République Algérienne Démocratique et Populaire

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

Université de Mohamed El Bachir El Ibrahimi de Bordj Bou Arreridj

Faculté des Mathématiques et d'Informatique

Département d'informatique

MEMOIRE

Présenté en vue de l'obtention du diplôme

Master en informatique

Spécialité : Technologie de l'Information et de la Communication

THEME

L'analyse des sentiments des tweets à propos de " ChatGpt "

Présenté par :

- DJAAFAR Yasmine
- BELHOUCHET Lilia

Soutenu publiquement le : 20 /06/2023

Devant le jury composé de:

- **Président :** Dr.Badaoui Atika
- **Examinateur :** Dr.Boutouhami Sara
- **Encadreur :** Mr. BENDIAF Messaoud

2022/2023

Remerciement

A l'issue de ce travail nous remercions Allah qui nous donne la patience et le courage durant ces longues années d'études.

Nous tenons à saisir cette occasion et adresser nos profonds remerciements et nos profondes reconnaissances à:

Notre encadreur Mr : Bendiaf Messaoud pour ses précieux conseils et son aide durant toute la période de travail.

Nos remerciements vont également à nos parents.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils sont eus porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.

Nous tenons à remercier toute personne qui a participé de près ou de loin à l'exécution de ce travail.

Dédicace

Je dédie sincèrement et affectueusement mon humble travail :

A mes chères parents ma mère et mon père pour leur patience, leur amour, leur soutient et leurs encouragements.

À mes amies et mes camarades sans oublier tous les professeurs.

Djâafar Yasmine.

Dédicace

Je dédie ce modeste travail :

À Ma très chère mère, pour ses sacrifices, son amour et son aide et à qui sa prière était le secret de mon succès.

À Mon très cher père, pour ses conseils et son soutien matériel et qui s'est toujours sacrifié pour me voir réussir.

À mon cher frère : Islam.

À mes chères sœurs : Sarah, Aya et Salsabile.

Je souhaite personnellement remercier mon ami proche Walid qui a toujours été là pour moi. Son soutien inconditionnel et ses encouragements ont été d'une grande aide.

Et à tous ceux qui me connaissent de près ou de loin.

Belhouchet Lilia,

Résumé

Dans les domaines politiques, de production et de services, l'analyse des textes est devenue un élément crucial. Avec l'omniprésence des réseaux sociaux, les internautes partagent leurs opinions sur divers sujets à travers des textes, ce qui rend la compréhension du contenu de ces derniers essentielle.

Il est crucial pour un gestionnaire efficace de considérer les opinions des citoyens et pour ce faire, l'analyse des sentiments revêt une grande importance afin de répondre adéquatement aux besoins de la population.

Nous allons utiliser trois algorithmes dans notre étude pour analyser et classifier un groupe de publications provenant des réseaux sociaux.

Ces derniers seront classés en trois catégories, c'est-à-dire la classe positive, négative et neutre.

A notre connaissance, il s'agit d'un des premiers travaux qui explore et compare plusieurs algorithmes de classification de commentaires sur Twitter.

Mots clés: fouille d'opinions, analyse des sentiments, web social, Twitter,Lexique de sentiments.

Abstract

In political, production, and service sectors, text analysis has become a crucial element. With the prevalence of social media, users share their opinions on various topics through texts, making understanding the content of these texts essential.

It's critical for effective management to consider citizens' opinions, and sentiment analysis plays a crucial role in adequately responding to the population's needs.

We will use three algorithms in our study to analyze and classify a group of publications from social media.

These will be classified into three classes: positive, negative, and neutral.

To our knowledge, this is one of the first studies that explores and compares several algorithms for classifying comments on Twitter.

Keywords : opinion mining, Sentiment Analysis, social web, Twitter, Lexicon of Sentiment.

الملخص

في القطاعات السياسية والإنتاجية والخدماتية، أصبح تحليل النص عنصرًا هاما.

مع انتشار وسائل التواصل الاجتماعي ، يتبادل المستخدمون آرائهم حول مواضيع مختلفة من خلال النصوص ، مما يجعل فهم محتوى هذه النصوص أمرًا ضروريًا.

من الأهمية بمكان للإدارة الفعالة أن تأخذ في الاعتبار آراء المواطنين ، ويلعب تحليل المشاعر دورًا أساسي في الاستجابة بشكل مناسب لاحتياجات السكان.

سنستخدم ثلاث خوارزميات في دراستنا لتحليل وتصنيف مجموعة من المنشورات من وسائل التواصل الاجتماعي.

سيتم تصنيفها إلى ثلاث فئات: إيجابية وسلبية ومحايدة.

على حد علمنا ، هذه إحدى الدراسات الأولى التي تستكشف وتقارن عدة خوارزميات لتصنيف التعليقات على تويتر

الكلمات المفتاحية : التنقيب في الآراء، تحليل المشاعر، قاموس المشاعر, الشبكات الاجتماعية، تويتر

Table des matières

Liste des figures

Liste des tableaux

Introduction générale

Aujourd'hui, l'information et son analyse sont au cœur de notre point de vue.

Grâce au développement du Web 2.0, cette information est de plus en plus disponible sous forme numérique, permettant aux gens de communiquer, partager et exprimer leurs opinions en ligne, en particulier dans les groupes de discussion, les blogs, les forums et les sites de critiques de produits.

L'impact des avis en ligne est considérable, comme en témoigne des sondages montrant que la plupart des utilisateurs font des recherches d'avis avant d'acheter un produit ou un service. Les entreprises prennent en compte ces retours pour la prise de décision.

La détection d'opinions est ainsi devenue une composante importante dans nombreux domaines, tels que le marketing politique, les études psychologiques, la santé ou encore le tourisme.

Depuis l'annonce de la mise en service de Chat GPT en novembre 2022, et déjà upgradé dans une version améliorée, GPT-4, l'intelligence artificielle fait grand bruit. Tout le monde en parle, l'essai et s'en amuse. La nouvelle version dépasse le million d'utilisateurs en moins d'une semaine, elle leur a facilité leurs épanouissement dans leurs vie. Certaines personnes craignent que ces technologies ne soient utilisées pour remplacer des emplois humains et perturber d'importants secteurs de l'économie, et pour faciliter le harcèlement en ligne ou la violation de la vie privée.

L'objectif de ce travail est d'étudier les opinions publiques en anglais sur le ChatGpt, et il se concentre sur les sentiments exprimés dans les commentaires Twitter.

Ce mémoire est divisé en quatre chapitres,

Le premier chapitre Nous avons commencé notre mémoire avec une introduction générale puis nous avons présenté la nouvelle technologie "chat GPT", ses avantages et inconvénients, ainsi que la concurrence entre Open AI et Google.

Le deuxième chapitre offre un aperçu des médias sociaux, des réseaux sociaux, des blogs, du micro blogs et de Twitter, avec différentes informations sur chacun d'eux.

Le troisième chapitre traite de l'analyse de sentiment, y compris ses types, son fonctionnement et son approche.

Le quatrième chapitre décrit les étapes suivies dans notre travail.

Conclusion générale et perspectives.

Chapitre I:

Etat de l'Art

Chapitre Ⅰ: Etat de lart

I.1 Définition

ChatGpt est un modèle de langage développé par Open AI, qui utilise un artifice intelligent disponible pour générer du texte à la demande.

Le nom « ChatGpt » est la combinaison des termes « chat » et « GPT », qui est un signe de « conversation » et de « modèle de langage prédictif », faisant référence à la capacité de ChatGpt en tant que simulateur de conversations humaines d'utilisateurs réalistes.

I.2 Les avantages de ChatGpt

Le chats GPT offrent plusieurs avantages dans leur utilisation :

- ✓ Capacité de générer du texte de manière autonome : les chats GPT peuvent générer du texte de manière autonome, sans intervention humaine. Cela permet de créer des textes rapidement et efficacement.
- ✓ Compréhension et production de langage naturel : les chats GPT ont la capacité de comprendre et de produire un langage naturel à un niveau de complexité et de variété inédits.
- ✓ Apprentissage de manière autonome : les chats GPT sont capables d'apprendre de manière autonome à partir des données d'entraînement qui leur sont fournies, ce qui leur permet d'améliorer sans cesse leur performance.
- ✓ Adaptation à différents domaines : les chats GPT peuvent être entraînés pour répondre à des besoins spécifiques dans différents domaines, comme la finance, la médecine, l'éducation, ou la recherche scientifique.
- ✓ Réduction des coûts : l'utilisation de chats GPT peut réduire les coûts de production de textes, de traductions, de résumés, et de synthèses, en automatisant ces tâches.

I.3 Les inconvénients de ChatGpt

Bien que les chats GPT présentent de nombreux avantages, ils présentent également des inconvénients et des risques, tels que :

- Risque de biais : les chats GPT peuvent être biaisés en fonction des données sur lesquelles ils ont été entraînés, ce qui peut influencer les réponses et les recommandations qu'ils fournissent.
- Manque de transparence et d'explicabilité : les chats GPT sont souvent considérés comme des boîtes noires, ce qui rend leur fonctionnement interne difficile à comprendre ou à expliquer.
- Risque de propagation de fausses informations : les chats GPT peuvent produire du texte de manière autonome, mais cela peut aussi les conduire à générer des fausses informations et des désinformations.
- Utilisation abusive : les chats GPT peuvent également être mal utilisés pour propager des discours de haine, des messages toxiques, ou pour tromper les utilisateurs.
- Le plus grand inconvénient est le risque de chômage qui pourrait être créé par cette technologie.Plusieurs métiers pourraient être menacés par cette intelligence artificielle capable de travailler plus rapidement et efficacement, sans coûts salariaux.**[1]**

I.4 Open AI vs Google

L'IA est en train de transformer notre monde et de reconfigurer la manière dont nous vivons, travaillons et interagissons. Google et Open AI figurent parmi les entreprises les plus importantes dans ce domaine, chacune ayant apporté des contributions uniques à l'IA. Cependant, lorsque Google a récemment lancé son propre assistant virtuel sur le marché, cela a créé une concurrence directe entre ces deux géants technologiques.

I.4.1 Open AI

La fonctionnalité de GPT-4 permet de créer et de déboguer des codes sophistiqués, tout en étant également capable de supporter des expressions mathématiques complexes grâce à la prise en charge du balisage Latex. L'intelligence artificielle intégrée dans Bing utilise une base de connaissances pour fournir toutes les réponses à des requêtes. Le GPT-4 accepte à la fois du texte en entrée.

I.4.2 Bard de Google

La mise à jour d'avril a doté Bard de fonctions telles que l'aide à la génération, au débogage et à l'explication de codes, ainsi que la possibilité d'écrire des fonctions pour Google Sheets. Bard est également capable de trouver des réponses sur internet et de fournir des liens directs vers des sites web. Cependant, l'ajout d'images n'est pas encore possible sur Bard.

La précision des réponses offertes par Bard est limitée, Google a déclaré que « sa capacité à conserver le contexte est volontairement limitée pour le moment ». En outre, Bard propose plusieurs options de réponses pour chaque requête, mais son utilisation des citations est encore limitée.

Quant à la disponibilité, Bard est accessible au public via un site web indépendant créé par Google, mais il n'est disponible qu'aux États-Unis et au Royaume-Uni et uniquement en anglais

FIGURE 1: OPEN AI VS GOOGLE BARD

I.5 Conclusion

Le Chat GPT est une technologie fascinante qui exploite la puissance de l'intelligence artificielle pour simuler des conversations humaines convaincantes et générer du texte à la demande. Bien qu'il soit encore relativement nouveau, il montre un grand potentiel dans les domaines de la communication, de l'assistance client, de la recherche et du marketing en ligne. Cependant, il soulève certaines questions sur la manipulation de l'information et la confidentialité des utilisateurs.

La technologie continue d'évoluer et ChatGpt est destiné à devenir un outil essentiel pour améliorer la communication et les interactions dans notre monde de plus en plus numérique.

Chapitre Ⅱ :

Les réseaux sociaux

Chapitre Ⅱ : Les réseaux sociaux

II.1 Introduction

Les réseaux sociaux ont exercé une influence majeure sur notre vie, transformant la façon dont nous communiquons, interagissons et consommons de l'information. Ils ont créé de nouvelles opportunités de connexion avec des amis, des connaissances et des communautés partageant les mêmes intérêts, tout en permettant un partage instantané de contenu à travers le monde. Cependant, cette omniprésence des réseaux sociaux a également suscité des préoccupations concernant la vie privée, la cyberdépendance et l'impact sur la santé mentale. Il est essentiel de prendre du recul, de développer une utilisation consciente et équilibrée des réseaux sociaux afin de maximiser les avantages tout en minimisant les risques potentiels. .

II.2 Pourquoi utiliser les réseaux sociaux

Des milliards de personnes dans le monde utilisent les réseaux sociaux pour échanger de l'information et établir des liens. Sur le plan personnel, les réseaux sociaux nous permettent de communiquer avec nos amis et notre famille, d'apprendre de nouvelles choses, de développer nos intérêts et de nous divertir. Sur le plan professionnel, nous pouvons utiliser les réseaux sociaux pour élargir nos connaissances dans un domaine particulier et bâtir notre réseau professionnel en établissant des liens avec d'autres professionnels de notre industrie. Au niveau de l'entreprise, les réseaux sociaux nous permettent d'avoir une conversation avec notre public, d'obtenir des commentaires des clients et d'améliorer notre marque.

II.3 Les avantages des réseaux sociaux

Les réseaux sociaux offrent de nombreux avantages, notamment

- **Communication et connectivité :** Les réseaux sociaux permettent de communiquer facilement avec des amis, des collègues, des membres de la famille et des personnes du monde entier. Ils offrent également des moyens de se connecter avec des communautés de personnes partageant les mêmes intérêts ou préoccupations.
- **Marketing et promotion :** Les entreprises peuvent utiliser les réseaux sociaux pour promouvoir leurs produits et services, ainsi que pour établir leur présence en ligne et renforcer leur image de marque. Les plateformes sociales permettent aux entreprises de toucher un public plus large à moindre coût que les canaux de marketing traditionnels.
- **Partage d'information :** Les réseaux sociaux permettent aux utilisateurs de partager facilement du contenu tel que des articles, des vidéos, des photos et

des blogs. Cela offre un moyen efficace de diffuser de l'information et de sensibiliser à des sujets d'intérêt public.

- **Opportunités de réseautage :** Les réseaux sociaux offrent des opportunités de réseautage professionnel et personnel. Les professionnels peuvent utiliser les plateformes sociales pour se connecter avec des collègues et des clients potentiels, échanger des idées et trouver de nouvelles opportunités de carrière.
- **Divertissement et loisirs :** Les réseaux sociaux offrent des possibilités de divertissement et de loisirs. Les utilisateurs peuvent regarder des vidéos, écouter de la musique, jouer des jeux en ligne, suivre des événements en direct, etc.

II.4 Les inconvénients des réseaux sociaux

Tout comme pour beaucoup d'autres choses, les réseaux sociaux doivent être utilisés de manière responsable si l'on ne veut pas s'exposer aux risques éventuels que comporte leur utilisation. Les risques éventuels de l'utilisation de réseaux sociaux sont les suivants :

- Être confronté à des contenus indésirables tels que des messages haineux ou violents.
- La perte de confidentialité due à une perte de contrôle sur les données personnelles.
- La baisse de notre rendement ou efficacité dans nos tâches ou activités. **[3]**

II.5 Les réseaux sociaux mobile

Les réseaux sociaux mobiles font référence à l'utilisation des réseaux sociaux sur les appareils mobiles comme les téléphones intelligents et les tablettes .Les réseaux sociaux mobiles sont des applications utiles du marketing mobile car la création, l'échange et la diffusion de contenu généré par les utilisateurs peuvent aider les entreprises à mener des recherches marketing, à communiquer et à établir des relations.

Les réseaux sociaux mobiles diffèrent des autres parce qu'ils intègrent l'emplacement actuel de l'utilisateur (sensibilité à l'emplacement) ou le délai entre l'envoi et la réception des messages. Selon Andreas Kaplan, les applications de réseaux sociaux mobiles peuvent être différenciées entre quatre types:

- ***Space-timers* (location and time-sensitive):** Échange de messages pertinents principalement pour un endroit spécifique à un point spécifique dans le temps (ex . Facebook Places , WhatsApp, Telegram, Foursquare)
- ***Space-locators* (only location sensitive**): Échange de messages pertinents pour un emplacement spécifique, qui est étiqueté à un certain endroit et lu plus tard par d'autres (p. ex., Yelp, Qype , Tumblr, Fishbrain)
- ***Quick-timers* (only time sensitive**): Transfert d'applications mobiles traditionnelles des médias sociaux pour accroître l'immédiateté (p.ex., affichage sur Twitter ou mises à jour sur Facebook)
- ***Slow-timers* (neither location nor time sensitive)*:*** Transfert d'applications de médias sociaux traditionnels vers des appareils mobiles. **[4]**

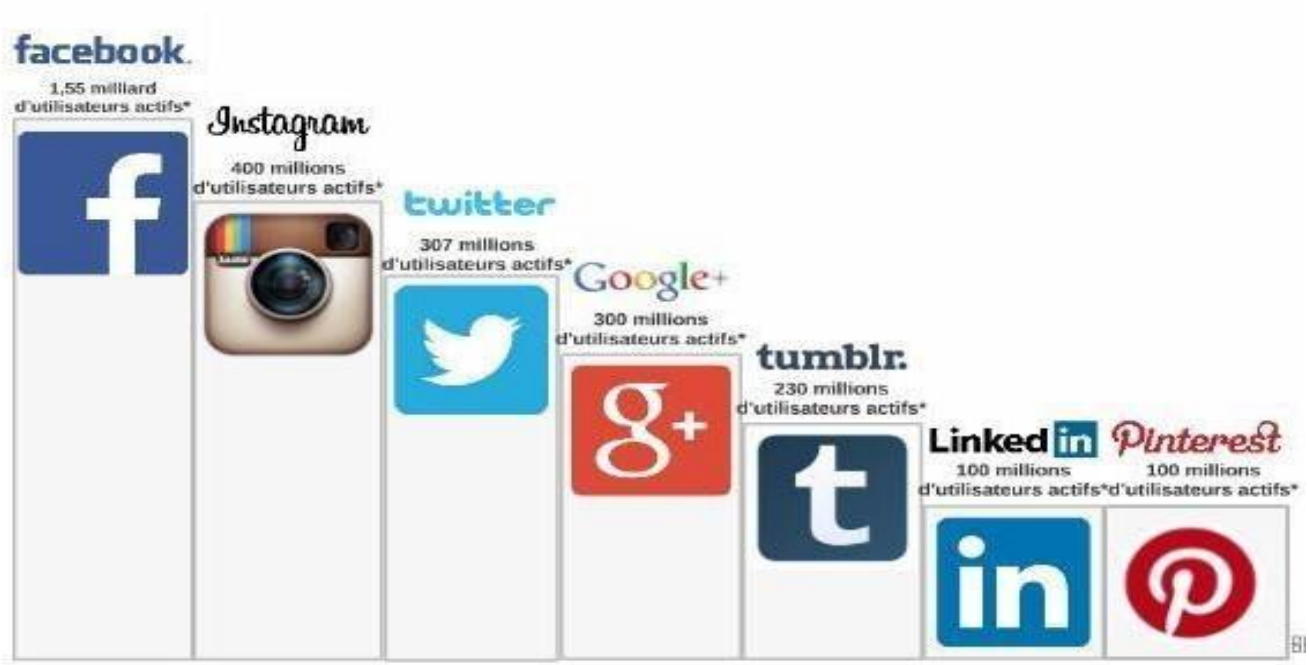


FIGURE 2 : LES RESEAUX SOCIAUX EN CHIFFRE

II.6 Exemples de réseaux sociaux

- **YouTube** est une plateforme sociale spécialisée dans les vidéos.
- **Facebook** est un réseau social qui compte 2,2 milliards d'utilisateurs.
- **Twitter** est une plateforme de microblogging.
- **LinkedIn** est un réseau social destiné aux professionnels.
- **Instagram** est un réseau social centré sur l'image.
- **Pinterest** est un réseau social où l'on partage des tableaux et des citations.
- **TikTok** est une plateforme pour les courtes vidéos musicales.

- **Snapchat** est un réseau social éphémère.
- **Dailymotion** est un concurrent direct de YouTube.
- **Reddit** est un réseau social basé sur le système d'upvote.
- **Twitch** est un réseau social destiné aux joueurs.
- **WhatsApp** est un réseau social pour les conversations.

II.7 Blog

Un blog est une page web tenue par un individu, appelé blogueur, qui partage ses idées et impressions sur divers sujets en ligne. Le contenu est souvent accompagné de liens externes, de photos, de dessins et/ou de sons qu'il souhaite partager. Les visiteurs ont également la possibilité de commenter ou de compléter les informations fournies.

II.8 Twitter

Twitter est une plateforme de réseau social gratuite où les utilisateurs publient des messages courts appels "tweets". Ces tweets peuvent contenir du texte, des vidéos, des photos ou des liens. Pour accéder à Twitter, une connexion Internet ou un smartphone est nécessaire pour utiliser l'application ou le site Web Twitter.com.

Il s'agit d'un service de microblogging, une combinaison de blogging et de messagerie instantanée, permettant aux utilisateurs enregistrés de publier, de partager, d'aimer et de répondre à des tweets avec de courts messages.

Les utilisateurs non enregistrés ne peuvent que lire les tweets. **[5]**

FIGURE 3 : LOGO TWITTER

II.9 Conclusion

Les réseaux sociaux ont révolutionné notre façon de communiquer et de partager des informations, créant une connectivité mondiale instantanée. Cependant, ils peuvent également présenter des défis en termes de protection de la vie privée, de propagation de la désinformation et de dépendance excessive. Il est essentiel de trouver un équilibre entre les avantages et les inconvénients des réseaux sociaux pour une utilisation responsable et épanouissante.

Chapitre Ⅲ:

Analyse des sentiments

Chapitre Ⅲ: Analyse des sentiments

III.1 Introduction

L’analyse de sentiment est une technique d'analyse des opinions, des sentiments et des émotions exprimés dans un texte, Cette technique utilise des algorithmes pour Identifier les attitudes positives, négatives ou neutres dans le contenu. L'analyse de sentiment est utilisée dans plusieurs domaines, y compris le marketing, la politique, la finance et la recherche. Elle permet de comprendre les opinions et les préférences des clients, des électeurs, des investisseurs et des chercheurs.[6]

Dans cette analyse, nous allons examiner les principaux aspects de l'analyse de sentiment, notamment les méthodes et les outils utilisés, les avantages et les limites de cette technique, ainsi que les applications pratiques dans différents domaines.

III.2 L'objectif l'analyse des sentiments

L'analyse de sentiment est importante pour plusieurs raisons. Tout d'abord, elle permet de comprendre les opinions et les attitudes des clients envers une marque, un produit ou un service. Cela peut aider les entreprises à améliorer leur offre et leur communication en fonction des besoins et des préférences des clients. L'analyse de sentiment peut également aider les entreprises à surveiller leur réputation en ligne en identifiant rapidement les commentaires négatifs et en prenant des mesures pour y remédier. Elle peut également aider à identifier les tendances et les opinions du public sur des sujets spécifiques, ce qui peut être utile pour les campagnes de marketing et les décisions commerciales.

Enfin, l'analyse de sentiment peut être utilisée dans divers domaines, tels que la politique, la finance et la santé, pour surveiller les opinions publiques et les tendances émergentes. Cela peut aider les décideurs à prendre des décisions éclairées en fonction des besoins et des préférences des personnes concernées.

III.3 Caractéristiques de l'analyse des sentiments

L'analyse des sentiments est une technique d'analyse de données qui permet d'extraire des informations sur les opinions, les attitudes et les émotions des personnes à partir de textes, de commentaires, de critiques ou de tout autre contenu textuel. Voici quelques-unes des caractéristiques de l'analyse des sentiments:

- ✓ **Analyse du contexte :** L'analyse des sentiments prend en compte le contexte dans lequel les textes ont été écrits. Par exemple, les mêmes mots peuvent avoir des significations différentes.

- ✓ **Traitement du langage naturel :** L'analyse des sentiments utilise des techniques de traitement du langage naturel pour comprendre les textes. Elle peut identifier les mots clés, les phrases importantes et les relations entre les différentes parties du texte.
- ✓ **Évaluation de la polarité :** L'analyse des sentiments permet d'évaluer la polarité des textes, c'est-à-dire si les opinions exprimées sont positives, négatives ou neutres.
- ✓ **Utilisation de données de sources multiples :** L'analyse des sentiments peut être effectuée à partir de différentes sources de données, telles que les réseaux sociaux, les forums en ligne, les critiques de produits, les enquêtes ou les commentaires de clients.

III.4 Les limites d'analyse de sentiment

L'analyse de sentiment a ses mites. Elle ne peut pas identifier les nuances de langage, comme l'ironie, le sarcasme ou Humour. Elle ne peut pas non plus comprendre le contexte dans lequel le texte a été écrit Par exemple, un commentaire négatif sur un produit pout être lié à une mauvaise expérience de livraison plutôt qu'à une mauvaise qualité du produit, Il est donc important de prendre en compte les limites de l'analyse de sentiment lors de son utilisation, l'est recommandé d'utiliser cette technique.

Conjointement avec d'autres méthodes d'analyse pour obtenir une vie d'ensemble plus complète du sujet étudié.

III.5 Niveaux d'analyses

Dans le cadre d'une étude d'analyse des sentiments, la première étape consiste à définir le texte qui sera analysé. Trois niveaux d'analyse sont généralement utilisés : le niveau du document, le niveau de la phrase et le niveau des aspects.

Ⅲ.5.1 Niveau du document

L'objectif est de déterminer la polarité globale du texte, en supposant que le texte ne contient qu'une seule opinion sur une seule entité (par exemple, un produit).

Ⅲ.5.2 Niveau de la phrase

Chaque phrase est analysée indépendamment pour déterminer sa polarité, en supposant que chaque phrase exprime une opinion unique sur une entité unique.

Ⅲ.5.3 Niveau des aspects

L’analyse est plus fine et repose sur l'hypothèse qu'une opinion se compose d'un sentiment et d'une cible d'opinion. **[7]**

III.6 Les types d'analyse de sentiment

Les types d'analyse des sentiments qui reposent sur l'intelligence artificielle (IA) et l'apprentissage automatique pour formuler des jugements et des prédictions.

- **L'analyse détaillée des sentiments** interprète la polarité de l'opinion publique, pouvant aller du simple sentiment binaire "j'aime/je n'aime pas" à une différenciation positive/négative ou à un système de notation complexe mesurant un fort accord ou un fort désaccord sur des questions comportementales.
- **L'analyse basée sur les aspects** permet de savoir ce que les clients pensent d'un élément spécifique d'un produit.
- **l'analyse des intentions** est utilisée pour déterminer l'intention derrière le message d'un client dans les services d'assistance à la clientèle.

Ⅲ.7 Les approches de l’analyse des sentiments

Il existe deux approches établies de l’analyse des sentiments :

- **L’approche fondée sur des règles**

Utilise un algorithme pour déterminer une description claire et précise de l’opinion en identifiant la subjectivité, la polarité…

Cette approche repose sur le traitement de base du langage naturel, incluant la recherche du radical, l’analyse syntaxique, la tokénisation, le marquage des parties du discours et l’analyse du langage.Elle fonctionne en comparant les mots du texte à des listes prédéfinies de mots positifs et négatifs pour calculer la polarité de l’opinion.

- **L’approche de l’analyse automatique des sentiments**

Consiste à extraire des informations exploitables à partir de textes en utilisant le machine learning plutôt que des règles prédéfinies.

Cette approche utilise des algorithmes de classification supervisée par machine learning pour comprendre la signification globale d’un message en fonction de plusieurs critères. Elle permet d’obtenir un niveau élevé de précision et d’exactitude tout en traitant rapidement les données. L’analyse des sentiments utilise différents types d’algorithmes de classification tels que la régression linéaire, les machines à vecteur de support, les bayésiens naïfs.

III.8 Conclusion

L'analyse de sentiment est une méthode importante pour comprendre les opinions et les sentiments des utilisateurs.

Les différents types d'analyse de sentiment ont leurs propres avantages et inconvénients, et peuvent être utilisés pour des tâches spécifiques.

En utilisant les bonnes techniques d’analyse de sentiment, les entreprises peuvent mieux comprendre les besoins et les préférences de leurs clients, améliorer leurs produits et services, et renforcer leur relation avec les clients.

Chapitre Ⅳ:

Implémentation et

Tests

Chapitre Ⅳ: Implémentation et tests

IV.1 Introduction

Dans différents domaines, la collecte et l'analyse des opinions des individus ont acquis une importance croissante en tant que sources d'informations précieuses.

Dans ce chapitre, nous décrirons en détail notre approche, ainsi que les différentes étapes que nous avons suivies pour effectuer une analyse des sentiments à partir de tweets. Nous avons traité un ensemble de tweets afin d'extraire la polarité des opinions exprimées, qu'elle soit négative, neutre ou positive.

Les données d'entrée que nous avons utilisées sont des tweets extraits d'un ensemble de données, afin d'entrainer et tester notre approche en temps réel.

IV.2 Approche proposée

Dans notre projet, nous nous concentrons sur l'analyse des sentiments dans le domaine de l'intelligence artificielle (IA).

Notre objectif est d'appliquer une analyse des commentaires sur ChatGpt. Pour atteindre cet objectif, nous avons identifié plusieurs étapes cruciales qui doivent être suivies afin d'obtenir des résultats optimaux. Ces étapes comprennent la collecte des données, le prétraitement des données, la classification des sentiments et l'évaluation des sentiments.

La figure ci-dessous illustre la méthodologie proposée ainsi que les étapes distinctes qui la composent. Tout d'abord, les données sont importées depuis Kaggle sous forme d'un ensemble de données (Dataset). Ensuite, les données textuelles sont nettoyées, ce qui implique de filtrer les données importées afin d'éliminer toute duplication, donnée manquante ou aberrante, tout en appliquant un prétraitement. Ensuite, les données sont traitées à l'aide de techniques de fouille de texte (Text Mining). Enfin, la précision, le rappel et le score F1 sont calculés en utilisant les trois classifieurs présents dans notre système : SVM, KNN et Naïve Bayes. **[8]**

IV.3 Corpus

Un corpus est un ensemble de messages recueillis manuellement ou automatiquement à partir d'une source spécifique telle qu'un journal, un réseau social ou un site de critiques, dans un domaine précis et dans un objectif déterminé, dans notre cas, l'analyse des opinions. Les corpus jouent un rôle essentiel dans les méthodes d'apprentissage automatique. En effet, ils fournissent la matière première nécessaire au classifieur, qui a besoin d'un grand nombre de messages annotés pour construire un modèle de classification et prédire la classe des nouveaux messages. Ainsi, plus la taille du corpus est grande, plus le modèle Construit à partir de l'apprentissage sur le corpus est de meilleures qualités. Dans la méthode d'apprentissage automatique, le corpus est généralement divisé en deux parties, comme nous le présenterons par la suite. **[9]**

IV.3.1 Corpus d'apprentissage

C'est avec ce corpus annoté que le classifieur fera l'apprentissage pour construire le modèle de classification. Il doit être de taille importante, de façon à bien modéliser le modèle de classification qu'il traite le maximum de cas possible. Ce corpus représente 80 % du corpus total.

IV.3.2 Corpus de test

Ce corpus sert à évaluer la qualité du modèle de classification construit dans la phase d'apprentissage avec des métriques dévaluations, ce corpus représente de 10% à 20% du corpus total.

IV.4 Collecte des tweets

Nous avons utilisés la Platform web **Kaggle** pour collecter l'ensemble des données.

Kaggle, une filiale de Google, est une communauté en ligne de scientifiques des données et d'ingénieurs en apprentissage automatique.

Kaggle permet aux utilisateurs de trouver les ensembles de données qu'ils souhaitent utiliser pour construire des modèles d'intelligence artificielle, de publier des ensembles de données, de collaborer avec d'autres scientifiques des données et ingénieurs en apprentissage automatique, et de participer à des compétitions visant à résoudre des défis en science des données.

Kaggle a été lancé en 2010 en proposant des compétitions en apprentissage automatique et en science des données, ainsi qu'une plateforme publique de données et de services cloud dédiée à l'éducation en science des données et en intelligence artificielle. **[10]**

IV.5 A propos du Dataset

Cette ensemble de données contient une collection de tweets avec le hashtag #ChatGpt. Les tweets ont été extraits de Twitter et couvrent une gamme de sujets liés au modèle linguistique ChatGpt. L'ensemble de données comprend les informations suivantes pour chaque tweet :

- Numéro de série
- Texte du tweet
- Nom d'utilisateur
- Sentiment (bad, good, neutral)

L'ensemble de données donne un aperçu de modèle linguistique ChatGpt et peut être utilisé pour diverses tâches de traitement automatique du langage naturel et d'apprentissage automatique, telles que l'analyse de sentiment, la modélisation de sujets, et plus encore. Il permet de comprendre la communauté, le niveau d'intérêt et l'utilisation de ChatGpt. **[11]**

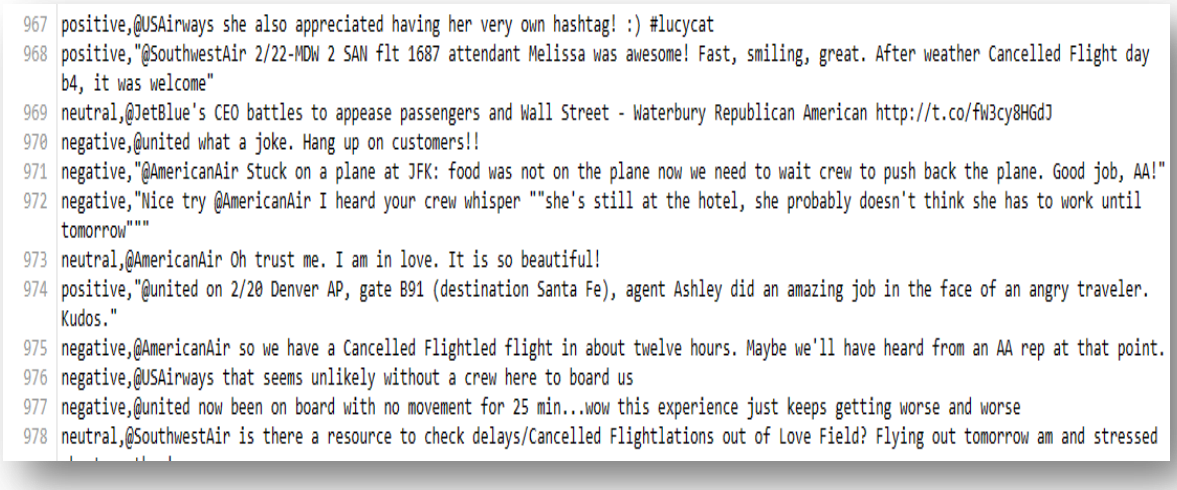


FIGURE 4 : DATASET

IV.6 Prétraitement

Cette étape essentielle comprend l'application de différentes techniques pour structurer et faciliter l'utilisation des messages.

Le prétraitement des données textuelles consiste en plusieurs étapes, qui sont les suivantes:

IV.6.1 Convertir les données textuelles en minuscules

Dans cette étape, nous avons appliqué une conversion en minuscules à tous les tweets. Pour ce faire, nous avons utilisé la fonction **Lower ()** en Python, qui permet de convertir tous les caractères majuscules en caractères minuscules.

Les figures ci-dessous représentent les résultats avant et après lapplication de cette opération.

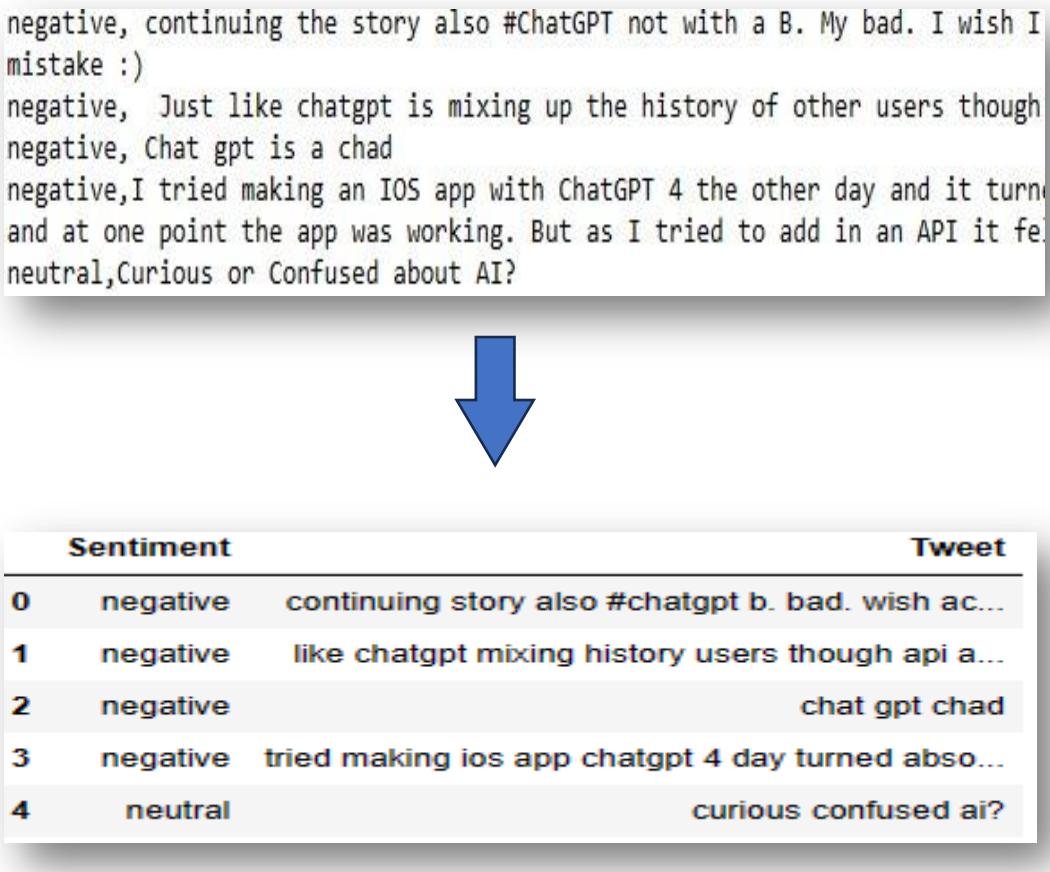


FIGURE 5 : LA CONVERSION EN MINUSCULE

IV.6.2 Nettoyage de données

Le nettoyage des données consiste à détecter et corriger les données altérées, inexactes ou non pertinentes. Cette étape essentielle du prétraitement des données permet d'améliorer la cohérence, la fiabilité et la valeur des données.

Ces données sont : les hashtags, les mots vides, les caractères spéciaux, les URL, les noms d'utilisateurs, etc. Pour faire cette étape on a besoin d'importer la bibliothèque **re** en python.

IV.6.2.1 Suppression des mots vides (stop words)

Dans ce cas, nous procédons à l'élimination des mots qui n'ont aucun impact sur l'opinion exprimée dans le message, et qui, de plus, augmentent de manière considérable et inutile le nombre de mots dans le vocabulaire.

Voici un exemple de ces mots :

(I, I'm, me, my, myself, we, ours, ourselves, you're, you've, you'll, you'd, your,

Yours, yourself, yourselves, he, Him, his, himself, she, her, hers, herself…etc.)

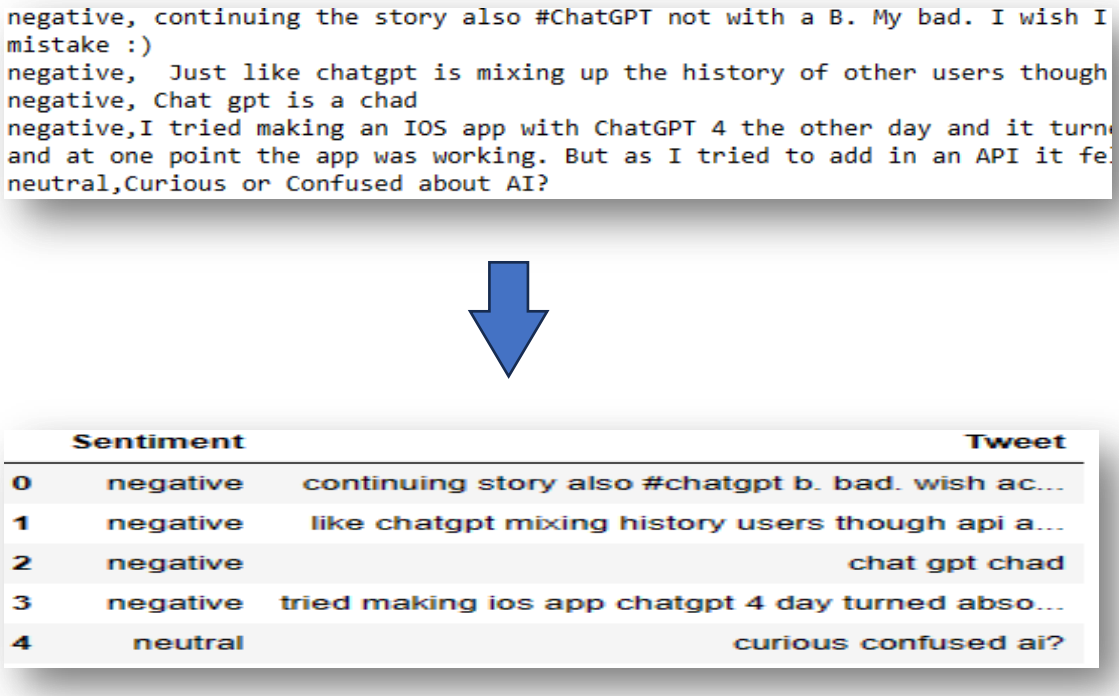


FIGURE 6 : LA SUPPRESSION DU MOTS VIDES

IV.6.2.2 Suppression des caractères spéciaux

Cette opération est pour supprimer les caractères spéciaux.

Tel que [“!”,””’,”%”,”&”,”amp, “{“,”|”,”}”,”~”,”–“,”$””””,”(“,”)”, “*”,”+”,”,”,”,”-“,”.”,

[…”~”,”–“,”@”,”#”,”/”,”:”,”;”,”<”,”=”,”>”,”?”,” [“,”\\”,”]”,”^”,”_”,…etc.]

IV.6.2.3 Suppression des caractères non alphabétiques

Cette opération est pour supprimer des caractères potentiellement nuisibles entres dans un champ de texte.

Parmi ces caractères, les non alphabétiques, exemples :

- Les liens (http?//[^\s<>”]+|www\.[^\s<>”]+)
- Les noms d’utilisateurs (@ [A-Za-z0-9] +…)
- Hashtags (\B#\w*[a-zA-Z]+\w*)
- Les emoticons (, , …)

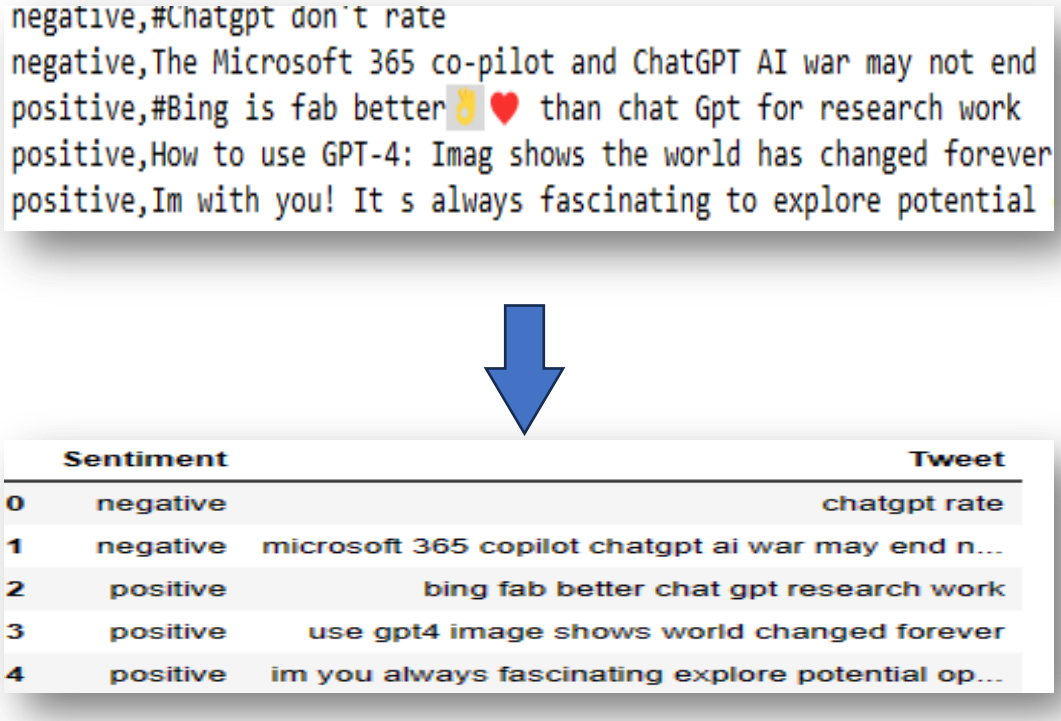


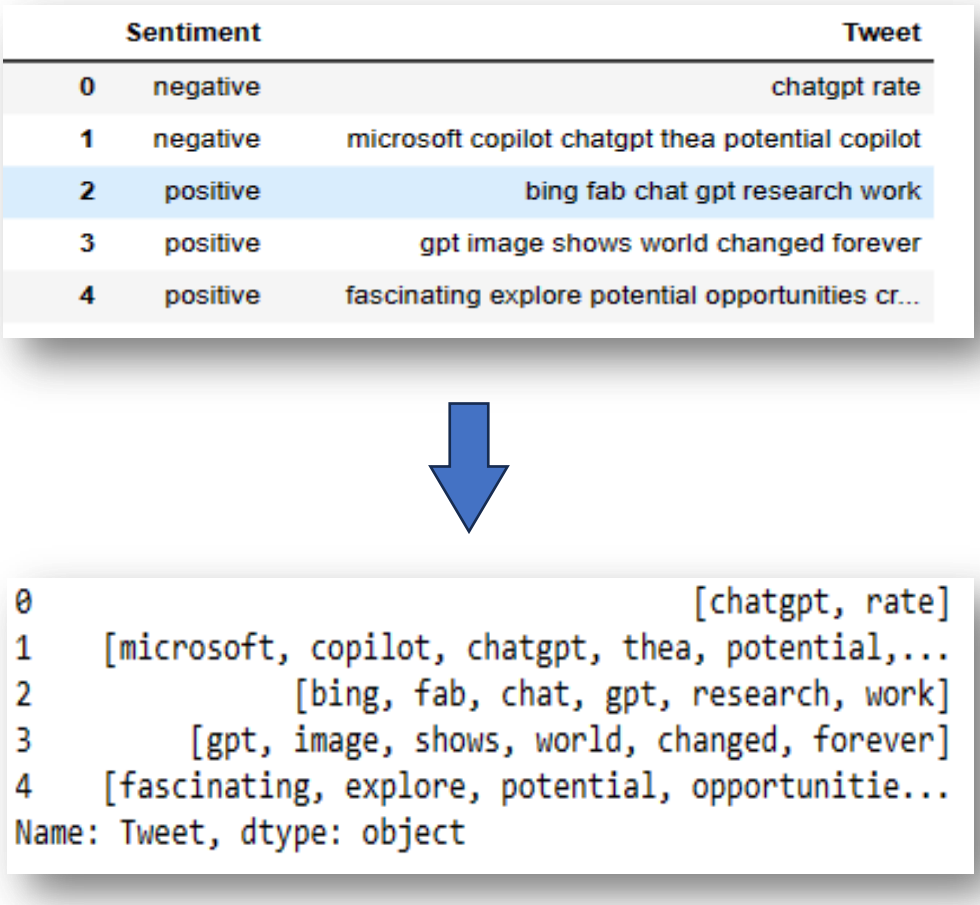
FIGURE 7 : LA SUPPRESSION DES CARACTERE SPECIAUX ET NON ALPHABETIQUES

IV.6.3 Tokenisation

La Tokenisation consiste à diviser un texte en entités plus petites appelées tokens. La définition d'un token peut varier en fonction du tokenizer utilisé. Un token peut représenter un mot, un caractère ou même un sous mot.

De plus, la ponctuation telle"!", ".", et ";" peut également être considérée comme des tokens.

La Tokenisation est une étape essentielle dans toutes les opérations de traitement du langage naturel (NLP). Étant donné les différentes structures linguistiques présentes, la Tokenisation varie d'une langue à l'autre. **[12]**



	Sentiment	Tweet
0	negative	chatgpt rate
1	negative	microsoft copilot chatgpt thea potential copilot
2	positive	bing fab chat gpt research work
3	positive	gpt image shows world changed forever
4	positive	fascinating explore potential opportunities cr...

```
0                                        [chatgpt, rate]
1    [microsoft, copilot, chatgpt, thea, potential,...
2                [bing, fab, chat, gpt, research, work]
3           [gpt, image, shows, world, changed, forever]
4    [fascinating, explore, potential, opportunitie...
Name: Tweet, dtype: object
```

FIGURE 8 : AVANT ET APRES LA TOKENISATION

IV.6.4 Lemmatisation

La lemmatisation, en linguistique, est le processus de regroupement des formes fléchies d'un mot afin de les analyser comme un seul élément, identifié par la forme canonique ou le lemme du mot, tel qu'il apparaît dans un dictionnaire.

En linguistique computationnelle, la lemmatisation est le processus algorithmique qui consiste à déterminer le lemme d'un mot en se basant sur son sens intentionnel. Contrairement au radical, la lemmatisation dépend de l'identification correcte de la catégorie grammaticale et du sens voulu d'un mot dans une phrase, ainsi que du contexte plus large entourant cette phrase, comme les phrases voisines ou même un document entier. Par conséquent, le développement d'algorithmes de lemmatisation efficaces est un domaine de recherche ouvert. **[13]**

	Tweet	Sentiment	no_sw	wo_stopfreq	wo_stopfreq_lem
0	chatgpt dont rate	negative	chatgpt rate	chatgpt rate	chatgpt rate
1	the microsoft copilot and chatgpt ai war may ...	negative	microsoft copilot chatgpt thea potential copilot	microsoft copilot chatgpt thea potential copilot	microsoft copilot chatgpt thea potential copilot
2	bing is fab better than chat gpt for research ...	positive	bing fab chat gpt research work	bing fab chat gpt research work	bing fab chat gpt research work
3	how to use gpt image shows the world has chang...	positive	gpt image shows world changed forever	gpt image shows world changed forever	gpt image shows world changed forever
4	im with you it s always fascinating to explore...	positive	fascinating explore potential opportunities cr...	fascinating explore potential opportunities cr...	fascinating explore potential opportunities cr...

FIGURE 9 : LA LEMMATISATION

IV.7 Techniques de classification

IV.7.1 Classifieur SVM

Sous la catégorie des Techniques de classification en apprentissage supervisé, l'algorithme SVM se distingue par sa capacité à prédire un hyperplan optimal dans un espace à n dimensions à partir d'un ensemble de données d'apprentissage.

cette technique divise l'ensemble de données d'entraînement en deux classes à l'aide de l'hyperplan.

Les SVM sont également capables de classifier les données dans des plans bidimensionnels ainsi que des hyperplans multidimensionnels, où les noyaux sont utilisés pour segmenter les données multidimensionnelles.**[14]**

IV.7.2 Classifieur KNN

L'algorithme kN-voisin (KNN) est un algorithme simple d'apprentissage automatique supervisé utilisé pour résoudre des problèmes de classification et de régression.

KNN fonctionne en recherchant les distances entre une donnée inconnue et toutes les données de la base d'apprentissage, sélectionnant ensuite les K exemples les plus proches de la requête, puis votant pour l'étiquette la plus fréquente (classification) ou en prenant la moyenne des étiquettes (régression).

Bien que facile à comprendre et à mettre en œuvre, KNN présente l'inconvénient majeur de ralentir considérablement à mesure que la taille des données utilisées augmente.

Pour choisir le bon K pour nos données, il est nécessaire d'essayer plusieurs K et de choisir celui qui fonctionne le mieux, que ce soit pour la classification ou pour la régression. **[15]**

IV.7.3 Classifieur Naïve Bayse

La méthode de classification naïve bayésienne est un algorithme de machine learning qui classe des observations grâce à des règles établies automatiquement. Pour cela, l'outil doit être entraîné sur des données, élaborant ainsi ses règles de classification. Cette méthode nécessite un jeu de données d'apprentissage pour donner des classes attendues en fonction des entrées.

L'algorithme suppose que les classes sont connues et fournies, caractérisant son caractère supervisé. Historiquement utilisé pour la classification de documents et les filtres anti-spam, cette méthode est aujourd'hui reconnue dans de nombreux domaines grâce à son apprentissage rapide et son exécution rapide par rapport à d'autres méthodes plus complexes.

La classification naïve bayésienne est basée sur le théorème de Bayes avec une hypothèse naïve d'indépendance entre toutes les paires de variables. Malgré cette hypothèse, elle donne des résultats remarquables dans de nombreux domaines de la vie courante. **[16]**

IV.8 Technologies utilisées

IV.8.1 Anaconda

Anaconda est une distribution libre et open source de logiciels pour la science des données, le traitement de données en gros volume, l'analyse et le calcul scientifique, ainsi que l'apprentissage automatique. Il a été créé par la société Anaconda, Inc. Anaconda contient une collection de plus de 1500 paquets open source et permet aux développeurs et aux scientifiques de travailler avec plusieurs langages de programmation, notamment Python, R et Julia.

Anaconda est un environnement complet de développement, qui comprend un gestionnaire de paquets, des éditeurs de code, des environnements virtuels, des outils de développement, des notebooks interactifs, des bibliothèques de visualisation de données, des bibliothèques pour l'apprentissage automatique et d'autres outils utiles pour le développement de projets de science des données. Anaconda rend l'installation de ces logiciels et bibliothèques plus faciles et plus rapides et offre un environnement intégré pour exécuter des scripts et expérimenter avec différents packages et librairies.

IV.8.2 Jupyter notebook

Jupyter Notebook (anciennement Python Notebook) est une application web open source pour la création, l'exécution et la partage de documents qui contiennent du code, des équations, des visualisations et du texte narratif. Il est souvent utilisé pour l'analyse de données, la simulation numérique, la modélisation mathématique et l'apprentissage automatique.

Dans un notebook Jupyter, les sections de code et de texte sont organisées en cellules, ce qui permet aux utilisateurs de traiter des données, d'expérimenter avec différents algorithmes et de générer des graphiques et des visualisations interactives.

Les résultats des cellules de code sont affichés directement dans le notebook. Le texte narratif permet d'expliquer les résultats, les hypothèses et les observations.

Jupyter Notebook prend en charge de nombreux langages de programmation, notamment Python, R, Julia, Matlab et d'autres langages populaires de science des données. Il permet également l'extension avec des bibliothèques tierces, telles que Pandas, Numpy, Scikit-learn, Matplotlib et Plotly, pour n'en nommer que quelques-unes.

Grâce à sa facilité d'utilisation, Jupyter Notebook est largement utilisé dans l'enseignement et la formation pour l'enseignement de la programmation et de la science des données. Il est également utilisé dans l'industrie pour l'exploration, l'analyse et la visualisation de données, ainsi que pour le prototypage d'applications et l'analyse de données en temps réel.

FIGURE 10 : LOGO D'ANACONDA ET JUPYTER NOTEBOOK

IV.9 Bibliothèque utilisés

NLTK	• est une suite de bibliothèques logicielles et de programmes conçue pour le traitement symbolique et statistique du langage anglais, le tout réalisé en utilisant le langage de programmation Python. • C'est une bibliothèque performante pour le traitement naturel du langage qui intègre plusieurs algorithmes clés tels que la segmentation de texte, l'étiquetage des parties de discours, la racinisation, l'analyse de sentiment… **[17]**
Pandas	• L'une des bibliothèques les plus prisées dans le domaine de l'analyse de données pour les Data Scientistes est conçue dans l'optique de simplifier et d'optimiser la manipulation des données, notamment celle sous forme de tables et de séries temporelles. • Elle offre une solution pratique pour l'analyse et la manipulation des données, **[18]**
Numpy	• Numpy, également appelé Numerical Python, est actuellement la bibliothèque de référence pour le calcul scientifique en Python. • sa capacité à effectuer des calculs numériques de base et la manipulation aisée des tableaux multidimensionnels. • Cette solution est souvent utilisée pour des tâches complexes telles que l'analyse numérique, l'algèbre linéaire et le calcul matriciel. **[19]**
Regular expression	• Re est une bibliothèque Python qui fournit des outils pour manipuler des chaînes de caractères en utilisant des expressions régulières, ce qui est utile dans la manipulation de données textuelles pour extraire des informations spécifiques.
Scipy	• La bibliothèque SciPy propose une multitude de sous-librairies dédiées à la résolution de divers problèmes mathématiques et numériques. • Parmi elles, on peut notamment citer : l'intégration numérique, la résolution d'équations et l'optimisation, l'algèbre linéaire, l'interpolation, la classification, les statistiques et enfin le traitement d'images et de signaux. **[20]**

Matplotlib	• L'objectif de Matplotlib est de permettre aux utilisateurs de créer des visualisations de données de qualité professionnelle en utilisant Python de manière simple, efficace et flexible. Cela inclut la création de graphiques, de diagrammes, de diagrammes de dispersion, de diagrammes en boîte, de graphiques à barres, de graphiques en radar et de nombreux autres types de visualisations. • Matplotlib peut être utilisé pour créer des visualisations statiques ou interactives, et il peut être intégré à des applications web, des tableurs, des outils d'analyse de données et d'autres programmes.
Sklearn	• Le but de scikit-learn est de fournir une interface simple et cohérente pour les tâches d'apprentissage automatique, tout en offrant des performances de traitement très rapides. Il est utilisé dans différents domaines tels que la science des données, la bio-informatique, la finance, l'analyse de sentiments, la reconnaissance de la parole, l'analyse d'image et bien d'autres. • scikit-learn permet de simplifier le processus de travail avec des données volumineuses et complexes pour les appliquer à des modèles d'apprentissage automatique.

TABLEAU 1 : LES BIBLIOTHEQUES UTILISEES

IV.10 Evaluation

Pour évaluer la performance d'un modèle de classification, on utilise des mesures comme la précision, le rappel, L'exactitudes et le score f1, qui sont calculées à partir de la matrice de confusion.

Cette matrice est un tableau croisé entre les valeurs réelles et les prédictions, qui permettent de se faire une idée des performances de notre modèle. **[21]**

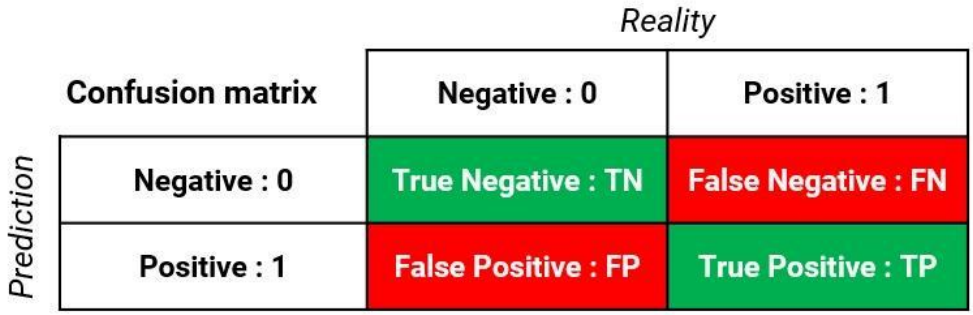


TABLEAU 2 : MATRICE DE CONFUSION

On distingue quatre catégories de résultats possibles dans cette matrice de confusion : les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs.

Vrai positif (VP) : Nombre de tweets le test déclare positifs et qui le sont réellement.

Faux positif (FP) : Nombre de tweets que le test déclare positifs et qui sont en réalité négatifs.

Vrai négatif (VN) : Nombre de tweets que le test déclare négatifs et qui sont en réalité négatifs.

Faux négatif (FN) : Nombre de tweets que le test déclare négatifs et qui sont en réalité positifs.

IV.10.1 La précision

(Precision en anglais) est une mesure de la qualité d'un modèle prédictif, indiquant la proportion de prédictions positives correctes parmi l'ensemble des prédictions positives. Elle est définie comme étant le ratio entre le nombre de vrais positifs (VP) et la somme des vrais positifs et des faux positifs (FP) : précision = VP / (VP + FP).

IV.10.2 Le rappel

(Recall en anglais) est une mesure de la sensibilité d'un modèle prédictif, indiquant la proportion de vrais positifs détectés parmi l'ensemble des vrais positifs et des faux négatifs.

Il est défini comme étant le ratio entre le nombre de vrais positifs et la somme des vrais positifs et des faux négatifs : rappel = VP / (VP + FN).

IV.10.3 L'exactitude

(Accuracy en anglais) est une mesure de la performance d'un modèle prédictif, indiquant la proportion de prédictions correctes parmi l'ensemble total des prédictions.

Elle est définie comme étant le ratio entre le nombre de prédictions correctes et le nombre total de prédictions : exactitude = (VP + VN) / (VP + VN + FP + FN).

IV.10.4 Le score F1

Mesure l'équilibre entre la précision et le rappel, et combine ces deux mesures en une seule valeur.

Il s'agit de la moyenne harmonique de la précision et du rappel, avec une pondération égale (c'est-à-dire, F1 score = 2*(Precision * rappel) / (Precision + rappel)).

IV.11 Matrice de confusion (algorithme KNN) :

```
Confusion Matrix for k = 1 is:

[[5349  609  346]
 [ 124 1833  225]
 [  55  134 1446]]

Classification Report for k = 1 is:

              precision    recall  f1-score   support

    negative       0.97      0.85      0.90      6304
     neutral       0.71      0.84      0.77      2182
    positive       0.72      0.88      0.79      1635

    accuracy                           0.85     10121
```

FIGURE 11 : MATRICE DE CONFUSION (ALGORITHME KNN)

IV.12 Matrice de confusion (algorithme naïve bayes) :

```
Confusion Matrix:
      0    1    2
0  1080  106   83
1   148  229   54
2    77   49  199
-------------------------------------------------
Classification Report:
              precision    recall  f1-score   support

    negative       0.83      0.85      0.84      1269
     neutral       0.60      0.53      0.56       431
    positive       0.59      0.61      0.60       325

    accuracy                           0.74      2025
```

FIGURE 12 : MATRICE DE CONFUSION (ALGORITHME NAÏVE BAYES**)**

IV.13 Matrice de confusion (algorithme SVM) :

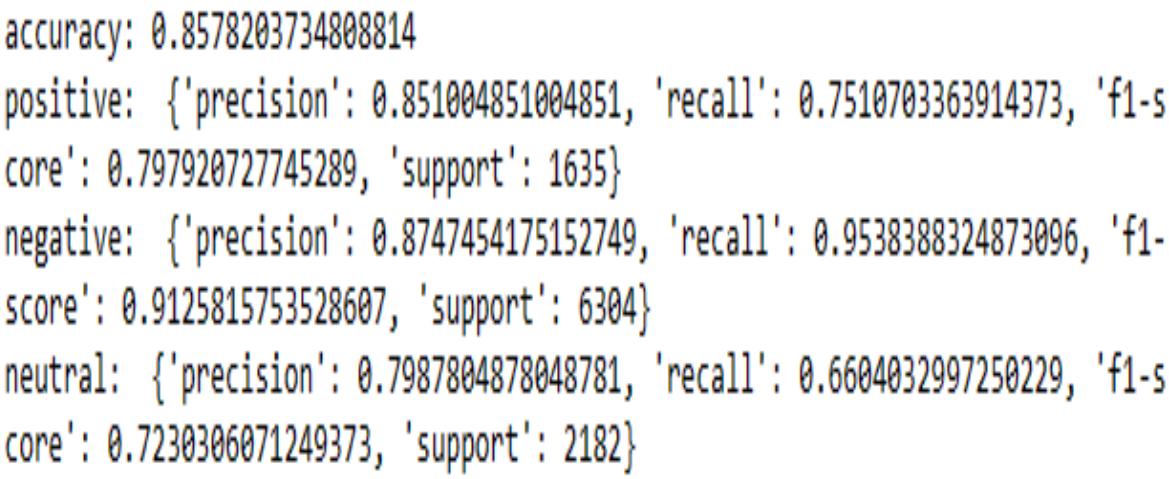


FIGURE 13 : MATRICE DE CONFUSION (ALGORITHME SVM)

IV.14 Classification et Comparaison

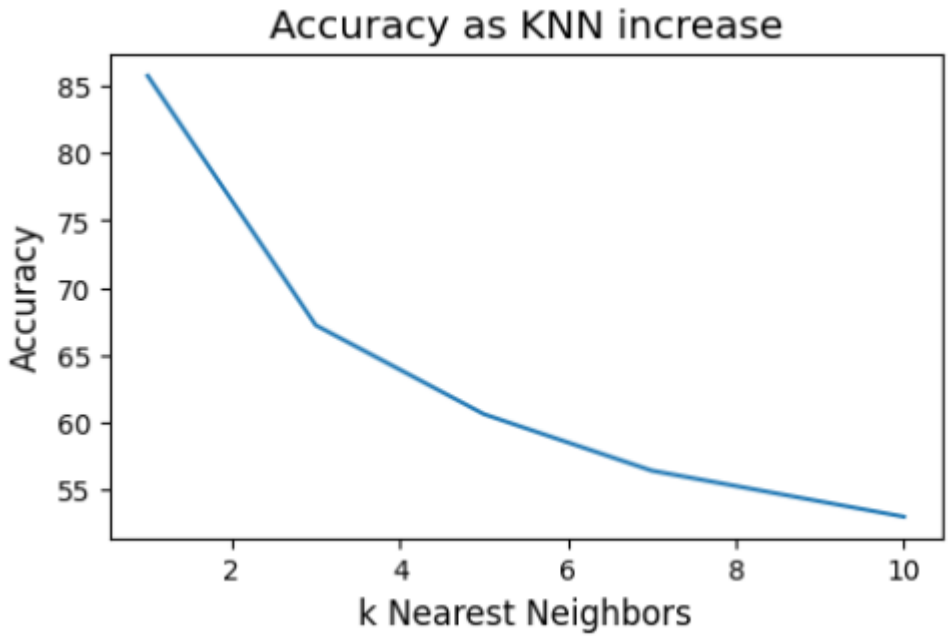
Nous avons utilisés trois classificateurs : SVM (Support Vector Machine) et Naïve Bayes, KNN.

Pour évaluer leur performance, nous avons mesuré le rappel, la précision, L'exactitude et le score F1. Par la suite, nous avons observé les résultats suivants.

Classificateurs	Accuracy	Classe positive			Classe négative			Classe neutre		
		Précision	Recall	F1	Précision	Recall	F1	Précision	Recall	F1
KNN	0.85	0.72	0.88	0.79	0.97	0.85	0.90	0.71	0.84	0.77
SVM	0.86	0.85	0.75	0.79	0.87	0.95	0.91	0.79	0.66	0.72
NB	0.74	0.59	0.61	0.60	0.83	0.85	0.84	0.60	0.53	0.56

TABLEAU 3 : LE RESULTAT D'EXACTITUDE DES CLASSIFIEURES

Lorsqu'on a utilisé le KNN, il s'est avéré que si la valeur de k est grande sa précision diminue.



Pour la performance, le Tableau 4 résume les performances selon les classificateurs utilisés.

Donc, le Support Vector Machine (86%) est plus performant en termes de précision et de spécificité.

En outre, il est également à noter que a donné la sensibilité la plus petite Naïve bayes (74%).

Les résultats obtenus montrent que la classification utilisant SVM donne de meilleurs résultats.

Classificateur	Accuracy
SVM	**86%**
KNN	85%
NB	74%

TABLEAU 4 : LES PERFORMANCES SELON LES CLASSIFIEURES UTILISES

IV.15 Conclusion

Dans ce chapitre, nous avons présenté le processus de collecte et d'annotation des tweets de notre corpus.

Nous avons également fourni une explication détaillée des méthodes de prétraitement utilisées sur notre corpus annoté et présenté les résultats en termes de précision et de rappel pour les trois classificateurs que nous avons utilisés pour détecter les émotions dans les tweets.

Nos résultats démontrent que SVM depasse l'autre classifieur.

Conclusion

Générale

Conclusion générale

Notre étude s'est concentrée sur la réalisation d'analyses automatiques et intelligentes des sentiments et sur la recherche d'opinions sur les médias sociaux et le Web sur le ChatGpt.

Nous avons commencé par faire des recherches sur le ChatGpt, les médias sociaux et les réseaux sociaux, avec un accent particulier sur Twitter.

De plus, nous avons exploré les méthodes actuelles utilisées dans l'analyse des sentiments.

Ensuite, nous avons décrit les étapes de notre étude, qui comprend la collecte des tweets, la Tokenisation, la suppression des mots vides et la lemmatisation.

Enfin, nous avons annoté notre corpus et classé les données.

Nos résultats révèlent que SVM affiche les meilleures performances.

Références

Les références

[1] Lewepedagogique.Chat GPT : pourquoi ce logiciel est dangereux et fabuleux à la fois. [En ligne]. Accéder le : 10/05/2023

https://lewebpedagogique.com/valmy/2023/01/26/chat-gpt-pourquoi-ce-logiciel-est-dangereux-et-fabuleux-a-la-fois/

[2] Oncrawl. OpenAI ChatGPT versus Google Bard. [En ligne]. Accéder le : 10/05/2023

https://fr.oncrawl.com/infographie/openai-chatgpt-versus-google-bard

[3] Je decide.be, Les inconvénients des media sociaux, [En ligne]. (31/03/2022). Accéder le : 15/04/2023

https://www.jedecide.be/les-parents-et-lenseignement/la-vie-privee-en-ligne/les-reseaux-sociaux-les-inconvénients

[4] Wikipédia. Les Médias sociaux mobiles. *Wikipédia.* [En ligne]. (08/11/2020). Accéder le : 03/06/2023

https://en.wikipedia.org/wiki/Social_media#Mobile_social_media

[5] L'Ecole Française. Les principaux réseaux sociaux. [En ligne]. Accéder le : 03/06/2023

https://lecolefrancaise.fr/a-quoi-servent-les-reseaux-sociaux/

[6] Das et Chen, 2001; Morinaga et al. 2002; Pang, Lee et Vaithyanathan, 2002;Tong, 2001, Tourney, 2002; Wiebe, 2000.

[7] HADJI Mehdi, Analyse des sentiments. [En ligne]. Accéder le : 10/05/2023

https://medium.com/@mehdihadji/analyse-des-sentiments

[8] Tibco. Qu'est-ce que l'analyse des sentiments. [En ligne]. Accéder le : 15/05/2023

https://www.tibco.com/fr/referjence-center/what-is-sentiment-analysis

[9] ATMANOU. Siham, MILI. Ferial Analyse des sentiments sur les avis des clients dans le E-Commerce. (2021).

[10] MADANI.Riad, Fouille d'opinions en utilisant l'apprentissage automatique supervisé. Corpus (2018).

[11] Theastrologypage. Kaggle. [En ligne]. Accéder le : 22/05/2023

https://fr.theastrologypage.com/kaggle

[12] Kaggle. Datasets and Machine Learning Projects - Kaggle. **CHARUNI. Sa** ChatGpt sentiment analysis. [En ligne]. Accéder le: 12/04/2023

https://www.kaggle.com/datasets

[13] NLP Cloud. API de Tokenisation ET de Lemmatisation. [En ligne]. Accéder le: 12/05/2023

https://nlpcloud.com/fr/nlp-tokenization-api.html

[14] Hamdi. Ikram, Automatic and Intelligent Sentiment Analysis and Opinion Mining on Social Medias. (2020).

[15] Myservername. Qu'est-ce que la machine vectorielle de support (SVM) dans l'apprentissage automatique. [En ligne]. Accéder le: 22/04/2023

https://fre.myservername.com/what-is-support-vector-machine-machine-learning

[16] Isnbreizh. Algorithme des k voisins les plus proches (knn). [En ligne]. Accéder le: 01/06/2023

https://www.isnbreizh.fr/nsi/activity/algoRefKnn/index.html

[17] Xlstat. classifieur bayésien naif. [En ligne]. Accéder le: 01/06/2023

https://www.xlstat.com/fr/solutions/fonctionnalites/classifieur-bayesien-naif

[18] Datascientest. NTLK : guide de l'outil de Traitement Naturel du Langage en Python. [En ligne]. Accéder le: 01/06/2023

https://datascientest.com/nltk

[19] Editions. C'est quoi, Pandas. [En ligne]. Accéder le: 01/06/2023

https://www.editions-eni.fr/open/mediabook.aspx?

[20] Ledatascientist. A La Découverte De La Célèbre Librairie NumPy. [En ligne]. Accéder le: 01/06/2023

https://www.google.com/amp/s/ledatascientist.com/amp/decouvrir-numpy/

[21] Morgan Morancey. Tutoriel d'introduction à Python 3. [En ligne]. Accéder le: 26/05/2023

https://mmorancey.perso.math.cnrs.fr/TutorielPython.html

[22] Kobia.Confusion, qu'est-ce que c'est [En ligne]. Accéder le: 02/06/2023https://kobia.fr/classification-metrics-matrice-de-confusion/