12.1 Natural language processing (NLP)

- **Définition de NLP**

Natural language processing (NLP), ou traitement automatique des langues (TALN), est une branche de l'intelligence artificielle qui s'attache à donner la capacité aux machines de comprendre, générer ou traduire le langage humain tel qu'il est écrit et/ou parlé. Les chatbots

figurent par parmi les logiciels de NLP les plus populaires. Autres applications de NLP célèbres : les assistants vocaux Alexa, Google Home ou encore Siri.[11]

- **Utilisations de NLP**

Le natural language processing a pour but de doter les logiciels de processus de traitement automatique du langage vocal ou textuel. Partant de là, il recouvre de nombreux cas d'usage plus ou moins élaboré :

- ✓ La classification de texte.
- ✓ La reconnaissance de texte.
- ✓ Le résumé automatique.
- ✓ La traduction automatique.
- ✓ Les chatbot, voicebot ou callbot.
- ✓ Les assistants intelligents.[11]

- **Fonctionnel de NLP**

Le NLP combine intelligence artificielle et traitement linguistique. La dernière génération des technologies de NLP s'adosse à des réseaux de neurones artificiels ou de simples modèles de machine learning statistiques. Des modèles d'apprentissage auront été entraînés sur des volumes importants de texte.

L'objectif peut cibler plusieurs types de traitement automatique : le speech-to-text et le text-to-speech, la reconnaissance d'entités nommées (noms de personnes, lieux...), l'analyse de sentiments (positif, négatif, neutre), la synthèse de texte, l'extraction d'aspects (ciblage de l'intention du texte) ou encore la modélisation de sujets.[11]

- **Deux types de modèles de Machine Learning en NLP**

Globalement, le natural language processing se décline en deux grandes catégories de modèles de Machine Learning :

- Les modèles de Machine Learning orientés NLU (natural language understanding) qui s'attachent à saisir le sens d'une langue et d'un discours dans son contexte.
- Les modèles de Machine Learning orientés NLG (natural language generation) qui ont pour but de générer un texte à la manière d'un humain.[11]

II.12.2 L'apprentissage e profond (Deep Learning)

- **Qu'est-ce que l'apprentissage profond ?**

L'apprentissage profond est un sous-domaine de l'apprentissage automatique et représente un ensemble d'architectures de réseaux neuronaux qui permettent de résoudre des problèmes complexes et sophistiqués. Ces architectures (ou modèles) sont connues sous le nom de réseaux neuronaux convolutifs (CNN) et de mémoires à long terme (LSTM), entre autres.

Les architectures d'apprentissage profond sont appelées profondes parce qu'elles comportent de nombreuses couches. La profondeur des couches est importante, comme vous le verrez dans cette section. L'apprentissage en profondeur fait référence aux architectures de réseaux neuronaux qui comprennent de nombreuses couches et qui ont la capacité d'apprendre (par l'entraînement) à mettre en correspondance une entrée, telle qu'une image, avec une ou plusieurs sorties, telles qu'une classification. La classification peut indiquer si l'image contient ou non un chat. Au fur et à mesure que les entreprises expérimentent des problèmes à forte intensité de données tels que la synthèse vocale et la vision par ordinateur, les scientifiques des données se sont tournés vers l'apprentissage profond pour résoudre des problèmes commerciaux qui ne peuvent pas être résolus à l'aide d'algorithmes non linéaires.[12]

- **Types de réseaux neuronaux profonds**

Nous allons maintenant approfondir ces architectures pour voir comment elles sont décomposées, ainsi que les méthodes de formation utilisées pour chacune d'entre elles.[12]

1. Réseaux neuronaux récurrents

Les réseaux neuronaux récurrents (RNN) se présentent sous différents styles architecturaux, mais tous incluent le comportement de l'état interne, ce qui signifie qu'ils peuvent être appliqués à des problèmes dans le domaine temporel. Comme le montre la figure 4 :[12]

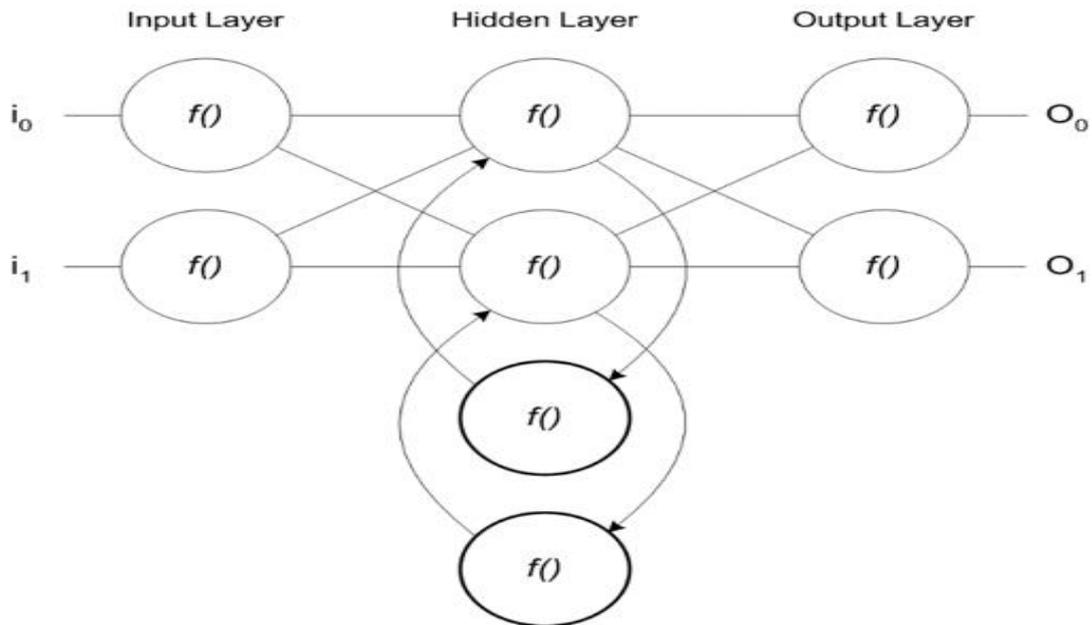


Figure 4: Déroulement de réseau neuronal récurrent.

2. Convolutional Neural Networks

Le CNN est une architecture de réseau neuronal profond qui excelle dans les tâches de classification d'images. Il fonctionne par le biais d'une série d'opérations de convolution et de mise en commun maximale sur une entrée (voir figure 5) :[12]

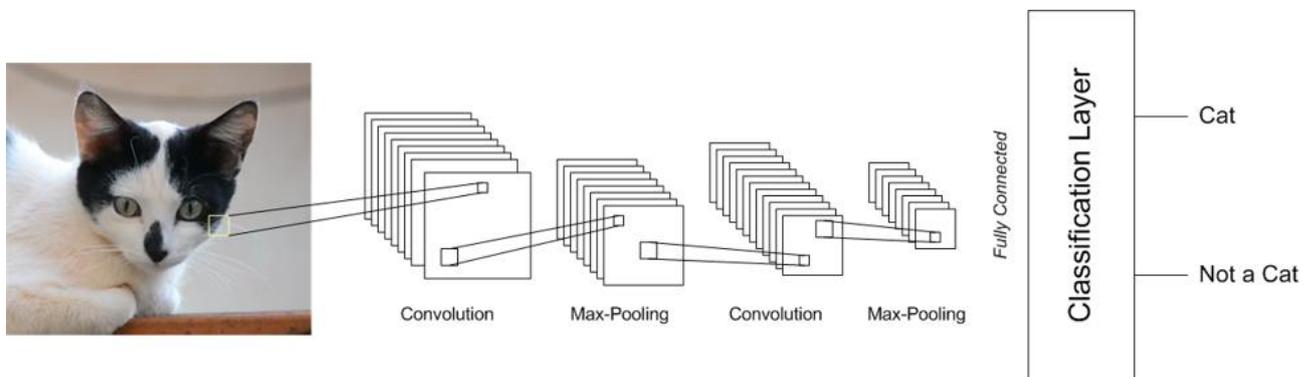


Figure 5: Convolutional Neural Network.

3. Long Short-Term Memory Networks

Les réseaux LSTM sont des RNN dotés d'une mémoire interne. Un bloc LSTM est utilisé pour construire un réseau qui existe en une dimension (comme le montre la figure 6) ou en plusieurs dimensions, où chaque bloc alimente les blocs situés à droite et au-dessus.[12]

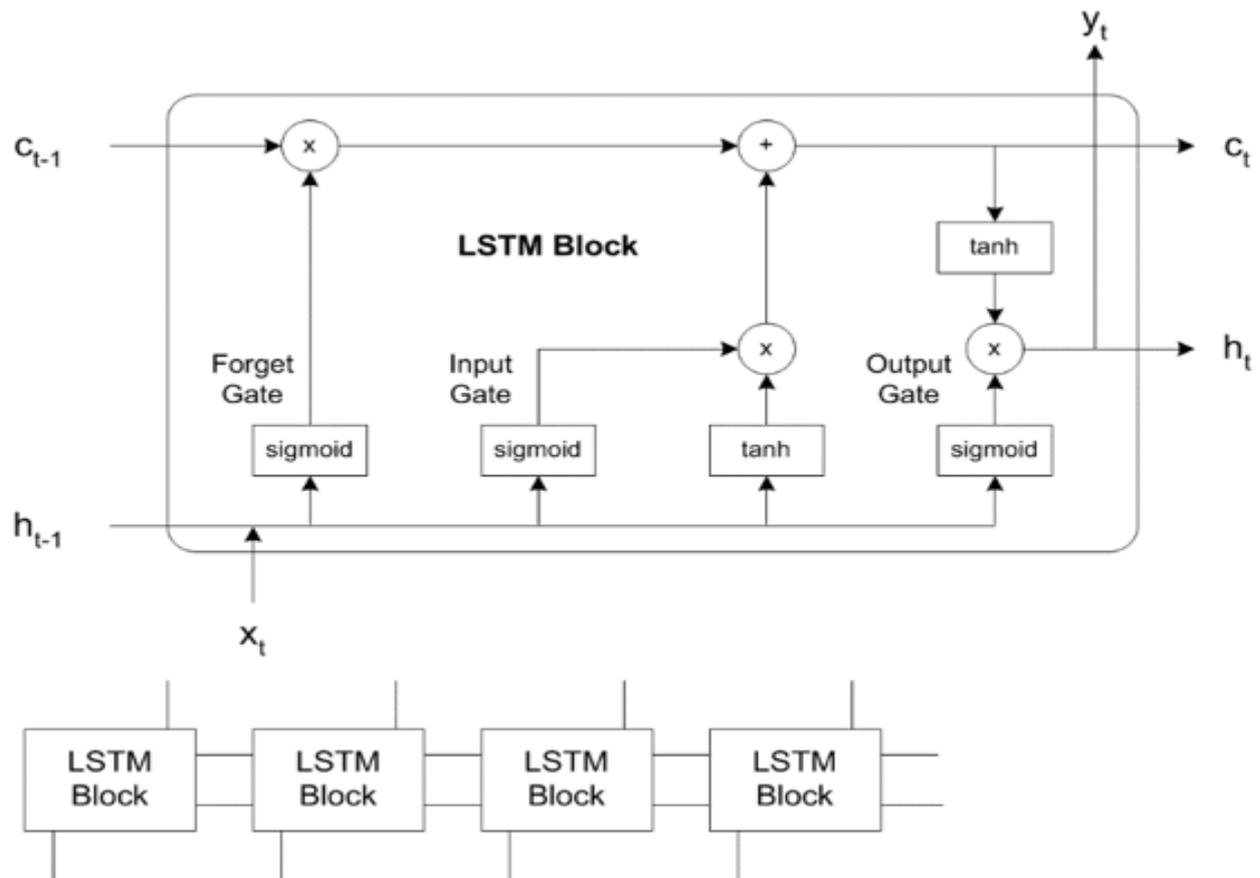


Figure 6: Long Short-Term Memory Network and Block.

• Techniques d'apprentissage profond

L'apprentissage profond représente un ensemble d'architectures fondées sur les idées des réseaux neuronaux. Les réseaux neuronaux sont des structures informatiques - des réseaux d'éléments informatiques qui peuvent être ajustés par le biais d'une formation, puis appliqués à des problèmes. Revenons à l'origine du Deep Learning et explorons les principes fondamentaux des réseaux neuronaux.[12]

✓ Des neurones aux réseaux multicouches

Un réseau neuronal est un réseau de neurones qui met en œuvre une fonction mathématique. Une entrée est introduite dans le réseau (généralement sous la forme d'un vecteur), les neurones individuels calculent leurs sorties dans chaque couche jusqu'à la sortie (figure 7).[12]

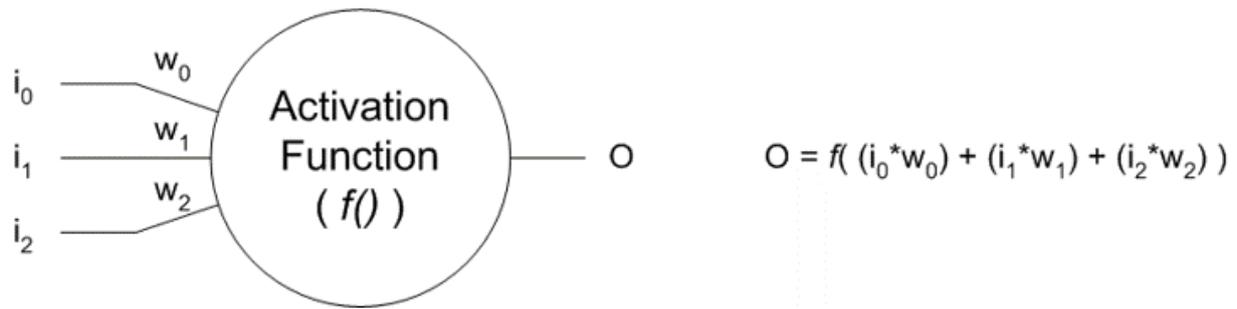


Figure 7: Un neurone unique et son équivalent mathématique.

Pour les réseaux multicouches, la fonction d'activation est choisie pour introduire la non-linéarité. L'introduction de la non-linéarité permet aux réseaux multicouches de résoudre des problèmes relativement complexes avec un petit nombre de neurones.[12]

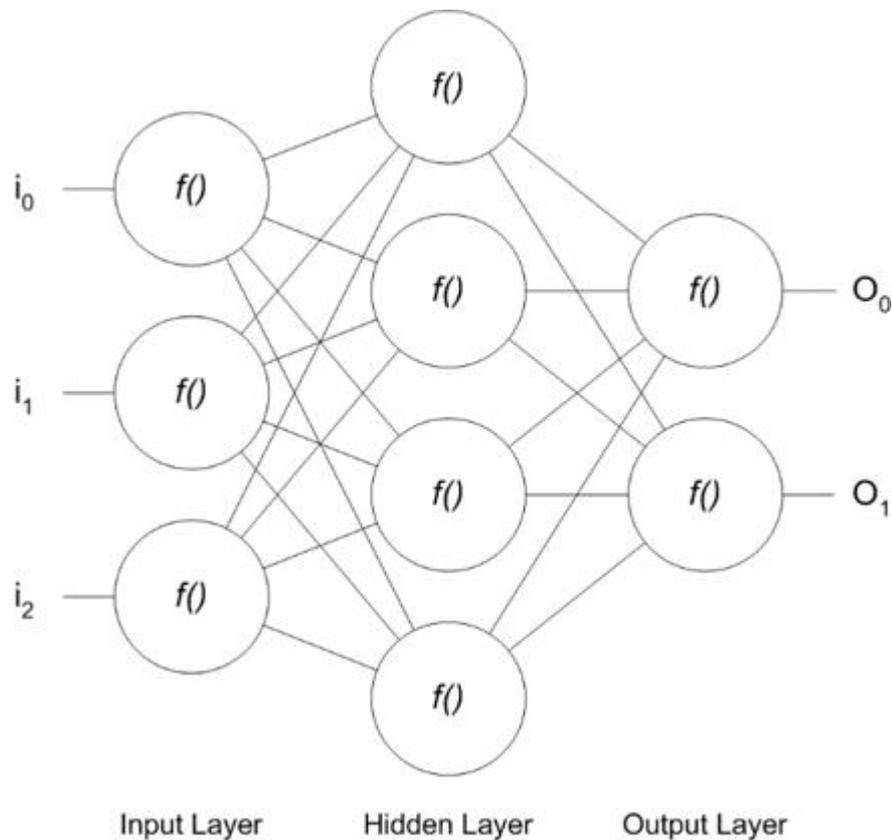


Figure 8: Un réseau neuronal multicouche (3 couches).

✓ **Formation**

Les poids du réseau sont définis individuellement et servent de base à la mise en correspondance des entrées et des sorties. La définition de ces poids constitue l'apprentissage,

qui s'effectue sur de nombreuses itérations afin d'ajuster les poids pour faire correspondre l'entrée à la sortie avec un niveau d'erreur acceptable.[12]

- **Cas d'utilisation de l'apprentissage profond**

Cette section explore certains des cas d'utilisation de pointe appliqués aujourd'hui à l'apprentissage profond :

- ✓ Reconnaissance d'images/d'objets
- ✓ Légende des images et des vidéos
- ✓ Reconnaissance de la parole
- ✓ Traduction linguistique
- ✓ Résumé automatique de texte
- ✓ Systèmes de questions-réponses .[12]

- **Apprentissage profond et réseaux de neurones (Deep Learning vs. Neural Networks)**

L'une des principales caractéristiques des réseaux neuronaux profonds (DNN) est la profondeur : le nombre de couches en plus de la largeur ou du nombre d'éléments de traitement au sein de chaque couche. Mais les réseaux neuronaux profonds (DNN) ont évolué par rapport aux réseaux multicouches typiques qui nous ont amenés à l'apprentissage profond (DP). Les architectures des réseaux neuronaux profonds (DNN) sont différentes de celles de leurs ancêtres multicouches. Les CNN, qui fonctionnent bien avec les données d'image, échantillonnent et regroupent les pixels de l'image pour les traiter. Les RNN, qui sont idéaux pour les données séquentielles telles que le texte, prennent en compte non seulement une entrée, mais aussi les entrées qui la précèdent et la suivent.

La structure des neurones a également changé au fur et à mesure de l'évolution de l'apprentissage profond (DL). Plutôt que de simples produits pondérés additionnés passant par une fonction d'activation, les neurones des nouveaux réseaux d'apprentissage profond, tels que les réseaux LSTM, comprennent des portes permettant de réguler le flux d'informations et même d'oublier des informations.[12]

II.13 Détection de la misogynie en ligne

II.13.1 Méthodes et outils de détection automatique du langage misogyne

Il est important de mentionner certaines des méthodes et certains des outils que nous avons décrits précédemment et qui ont été appliqués avec succès pour résoudre le problème de la catégorisation des textes, un exemple étant l'identification du discours misogyne dans les réseaux sociaux. D'un point de vue mathématique, nous pouvons essayer d'identifier les modèles qui fonctionnent mieux que d'autres.

Il existe différentes approches du problème de la catégorisation des textes, que nous mentionnerons ci-dessous. Aujourd'hui, outre les modèles non supervisés de classification du discours misogyne, deux approches principales sont couramment utilisées : L'approche par réseau neuronal et l'approche classique d'apprentissage automatique. Il existe également des combinaisons d'approches - combinaisons - qui impliquent la création de différents types de modèles (par exemple, une combinaison de différents modèles d'apprentissage automatique classique ou une combinaison de modèles d'apprentissage automatique classique et de modèles basés sur l'approche des réseaux neuronaux) les meilleurs modèles étant sélectionnés par la suite. Dans cette section, nous décrivons ces méthodes.[13]

- **Approche classique de l'apprentissage automatique (Machine Learning)**

Les modèles basés sur l'approche classique de l'apprentissage automatique sont populaires pour résoudre la catégorisation des textes et donnent de bons résultats. Dans l'apprentissage automatique classique, les chercheurs utilisent une quantité relativement faible de données et identifient les fonctions les plus importantes dans les données que l'algorithme doit prédire. Par modèles basés sur l'apprentissage automatique classique, nous entendons un ensemble de méthodes utilisées pour créer des modèles capables d'apprendre à partir d'observations et de faire des prédictions. Ces modèles utilisent des algorithmes, la régression et des sciences connexes pour donner un sens aux données. Ces algorithmes peuvent généralement être considérés comme des modèles statistiques. Nous énumérons ci-dessous quelques-uns des modèles les plus courants qui ont été appliqués à la catégorisation de textes :[13]

1. Régression logistique (Logistic Regression)

L'un des types de modèles d'apprentissage automatique les plus populaires et les plus efficaces est le modèle de régression logistique. Cette méthode s'applique bien aux problèmes de classification binaire (c'est-à-dire les problèmes pour lesquels nous obtenons l'une des deux classes en sortie). La régression logistique est utilisée pour prédire la probabilité d'un certain événement à partir des valeurs de nombreuses caractéristiques.[13]

2. Naive Bayes

L'algorithme de Naive Bayes est l'une des méthodes les plus efficaces dans le domaine de la classification des textes, mais ce n'est qu'avec un grand nombre d'échantillons d'entraînement qu'il peut obtenir des résultats plus précis. L'exigence d'un grand nombre d'échantillons entraîne non seulement une lourde charge de travail pour la classification manuelle précédente, mais aussi une demande accrue de ressources de stockage et de calcul lors du post-traitement informatique.[14]

3. Machines à vecteurs de support (Support Vector Machines)

Une méthode populaire est la machine à vecteur de support (SVM), qui résout les tâches de classification et de régression en construisant un plan non linéaire séparant les solutions. En raison de la nature de l'espace de caractéristiques dans lequel les limites de la solution sont construites, la SVM offre une grande flexibilité dans la résolution des problèmes de classification de différents niveaux de complexité.[13]

- **Approche par réseaux neuronaux (NN) (Neural Network)**

Une autre catégorie de méthodes de détection du langage misogyne repose sur l'utilisation de réseaux neuronaux. Les réseaux neuronaux (RN) nous permettent de trouver des connexions et des modèles cachés dans les textes, mais ces connexions ne peuvent pas être représentées sous une forme explicite. Plusieurs raisons expliquent l'intérêt croissant des chercheurs pour les réseaux neuronaux. Tout d'abord, l'utilisation des réseaux neuronaux améliore considérablement la qualité de la résolution de certains problèmes standard de classification de textes et de séquences.

Problèmes de classification de textes et de séquences standard. Deuxièmement, l'utilisation de réseaux neuronaux réduit la complexité du travail direct sur le texte. Troisièmement, les réseaux neuronaux nous permettent de résoudre de nouveaux problèmes (par exemple, la

création de chatbots). En même temps, les réseaux neuronaux ne peuvent pas être considérés comme un mécanisme totalement indépendant pour résoudre les problèmes linguistiques.[13]

1. Réseaux neuronaux convolutifs (Convolutional neural networks (CNN))

Les algorithmes basés sur un réseau de neurones convolutifs (CNN), qui est une branche de l'apprentissage profond (Deep Learning), ont connu des progrès significatifs dans les analyses d'images et de vidéos ces dernières années. Le succès de ces nouveaux modèles a conduit à leur utilisation généralisée dans divers domaines, notamment la fouille de textes et les données de séries temporelles.[13]

2. Réseaux neuronaux récurrents (Recurrent neural networks (RNN))

Les réseaux neuronaux récurrents (RNN) constituent un autre type de réseaux neuronaux. Il est nécessaire de disposer de grands corpus pour étudier les modèles linguistiques. Est nécessaire d'avoir de grands corpus pour étudier les modèles de langage, donc plus le corpus d'entraînement est grand, plus le modèle "connaît" de paires de mots. L'utilisation de réseaux neuronaux pour développer des modèles linguistiques réduit la quantité de données stockées. En fonction du nombre de couches cachées et du nombre de neurones qu'elles contiennent, le réseau formé peut être stocké sous la forme d'un certain nombre de matrices denses de dimension relativement faible. Cependant, un tel modèle neuronal de langage ne permet pas de prendre en compte les longues connexions entre les mots. Ce problème est résolu par les RNN dans lesquels l'état interne de la couche cachée n'est pas seulement mis à jour après l'arrivée d'un nouveau mot à l'entrée, mais est également transféré à l'étape suivante. Ainsi, la couche cachée du réseau récurrent accepte deux types d'entrées : l'état de la couche cachée à l'étape précédente et le nouveau mot. Si un RNN traite une phrase, les états cachés lui permettent de mémoriser et de transmettre de longues connexions dans les phrases. Comme exemple de mise en œuvre réussie d'un RNN, on peut citer le travail où le RNN a été mis en œuvre en utilisant des caractères comme entrée au lieu de mots, ce qui a permis d'augmenter d'environ 8 % la précision moyenne de la classe. De même, les auteurs ont montré la réussite de l'implémentation du RNN dans un cas de classification de messages en langue grecque provenant de Facebook.[13]

3. Mémoire à long et court terme (Long-Short-term memory (LSTM))

Le dernier type de réseaux neuronaux à noter dans le cas de la classification de textes est la mémoire à long et court terme (LSTM), qui est un type spécifique d'architecture RNN, capable d'apprendre les dépendances à long terme. Ce type de réseaux neuronaux est adapté à la résolution d'un certain nombre de problèmes et est aujourd'hui largement utilisé. Les auteurs ont appliqué la LSTM pour prédire les polarités des tweets et ont obtenu une précision supérieure de 1 % par rapport au modèle RNN standard. Un LSTM bidirectionnel, composé de deux LSTM fonctionnant en parallèle, a obtenu de bons résultats dans la classification de textes.[13]

II.13.2 Identification automatique de la misogynie

La forme spécifique de discours de haine dans les réseaux sociaux étudiée dans cette thèse est la misogynie et ses diverses manifestations. La misogynie est un concept qui signifie la haine, l'hostilité, un préjugé enraciné contre les femmes. Elle peut se manifester de différentes manières, et l'une des manifestations les plus courantes est le sexisme, c'est-à-dire un ensemble de stéréotypes et de préjugés à l'égard des personnes en fonction de leur sexe. L'idéologie du sexisme divise les gens en hommes et en femmes, les oppose les uns aux autres et affirme directement ou indirectement la supériorité des hommes sur les femmes. Elle explique l'inégalité économique, sociale et politique entre eux par des différences de nature. La misogynie peut également s'exprimer sous la forme d'une objectivation sexuelle des femmes par rapport à une autre personne, exclusivement comme un instrument (objet) pour leur propre satisfaction sexuelle. Cette objectivation peut se manifester tant au niveau de la société qu'au niveau de la communication individuelle. La misogynie comprend également la violence à l'égard des femmes et l'humiliation des femmes, c'est-à-dire les conséquences directes de la haine. Bien qu'il s'agisse dans ce cas d'un espace virtuel - les plateformes Internet - plutôt que d'un contact physique direct, il convient de noter que les messages contenant des signes de misogynie sont insultants et peuvent nuire à leur destinataire. Il est donc important de protéger les utilisateurs des messages misogynes. Le développement d'approches permettant l'identification automatique de la misogynie est devenu une nécessité urgente dans les réseaux sociaux, en raison de la vitesse à laquelle ces messages sont envoyés et de l'impossibilité de contrôler et de supprimer les messages misogynes "à la main". Un certain nombre d'approches ont été publiées ces dernières années pour aider à détecter les messages misogynes. Nous

décrivons plus en détail les approches les plus importantes pour l'identification automatique de la misogynie dans ce qui suit.[13]

1. L'approche de Waseem et Hovy

Waseem et Hovy ont créé un ensemble de données composé de tweets sexistes. Dans le cadre de leurs travaux, ils ont proposé une définition des conditions dans lesquelles les messages sont considérés comme misogynes. Nous utilisons cette classification des messages misogynes dans notre travail. Un message est qualifié de misogyne si :

1. une insulte sexiste a été utilisée.
2. une minorité a été attaquée.
3. l'objectif était de chercher à faire taire une minorité.
4. une minorité a été critiquée sans argument fondé.
5. un discours de haine ou un crime violent a été promu mais n'a pas été directement utilisé.
6. une minorité a été critiquée et un argument de paille a été utilisé.
7. la vérité a été déformée de manière flagrante ou l'objectif était de chercher à déformer les opinions sur une minorité avec des affirmations non fondées.
8. des hashtags problématiques ont été utilisés.
9. les stéréotypes d'une minorité sont négatifs.
10. le sexisme a été défendu.
11. un pseudonyme offensant a été utilisé.

Les auteurs ont extrait un certain nombre de caractéristiques à l'aide des métadonnées des messages. Ils ont mis en évidence le sexe des utilisateurs en examinant les noms dans les profils et ont constaté qu'environ la moitié des messages étaient rédigés par des hommes, 2,26 % étaient rédigés par des femmes et 47,64 % ne comportaient aucune indication. En utilisant les fuseaux horaires indiqués dans les tweets, les auteurs ont créé une caractéristique de distribution géographique et ont également noté la longueur des tweets. Un modèle a été créé sur la base d'une régression logistique qui a donné les meilleurs résultats dans les cas où les caractéristiques de genre et de localisation géographique étaient prises en compte. Il est intéressant de noter que

dans les cas où le sexe, la localisation géographique et la longueur du tweet étaient pris en compte en même temps, les résultats n'étaient pas aussi bons. Ce travail nous permet de mieux comprendre le processus de construction des ensembles de données sur les discours de haine et les caractéristiques de l'extraction des métadonnées.[13]

2. L'approche de Park et Fung

Park et Fung présentent trois modèles différents basés sur des réseaux neuronaux qui fonctionnent bien pour identifier les discours haineux. Le premier de ces modèles est le modèle CharCNN, qui est basé sur CNN, un réseau convolutionnel au niveau des caractères, et le deuxième modèle est le modèle WordCNN, qui est un réseau convolutionnel dans lequel une phrase segmentée est identifiée. WordCNN, qui est un réseau convolutionnel dans lequel une phrase segmentée est convertie en mots pour inclure un mot 2VC à 300 dimensions, entraîné sur 100 milliards de mots tirés de Google News. Le dernier modèle, HybridCNN, est une combinaison de CharCNN et de WordCNN. HybridCNN est une combinaison de CharCNN et de WordCNN avec deux entrées : Les caractères et les mots. L'idée de créer ce modèle est née de l'observation que les tweets offensants contiennent souvent des mots mal orthographiés, intentionnellement ou accidentellement. Les trois modèles comportent trois couches. Les auteurs ont comparé leurs résultats à ceux de modèles basés sur la régression logistique, le texte rapide et les machines à vecteurs de support à l'aide de la métrique F1. L'ensemble de données pour les expériences a été créé à partir de tweets sexistes, de messages racistes et de tweets non racistes et non sexistes. Les auteurs ont présenté deux types de classification : la classification en une étape et la classification en deux étapes. La classification en une étape consiste en une catégorisation de trois catégories différentes : "raciste", "sexiste" et "aucun". Dans ce cas, les meilleurs résultats ont été obtenus par le modèle HybridCNN avec un score F1 de 0,827, tandis que le meilleur résultat obtenu par le modèle classique (LogReg) était de 0,814. La classification en deux étapes consistait à diviser la première étape en deux catégories, "offensant" et "aucun", puis à diviser la catégorie "offensant" en catégories "sexiste" et "raciste" dans la deuxième étape. Dans ce type de catégorisation, le meilleur résultat a été obtenu par le modèle LogReg avec un score F1 de 0,826 score F1, tandis que le modèle HybridCNN a obtenu un score F1 de 0,807. Dans ce cas, les auteurs ont combiné HybridCNN et LogReg et ont amélioré le score F1 à 0,818. Ce travail a montré que la combinaison d'un modèle basé sur un réseau neuronal et d'une approche classique d'apprentissage automatique peut donner de très bons résultats pour la classification en deux étapes, et ce résultat est encourageant.[13]

3. L'approche de Hardaker et McGlashan

Autre travail intéressant lié aux caractéristiques linguistiques, les auteurs avaient deux objectifs principaux dans leur recherche : Premièrement, étudier le langage entourant les agressions sexuelles sur Twitter et, deuxièmement, trouver des communautés en réponse à ce langage sexuellement agressif. Ils ont constaté que le langage abusif visait principalement à dégrader les femmes, en utilisant des mots sexuellement agressifs (abus, viol et menaces étaient très fréquents) et était souvent utilisé avec des mots tels que get, receive, receive, et des adjectifs tels qu'horrible, lâche, honteux, méprisable, ignoble, photogénique abominable, répugnant et grotesque.

Il est également intéressant de noter que dans ce type de message injurieux, le sujet grammatical (celui qui commet l'action injurieuse) était absent, comme d'habitude, et que le message était centré sur la personne vilipendée.

Les auteurs ont étudié un groupe d'utilisateurs ayant reçu de nombreux messages dans lesquels l'image du « vrai homme » était incompatible avec un comportement abusif et menaçant.

Ils ont sélectionné trois types d'utilisateurs différents (à haut risque, à faible risque et sans risque) et ont essayé de comparer leurs messages pour déterminer si les utilisateurs à haut risque utilisaient un langage plus abusif. Les utilisateurs à haut risque incluaient dans leur profil Twitter des preuves de ce qui suit : Intention de faire craindre un préjudice (sexuel), harcèlement et comportement potentiellement illégal. Les profils des utilisateurs à faible risque contenaient peu de matériel offensant, d'insultes et de sarcasmes. Les utilisateurs à faible risque n'avaient aucune preuve de ces incidents. La sélection d'un groupe d'utilisateurs s'est avérée très complexe et il est nécessaire de tenir compte du comportement dynamique des comptes Twitter et du fait que le type d'utilisateur peut être différent d'un compte à l'autre.

Twitter et au fait que le type de compte peut passer d'un risque faible à un risque élevé au fil du temps.[13]

4. L'approche de Clark et Greif

L'article présente l'étude de l'importance de la variation linguistique fonctionnelle dans un corpus de tweets à caractère sexuel. Les auteurs ont analysé le rôle des caractéristiques lexicales et syntaxiques à l'aide de l'analyse multidimensionnelle, une méthode qui repose sur une analyse multifactorielle permettant de détecter les caractéristiques grammaticales et lexicales mesurées dans un corpus de textes. Les chercheurs ont sélectionné 81 caractéristiques apparaissant dans au moins 1 % des tweets et les ont soumises à une analyse des correspondances multiples

(ACM) en R à l'aide de Facto Miner, ce qui leur a permis d'obtenir une mesure positive et une mesure négative pour chaque caractéristique linguistique. Ils ont ainsi obtenu une mesure positive et une mesure négative pour chaque caractéristique linguistique, qui a mis en évidence les relations entre la fréquence d'utilisation de cette caractéristique dans les messages sexistes dans trois dimensions différentes, à savoir l'interactivité, l'hostilité et l'attitude.

La première dimension, l'interactivité, montre à quel point le message est interactif ou informatif. Le message visait-il à engager le lecteur dans une discussion ou à l'informer sur des faits ? Les attributs les mieux notés sont le « point d'interrogation » (lorsqu'il y a un « ? » dans le message), « est interrogatif » (lorsque le message commence par le mot « est »), « est » (lorsque le message commence par le mot « est ») et « cas de fraude » (lorsqu'un cas de fraude est utilisé dans le message), les scores les plus élevés étant « cas de fraude » (lorsqu'un cas de fraude est utilisé dans le message) et les scores les plus bas étant « cas de fraude » (lorsqu'un cas de fraude est utilisé dans le message). Les scores les plus bas étaient « cas accusatif », « adverbe de lieu » et « être comme verbe principal », et montraient que les tweets les plus engageants contenaient beaucoup de questions, tandis que les messages informatifs avaient tendance à présenter quelques faits.

La deuxième dimension, « Adversaire », montre l'attitude de l'utilisateur à l'égard de ses lecteurs : Sont-ils d'accord avec eux ou non ? Les traits les plus fréquents dans le cas où l'utilisateur est les traits les plus fréquents dans le cas où l'utilisateur est adversaire des autres utilisateurs sont les « pronoms interrogatifs », les « points d'interrogation », les « points d'interrogation », les « pronoms », et dans les cas où l'utilisateur est d'accord avec ses lecteurs, les « pronoms », les « pronoms ». Il convient de noter que l'expression « d'accord » signifie que les messages étaient toujours sexistes et que l'utilisateur communiquait avec ses amis qui partageaient son point de vue.

La dernière dimension était « attitudinale » et a été interprétée comme représentant le degré de jugement attitudinal manifesté par le tweet. Les attributs les plus courants étaient les « adjectifs structurels », l'« existentialisme » et l'« absence de prépositions », qui témoignaient de l'opinion de l'utilisateur, tandis que les caractéristiques les moins bien notées étaient les « adjectifs auxiliaires », l'« aspect progressif » et les « balises de hachage ». Cet article nous permet de conclure que les caractéristiques linguistiques les plus « communes » dans le langage

abusif sont les « points d'interrogation » et le « point d'interrogation ». Les auteurs de sexto ont tendance à écrire des tweets plus personnalisés, et nous devrions accorder plus d'attention à ce concept dans les futures recherches sur les sextos.[13]

- **Résumé des approches**

Dans cette section, nous résumons les travaux étudiés consacrés à l'identification automatique de la misogynie en termes de caractéristiques utilisées, d'algorithme de classification et de résultats principaux. Le résumé est présenté dans un tableau :[13]

Les auteurs de l'approche Approche	Caractéristiques	Classification Algorithme	Principaux résultats
Waseem Z, Hovy D.	Le sexe des utilisateurs, la répartition géographique a été créée et a également noté la longueur des tweets.	Régression logistique	Les meilleurs résultats ont été obtenus dans les cas où les caractéristiques du sexe et de la situation géographique ont été prises en compte. Géographique étaient prises en compte
FasoliF., CarnaghiA., Paladino M.	Indice d'offensivité et indice d'acceptabilité	Analyses factorielles	Les insultes ont plus d'acceptabilité sociale dans un contexte de relations affectives que dans un contexte de relations sociales. Les insultes émanant de femmes sont plus acceptables que celles émanant d'hommes.
Hardaker C., McGlashan, M.	Des listes de mots-clés	Linguistique de corpus	Le langage offensant avait surtout pour but d'insulter les

	positifs et négatifs		femmes, en utilisant des mots agressifs sur le plan sexuel (abus, viol, menaces) et en les associant souvent à des mots tels qu'obtenir, recevoir, recevoir, et à des adjectifs tels que terrible, lâche, honteux, méprisable, graphique, haineux et horrible. Dans ce type de messages abusifs, l'acteur grammatical (celui qui commet l'action abusive) est absent, comme d'habitude, et le message est centré sur la personne vilipendée
Clarke I., Grieve J.	81 caractéristiques linguistiques et grammaticales	Analyse multidimensionnelle, analyse des correspondances multiples	La caractéristique linguistique la plus « populaire » dans le langage offensant sont les « points d'interrogation » et les « questions ». Les auteurs de messages sexistes ont tendance à écrire des tweets plus personnels.

Tableau 2:Résumé des approches d'identification automatique de la misogynie.

II.14 Conclusion

Dans ce chapitre, nous avons abordé les concepts de l'intelligence artificielle (IA) et des exemples de ses domaines de travail, mentionné l'intelligence artificielle générale (AGI) avec ses types, ses avantages et ses inconvénients, et défini l'apprentissage automatique avec une explication de ses fonctions et de ses types, en plus des cinq principales applications de l'apprentissage automatique et des dix principales tendances de 2022 et de son rôle dans la compréhension et la lutte contre la misogynie sur les réseaux sociaux, nous avons également

mentionné les techniques d'apprentissage automatique, à savoir le traitement du langage naturel et l'apprentissage en profondeur, avec des explications détaillées.

Enfin, nous avons présenté la détection de la misogynie par la détection automatique du langage misogyne selon deux approches : l'approche classique de l'apprentissage automatique et l'approche du réseau neuronal (NN). Nous avons également mentionné l'identification automatique de la misogynie et, enfin, nous avons brièvement décrit quelques articles qui ont étudié et analysé le sujet de la misogynie dans les sites de réseaux sociaux auparavant.

Chapitre



Conception, Implémentation et résultats

III.1 Introduction

Dans ce chapitre, nous concentrons sur la création du programme que nous avons proposé après avoir étudié la détection de la misogynie dans les réseaux sociaux en général dans le chapitre précédent. Tout d'abord, nous avons commencé par utiliser (Google Colab) pour écrire notre programme plus facilement sans avoir besoin d'un ordinateur puissant pour télécharger les bibliothèques, et nous avons expliqué les étapes nécessaires à l'installation de tous les composants de base. En outre, nous avons expliqué le processus de catégorisation des commentaires en commentaires appréciés et non appréciés à l'aide de l'ensemble de données Let-Mi. Enfin, pour évaluer la performance de notre approche, des résultats de simulation ont été collectés et discutés.

III.2 Motivation

Récemment, nous avons remarqué que de plus en plus d'utilisateurs participent activement aux réseaux sociaux sous la forme de messages WhatsApp, de messages Facebook, de courts clips YouTube, de commentaires Twitter, etc. sur divers sujets, y compris les commentaires misogynes. Les gens partagent leurs opinions, ce qui génère une énorme quantité de données sur le web. En présentant diverses définitions des formes de misogynie, en les analysant et en soulignant la motivation derrière la détection du contenu misogyne, nous discutons brièvement des travaux récents dans ce domaine en termes de méthodologies, de méthodes, de performances, de différentes normes, etc. Nous soulignons également d'autres tendances futures qui motivent les chercheurs à détecter les contenus misogynes.

III.3 Dataset

Dans ce chapitre, nous présentons le Levant Arabic Twitter Misogyny Dataset (Let-Mi) comme le premier ensemble de données standardisé pour le corpus de misogynie en arabe. Il peut être décrit comme un ensemble de données politiques car il s'agit de réponses à des tweets extraits des lignes de temps de journalistes/reporters populaires lors des manifestations du 17 octobre au Liban. Il a été collecté à l'aide de l'API de Twitter 3.[15]

Data set	Total des Commentaires	Misogynie	Non-Misogynie
Let-Mi	5240	2654	2586

Tableau 3: Caractéristiques de l'ensemble de données utilisé.

Voici quelques exemples de commentaires Twitter dans notre ensemble de données (Let-Mi) :

Text	Misogynies
اي نحن ما منقلبها صرماية باجرنا...مبروك.ع راسكم	misogyny
نوال شو قصتك كل يوم بتاكلي قتله شو صاير معك شي اليوم طمنينا عنك	none
انشاللهيقبر قلبك عن قريب...يافهيمة عصرك	misogyny
يسلم لسانك من مصر يا ديما نحن معك	none
فعلاً الثورة أنثى، برافو ديما	none
إنك بكل تواضع سفيهة منحطة عديمة المبدأ، ويجوز فيك القول بل يجب إنك...وضيعة، لست إلا مأمورة غبية تحتاجين ال https://t.co/B68FXaFNex	misogyny
شتقنالک جيسیکا مش عم نشوفك كثير ♥	none

Tableau 4: Tweetez des exemples d'étiquettes d'annotation.

III.4 Les outils et bibliothèques utilisés

1. Les outils

- **Google Colab**

Google Colab, abréviation de Google Colaboratory, est une plate-forme cloud gratuite fournie par Google qui permet aux utilisateurs d'écrire et d'exécuter du code Python dans un environnement de bloc-notes Jupyter. Il est largement utilisé pour les projets d'analyse de données, d'apprentissage automatique et d'apprentissage profond en raison de son accessibilité et de l'intégration de ressources informatiques puissantes telles que les GPU et les TPU.[16]

- **Python**

Python est un ensemble d'instructions que nous donnons sous la forme d'un programme à notre ordinateur pour qu'il effectue une tâche spécifique. Il s'agit d'un langage de programmation interprété, orienté objet et de haut niveau. En raison de sa syntaxe conviviale,

il est devenu un choix évident pour les débutants qui veulent se lancer dans la programmation. L'objectif principal de sa création est de le rendre plus facile à lire et à comprendre pour les développeurs, tout en réduisant le nombre de lignes de code.[17]

2. Bibliothèques

Nous avons utilisé un groupe de bibliothèques Python afin d'obtenir un bon résultat pour notre logiciel, à savoir :

- **Tensorflow**

TensorFlow est une plate-forme et un framework open source pour l'apprentissage automatique, qui comprend des bibliothèques et des outils basés sur Python et Java, conçus dans le but de former des modèles d'apprentissage automatique et d'apprentissage profond sur les données.[18]

La bibliothèque TensorFlow contient un ensemble de fonctions que nous avons utilisées pour coder notre modèle :

1. Tokenization

Un module dans « tensorflow.keras.preprocessing.text » que nous avons utilisé pour convertir le texte en une séquence de nombres. Ce module identifie les mots et les convertit en nombres que les modèles peuvent traiter.

2. Pad-sequences

Est une fonction de « tensorflow.keras.preprocessing.sequence » que nous avons utilisée pour fixer la longueur de la séquence à un niveau égal en ajoutant du padding.

3. Sequential

Il s'agit d'une classe de « tensorflow.keras.models » que nous avons utilisée pour créer un modèle hiérarchique, où les couches sont ajoutées de manière séquentielle.

4. Dense

Une couche entièrement connectée trouvée dans « tensorflow.keras.layers » qui est utilisée dans les réseaux neuronaux.

5. Embedding

Nous avons utilisé cette couche pour créer des représentations vectorielles des mots. Convertit les mots en vecteurs à haute dimension.

6. LSTM

Il s'agit d'un type de réseau neuronal récurrent (RNN) que nous avons utilisé pour traiter les données séquentielles et la mémoire à long terme.

7. Bidirectional

Il s'agit d'une couche que nous avons utilisée pour rendre un réseau neuronal itératif (tel que LSTM) bidirectionnel, permettant au modèle d'apprendre à partir du contexte précédant et suivant dans les données séquentielles.

- **NumPy**

NumPy est une bibliothèque Python utilisée pour travailler avec des tableaux. Il dispose également de fonctions permettant de travailler dans le domaine de l'algèbre linéaire, de la transformée de Fourier et des matrices. NumPy a été créé en 2005 par Travis Oliphant. C'est un projet open source et vous pouvez l'utiliser librement. NumPy signifie Python numérique.[19]

- **NLTK**

NLTK est une plateforme de premier plan pour la création de programmes Python permettant de travailler avec des données sur le langage humain. Elle fournit des interfaces faciles à utiliser pour plus de 50 collections de ressources lexicales telles que WordNet, un ensemble de bibliothèques de traitement de texte pour la classification, la tokenisation, la segmentation, l'étiquetage, l'analyse syntaxique et l'inférence sémantique, des wrappers pour des bibliothèques NLP de qualité industrielle, et un forum de discussion actif.[20]

La bibliothèque NLTK contient un ensemble de ressources que nous avons téléchargées et utilisées pour coder notre modèle :

1. Stopwords

Une liste de stop words (mots courants qui n'ont pas de sens fort, comme « in » et « on » en arabe).

2. Wordnet

Une base de données de vocabulaire anglais que nous avons utilisée pour la lemmatisation. Il est défini comme suit :

→ **WordNet Lemmatizer**

Il s'agit d'un module issu de la bibliothèque NLTK (Natural Language Toolkit), que nous avons utilisé pour effectuer la lemmatisation, un processus qui convertit un mot en sa forme de base ou sa racine en se basant sur le dictionnaire WordNet.

3. Omw-1.4

Une base de données similaire à WordNet mais prenant en charge plusieurs langues avec Open Multilingual Wordnet version 1.4.

- **Pickle**

Le module pickle implémente des protocoles binaires de sérialisation et dé-sérialisation d'objets Python. La sérialisation est le procédé par lequel une hiérarchie d'objets Python est convertie en flux d'octets. La désérialisation est l'opération inverse, par laquelle un flux d'octets (à partir d'un binary file ou bytes-like object) est converti en hiérarchie d'objets.[21]

- **Pyarabic**

Une bibliothèque spécifique au langage arabe pour Python, qui fournit des fonctions de base pour manipuler les lettres et le texte arabes, comme la détection des lettres arabes, des groupes de lettres arabes et de leurs caractéristiques, la suppression des diacritiques, etc.[22]

- **Matplotlib**

Matplotlib est une bibliothèque complète permettant de créer des visualisations statiques, animées et interactives en Python. Matplotlib rend les choses faciles faciles et les choses difficiles possibles.[23]

Nous avons utilisé un ensemble de modules pour obtenir un bon résultat pour notre programme, à savoir :

- **CSV**

Un module Python qui fournit des fonctions pour lire et écrire des fichiers CSV (comma-separated values).

- **Re**

Un module Python qui fournit des fonctions pour gérer les expressions régulières. Nous l'avons utilisé parce qu'il est utile pour des tâches avancées de traitement de texte telles que la recherche, le mélange et la division.

- **String**

Un module de Python contenant un ensemble de constantes et de fonctions permettant de manipuler du texte.

III.5 Configuration expérimentale

Nous avons utilisé :

- Un ordinateur portable DELL.
- Mémoire vive installée : 4,00 Go.
- Type de système : Système d'exploitation 64 bits, processeur x64.
- Système d'exploitation : Windows 10 Professionnel 64 bits (version 22H2).
- Processeur : Intel ® Core™ i3-4010U CPU@ 1.70GHz.
- Carte graphique : Intel ® HD Graphics family.

III.6 Implémentation

III.6.1 Prétraitement

Le prétraitement est une étape importante pour préparer les données avant qu'elles ne soient introduites dans le modèle d'apprentissage profond. Le code effectue plusieurs étapes de prétraitement, passons-les en revue :

- Supprimer les liens du texte à l'aide d'une expression régulière.
- Suppression des noms d'utilisateur (ceux commençant par @).
- Suppression des chiffres.
- Supprimez les symboles ou les mots contenant des caractères inattendus à l'aide d'une expression régulière.
- Supprimez les émoticônes à l'aide d'une expression régulière.
- Supprime les signes diacritiques d'un texte arabe à l'aide de la bibliothèque Pyarabic.

III.6.3 Diviser les données

La division des données en ensembles de formation, de validation et de test est une étape importante du processus de construction et d'évaluation des modèles d'apprentissage automatique. Voici comment cette segmentation est mise en œuvre dans notre logiciel :

1. Ensemble d'entraînement (Training Set)

Il est utilisé pour former le modèle et représente la majeure partie des données (environ 80 % des données totales). Le modèle apprend des modèles et des paramètres à partir de ces données.

2. Ensemble de validation (Validation Set)

Il est utilisé pour évaluer les performances du modèle au cours du processus de formation et représente une petite partie des données (environ 10 % des données totales). Il permet d'ajuster les paramètres du modèle (tels que les taux d'apprentissage, le nombre de couches, etc.) et d'éviter le problème de l'adaptation excessive.

3. Ensemble de test (Test Set)

Cet ensemble est utilisé pour évaluer les performances du modèle après l'apprentissage. Il représente la partie restante des données (environ 10 % des données totales) et donne une indication de la capacité du modèle à traiter des données qu'il n'a jamais vues auparavant.

III.7 Résultats expérimentaux et discussion

Notre modèle a été entraîné et testé sur le jeu de données Let-Mi, en utilisant Google Colab avec Python 3 comme environnement, ainsi que Tensorflow, Keras et la bibliothèque NLTK. Le corpus a été divisé en 3 groupes : 80 % pour l'entraînement, 10 % pour les tests et 10 % pour la validation. Nous avons pris 10 % comme ensemble de validation. Lors de la division de l'ensemble de données, nous avons pris en compte la distribution égale des « commentaires misogynes » et des « commentaires non misogynes ». La figure 10 montre les statistiques relatives à la division de l'ensemble de données.

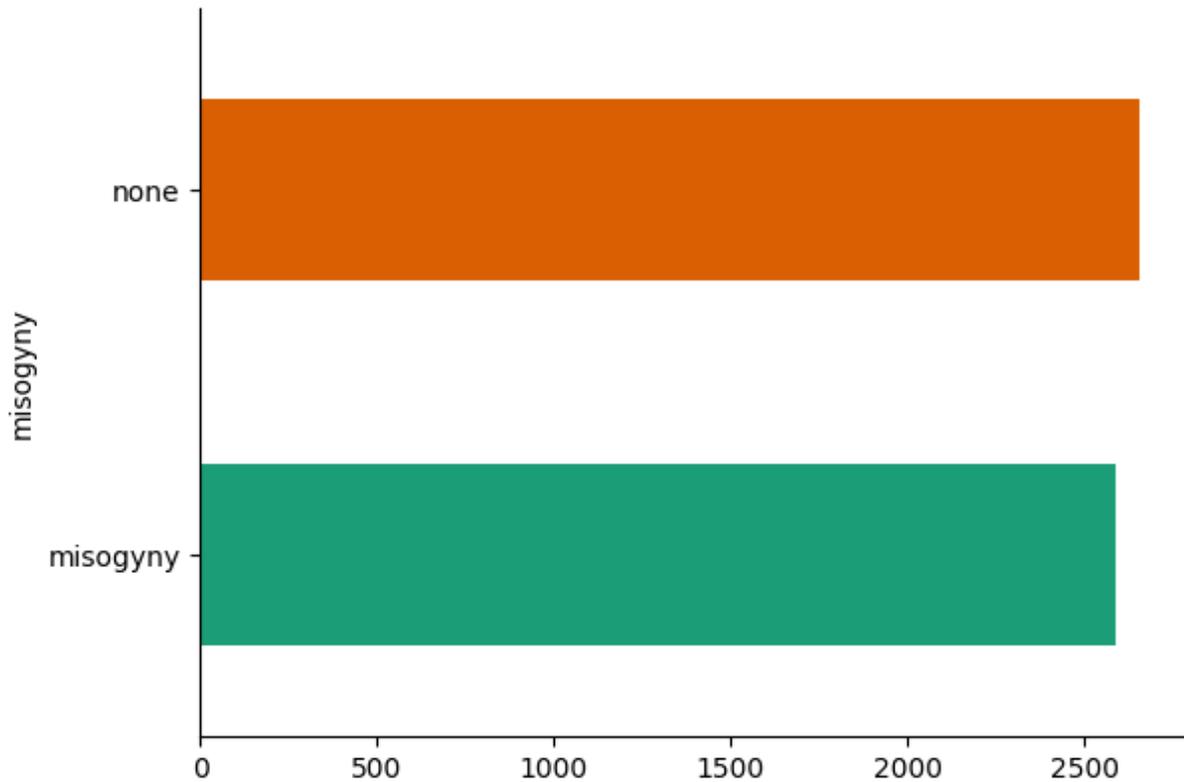


Figure 10: les statistiques relatives à la division de l'ensemble de données.

III.8 Résultats de l'expérimentation

Les résultats ont été évalués sur la base de l'exactitude (Accuracy), de la précision (Precision), du rappel (Recall) et du F-score :

- l'exactitude est utilisée pour estimer le pourcentage de prédictions exactes. et se calcule comme suit :

$$\text{Accuracy} = \frac{T_p + T_n}{2aT_p + F_p + F_n + T_n}$$

- La précision est le pourcentage de cas correctement prédits sur l'ensemble des cas prédits et est calculée à l'aide de la formule suivante :

$$\text{Precision} = \frac{T_p}{T_p + F_p}$$

- Le rappel est le rapport entre le nombre de cas correctement prédits le nombre de cas correctement prédits et le nombre total de cas pertinents. et se calcule comme suit :

$$\text{Recall} = \frac{T_p}{T_p + F_n}$$

- La métrique F1 est la moyenne harmonique du rappel et de la précision donnée par :

$$\text{F1score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Ici T_p , T_n , F_p et F_n représentent le nombre de réponses correctes positives, fausses positives, fausses négatives, fausses négatives et vraies négatives respectivement.[24]

-Le tableau 5 résume les résultats de la formation de l'ensemble des modèles de base :

Modèle	Accuracy	Precision	Recall	F1 score
LSTM	86.28	85.37	88.77	87.03
CNN	89.29	88.15	91.67	89.88

Tableau 5: Mesures d'évaluation des performances sur l'ensemble de données Let-Mi.

Le modèle CNN (89,29 %) a atteint la précision la plus élevée d'environ 3,01 points de pourcentage par rapport au modèle LSTM (86,28 %). En termes de précision positive, le modèle CNN (88,15 %) est supérieur d'environ 2,78 points de pourcentage au modèle LSTM (85,37 %). En extraction, le modèle CNN (91,67 %) est supérieur d'environ 2,90 points de pourcentage au modèle LSTM (88,77 %). En F1, le modèle CNN (89,88%) présente un score effectif supérieur d'environ 2,85 points de pourcentage au modèle LSTM (87,03%), tandis qu'en F2, le modèle CNN (89,88%) présente un score effectif supérieur d'environ 2,85 points de pourcentage au modèle LSTM (87,03%). Enfin, nous concluons que le modèle CNN présente de meilleures performances globales que le modèle LSTM dans ce cas de classification de texte, et le modèle CNN semble être en mesure d'extraire plus d'informations de modèle des données et d'améliorer les performances dans les classifications binaires par rapport au modèle LSTM. Cependant, nous devons également garder à l'esprit que les performances peuvent être affectées par de nombreux facteurs, y compris le type de données, la taille des données, les modifications du modèle et l'environnement d'entraînement et d'évaluation.

Le tableau 6 présente les résultats significatifs pour la misogynie obtenus par les chercheurs précédents sur l'ensemble de données Let-Mi. En examinant les résultats, nous remarquons

qu'ils ont utilisé des modèles différents des nôtres, et nous nous concentrons sur la comparaison du modèle que nous avons en commun, à savoir le modèle LSTM.

Modèle	Accuracy	Precision	Recall	F1 score
Majority class	0.52	0.26	0.50	0.34
BOW + TF-IDF	0.84	0.84	0.84	0.84
Frenda et al. (2018) model	0.85	0.86	0.85	0.85
LSTM	0.82	0.82	0.82	0.82
BERT	0.88	0.88	0.89	0.88

Tableau 6: Résultats de la tâche d'identification de la misogynie.[15]

En comparant les résultats du modèle LSTM dans les deux tableaux, nous constatons qu'il existe une différence significative entre les résultats, qui peut être due aux différents paramètres du modèle.

L'exactitude (Accuracy) dans le tableau 5 est plus élevée, ce qui indique une meilleure performance globale du second modèle, la précision (Precision) dans le tableau 5 est plus élevée, ce qui signifie que le modèle réduit mieux les faux positifs, le rappel (Recall) dans le tableau 5 est plus élevé, ce qui signifie que le modèle reconnaît mieux les positifs, et enfin, le score F1 dans le tableau 5 est beaucoup plus élevé, ce qui indique une meilleure performance dans l'équilibre entre la précision et le rappel. La comparaison nous permet de conclure que le modèle LSTM du tableau 5 présente une amélioration significative de toutes les mesures par rapport au modèle du tableau 6. Cela indique que le modèle du tableau 5 est plus performant que le modèle du tableau 6. Cela indique que le modèle du tableau 5 a été formé et optimisé à l'aide de techniques et de données différentes, ce qui se traduit par de meilleures performances.

III.9 Conclusion

Dans ce chapitre, nous avons expliqué notre logiciel étape par étape afin de comprendre ses algorithmes visant à identifier la misogynie sur les réseaux sociaux. Nous avons couvert les différentes étapes de l'installation du système, en particulier l'initialisation dans le cadre Google colab, nous avons donné un aperçu de l'ensemble de données Let-Mi que nous avons utilisé avec des exemples, et enfin nous avons évalué notre proposition et discuté des résultats obtenus.

Conclusion générale

Les réseaux sociaux jouent un rôle important dans notre vie quotidienne, en particulier après les récents développements dans le monde, bien que la plupart des gens aient au moins un compte sur les réseaux sociaux. Avec la diffusion de ces derniers, des habitudes négatives se sont également répandues, qui sont devenues une préoccupation majeure pour les utilisateurs de ces moyens, notamment la misogynie, qui est devenue un phénomène croissant qui représente un défi majeur pour les sociétés modernes.

Cette recherche met en lumière les formes, les causes et les effets de la misogynie sur les individus et la société dans son ensemble. Les réseaux sociaux jouent un double rôle : ils offrent une plateforme d'expression et de participation, mais facilitent en même temps la propagation de comportements négatifs et incendiaires à l'égard des femmes. Il est clair qu'il est urgent d'élaborer des stratégies globales pour lutter contre ce phénomène, notamment en renforçant les réglementations et les lois sur le harcèlement en ligne, en renforçant le rôle des plateformes de réseaux sociaux dans la surveillance et la suppression des contenus, ainsi qu'en sensibilisant et en éduquant les utilisateurs à l'importance du respect d'autrui et de la non-discrimination. Nous avons proposé une stratégie visant à créer un programme qui élimine ce phénomène en utilisant l'ensemble de données Let-Mi collectées par les chercheurs. Le logiciel télécharge et traite les données, puis divise les données en formation, vérification et test, puis convertit les textes en matrices numériques, construit et entraîne un modèle d'apprentissage profond pour classer les commentaires, et enfin sauvegarde le modèle entraîné et l'utilise pour analyser les commentaires via une interface de ligne de commande.

Enfin, en termes de perspectives d'avenir, on s'attend à ce que les plateformes de réseaux sociaux continuent à développer leurs politiques et leurs outils pour lutter contre les discours de haine et le harcèlement. Cela inclut l'introduction de nouvelles fonctionnalités pour protéger les utilisateurs, comme celle que nous avons proposée pour développer notre logiciel afin de détecter et de supprimer automatiquement les commentaires misogynes pour créer un environnement numérique plus sûr et plus inclusif.

Les références

- [1] “C’est quoi les réseaux sociaux ?” Accès le : May 10, 2024. [En ligne]. Available: <https://www.solocal.com/ressources/articles/definition-reseaux-sociaux>
- [2] وسائل التواصل الاجتماعي وأثرها على المجتمع, “هديل الخطيب,” Accès le : May 10, 2024. [En ligne]. Available: https://mawdoo3.com/وسائل_التواصل_الاجتماعي_وأثرها_على_المجتمع
- [3] O.-C. Silvi Saxena, MBA, MSW, LSW, CCTP, “Comportement antisocial : définition, exemples et impacts.” Accès le : May 21, 2024. [En ligne]. Available: <https://www.choosingtherapy.com/antisocial-behavior/>
- [4] N. Y. Times, T. Guardian, and Y. News, “Anti-Social Behaviour ASB and Misinformation in online communities,” pp. 1–31, 2021.
- [5] “Misogyny.” Accès le : May 01, 2024. [En ligne]. Available: <https://en.m.wikipedia.org/wiki/Misogyny>
- [6] “misogyny.” Accès le : May 01, 2024. [En ligne]. Available: <https://www.britannica.com/topic/misogyny#ref352689>
- [7] “مصطلحات-طبية/امراض-نفسية/كاره-النساء | Misogyn.” Accès le : May 01, 2024. [En ligne]. Available: <https://altibbi.com/مصطلحات-طبية/امراض-نفسية/كاره-النساء>
- [8] “Special issue on online misogyny.” Accès le : May 02, 2024. [En ligne]. Available: <https://www.tandfonline.com/doi/full/10.1080/14680777.2018.1447345>
- [9] Coursera Staff, “What Is Artificial Intelligence? Definition, Uses, and Types.” Accès le : Apr. 01, 2024. [En ligne]. Available: <https://www.coursera.org/articles/what-is-artificial-intelligence>
- [10] Vijay Kanade, “Qu’est-ce que l’apprentissage automatique ? Définition, types, applications et tendances.” Accès le : May 22, 2024. [En ligne]. Available: <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/>
- [11] Antoine Crochet-Damais, “Natural language processing (NLP) : définition et techniques.” Accès le : May 21, 2024. [En ligne]. Available: <https://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence->

artificielle/1501887-natural-language-processing-nlp/

- [12] Tim Jones, “Qu’est-ce que l’apprentissage profond ? Définition, techniques et cas d’utilisation.” Accès le : May 25, 2024. [En ligne]. Available: <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-deep-learning/>
- [13] E. Shushkevich, “Misogyny Detection in Social Media on the Twitter Platform Misogyny Detection in Social Media on the Twitter platform Elena Shushkevich Technological University Dublin Master,” 2020, doi: 10.21427/d1jc-vj32.
- [14] L. Li and Y. Huang, “Naive Bayes classification algorithm based on small sample set.” Accès le : May 10, 2024. [En ligne]. Available: <https://ieeexplore.ieee.org/abstract/document/6045027>
- [15] H. Mulki and B. Ghanem, “Let-Mi: An Arabic Levantine Twitter Dataset for Misogynistic Language,” *WANLP 2021 - 6th Arab. Nat. Lang. Process. Work. Proc. Work.*, no. October, pp. 154–163, 2021.
- [16] “Qu’est-ce que Google Colab ?” Accès le : Jun. 01, 2024. [En ligne]. Available: <https://www.ituonline.com/tech-definitions/what-is-google-colab/>
- [17] “What is Python? it’s Uses and Applications.” Accès le : Jun. 01, 2024. [En ligne]. Available: <https://www.geeksforgeeks.org/what-is-python/>
- [18] C. BasuMallick, “What Is TensorFlow? Meaning, Working, and Importance.” Accès le : Jun. 01, 2024. [En ligne]. Available: <https://www.spiceworks.com/tech/devops/articles/what-is-tensorflow/>
- [19] “Présentation de NumPy.” Accès le : Jun. 01, 2024. [En ligne]. Available: https://www.w3schools.com/python/numpy/numpy_intro.asp
- [20] “NLTK.” Accès le : Jun. 01, 2024. [En ligne]. Available: <https://www.nltk.org/>
- [21] “pickle— Sérialisation d’objets Python.” Accès le : Jun. 01, 2024. [En ligne]. Available: <https://docs.python.org/3/library/pickle.html>
- [22] “PyArabic 0.6.15.” Accès le : Jun. 01, 2024. [En ligne]. Available: <https://pypi.org/project/PyArabic/>
- [23] “Matplotlib : visualisation avec Python.” Accès le : Jun. 01, 2024. [En ligne].

Available: <https://matplotlib.org/>

- [24] C. Toumi and A. Bouramoul, “Ensemble learning-based model for fake news detection,” *4th Int. Conf. Pattern Anal. Intell. Syst. PAIS 2022 - Proc.*, pp. 1–8, 2022, doi: 10.1109/PAIS56586.2022.9946895.