

# L'analyse sentimentale et l'apprentissage automatique : Méthodes hybrides et perspectives futures

## Sentimental analysis and machine learning: hybrid methods and future prospects

**Bilal AZDINE, (Doctorant)**

*Laboratoire Recherche en finance, Audit et Gouvernance des Organisations–LARFAGO  
École Nationale de Commerce et de Gestion  
Université Hassan 1er de Settat, Maroc*

**Mounime EL KABBOURI, (Enseignant-Chercheur)**

*Laboratoire Recherche en finance, Audit et Gouvernance des Organisations–LARFAGO  
École Nationale de Commerce et de Gestion  
Université Hassan 1er de Settat, Maroc*

<b>Adresse de correspondance :</b>	École Nationale de Commerce et de Gestion Km 3, route de Casa BP 658, Settat Université Hassan 1er Maroc (Settat) 05237-23577- 05 23 40 12 02/05 23 40 13 62.
<b>Déclaration de divulgation :</b>	Les auteurs n'ont pas connaissance de quelconque financement qui pourrait affecter l'objectivité de cette étude et ils sont responsables de tout plagiat dans cet article.
<b>Conflit d'intérêts :</b>	Les auteurs ne signalent aucun conflit d'intérêts.
<b>Citer cet article</b>	AZDINE, B., & EL KABBOURI, M. ( (2025). L'analyse sentimentale et l'apprentissage automatique : Méthodes hybrides et perspectives futures. <i>International Journal of Accounting, Finance, Auditing, Management and Economics</i> , 6(1), 166-190. <a href="https://doi.org/10.5281/zenodo.14633276">https://doi.org/10.5281/zenodo.14633276</a>
<b>Licence</b>	<b>Cet article est publié en open Access sous licence CC BY-NC-ND</b>

Received: November 27, 2024

Accepted: January 09, 2025

**International Journal of Accounting, Finance, Auditing, Management and Economics - IJAFAME**

**ISSN: 2658-8455**

**Volume 6, Issue 01 (2025)**

## **L'analyse sentimentale et l'apprentissage automatique : Méthodes hybrides et perspectives futures**

### **Résumé :**

L'analyse de sentiment (SA) utilisant le langage naturel est l'un des sujets les plus passionnants de cette décennie en raison de l'utilisation généralisée des médias sociaux, qui permet d'analyser les interactions entre les individus et de déterminer les structures importantes dans ces communications. La SA est définie comme un domaine d'étude qui utilise des méthodes d'intelligence pour analyser, traiter et révéler les sentiments, émotions et ressentis des individus cachés derrière un texte ou une interaction grâce au traitement du langage naturel. La SA utilise des techniques d'apprentissage automatique (ML) pour tirer les inférences nécessaires des interactions des utilisateurs. Les algorithmes de ML tels que l'apprentissage non supervisé, supervisé, semi-supervisé, et les techniques d'apprentissage profond sont utilisés pour tirer des informations utiles des sentiments. Collecter les opinions des individus et prendre des décisions basées sur celles-ci peut être bénéfique pour de nombreuses personnes, et à travers cela, des informations commerciales utiles peuvent être tirées. Cependant, le traitement de données massives et multilingues, la détermination de la polarité de sentiment appropriée, le traitement du sarcasme et des emojis, et la sélection de la technique de ML appropriée pour construire le modèle d'analyse sont quelques-uns des défis auxquels est confrontée la SA. Ce chapitre fournit un aperçu détaillé des tendances et des défis de l'utilisation du ML dans la SA. Afin d'établir des objectifs futurs, ce chapitre examine les difficultés associées à la SA. Cet article utilise une approche méthodologique de type état de l'art. Il fournit un aperçu détaillé des tendances, des défis et des différentes méthodes d'analyse sentimentale, tout en intégrant des éléments de recherche théorique. Cette méthodologie permet de synthétiser les connaissances existantes sur le sujet et d'explorer les futures directions de recherche.

**Mots clés :** Apprentissage Automatique, Traitement du Langage Naturel, Opinion, Polarité, Analyse de sentiment.

**JEL Classification :** G41.

**Type du papier :** Recherche Théorique.

### **Abstract :**

Sentiment Analysis (SA) using natural language is one of the most exciting topics of this decade due to the widespread use of social media, which allows for the analysis of interactions between individuals and the determination of important structures in these communications. SA is defined as a field of study that utilizes intelligence methods to analyze, process, and reveal the feelings, emotions, and sentiments of individuals hidden behind text or interaction through natural language processing. SA employs machine learning (ML) techniques to draw necessary inferences from user interactions. ML algorithms such as unsupervised, supervised, semi-supervised learning, and deep learning techniques are used to extract useful information from sentiments. Collecting individuals' opinions and making decisions based on them can be beneficial for many people, and through this, useful business insights can be gleaned. However, handling massive and multilingual data, determining appropriate sentiment polarity, dealing with sarcasm and emojis, and selecting the appropriate ML technique to build the analysis model are some of the challenges faced by SA. This chapter provides a detailed overview of trends and challenges in using ML in SA. To establish future objectives, this chapter examines the difficulties associated with SA. This article uses a state-of-the-art methodological approach. It provides a detailed overview of the trends, challenges and different methods of sentimental analysis, while incorporating elements of theoretical research. This methodology makes it possible to synthesise existing knowledge on the subject and explore future research directions.

**Keywords:** Machine Learning; Natural Language Processing; Opinion, Polarity, Sentiment Analysis.

**Classification JEL:** G41.

**Paper type:** Theoretical Research.

## 1. Introduction

L'analyse de sentiment (SA) a gagné en importance grâce à l'explosion des médias sociaux et des plateformes générant des données massives. La SA se concentre sur l'identification et la classification des émotions, opinions et attitudes exprimées dans les textes, en utilisant des techniques de traitement du langage naturel. En revanche, l'exploration d'opinions vise spécifiquement à extraire des jugements ou évaluations implicites ou explicites, souvent associés à des aspects particuliers d'un produit ou d'un service. Bien que les deux concepts soient liés, l'analyse sentimentale est plus large et inclut des dimensions émotionnelles, tandis que l'exploration d'opinions se concentre davantage sur les aspects évaluatifs. Cette discipline se concentre sur la compréhension des opinions, émotions et ressentis exprimés par les individus, offrant ainsi des perspectives essentielles pour les décisions commerciales, politiques et sociales.

En effet, l'analyse sentimentale consiste à détecter le sentiment émotionnel conservé dans un texte par le biais d'une analyse sémantique spécialisée à des fins diverses, par exemple pour évaluer la qualité de l'accueil réservé par le marché à un nouveau produit et le retour global des clients ou pour estimer la popularité d'un produit ou d'une marque (Ghiassi, Skinner, & Zimbra, 2013 ; Mostafa, 2013) parmi les gens. Un certain nombre de recherches se concentrent sur l'analyse des sentiments ou sur ce que l'on appelle le « sondage d'opinion » (Balahur, Steinberger, Goot, Pouliquen, & Kabadjov, 2009 ; Cambria, Schuller, Yunqing, & Havasi, 2013 ; Hsinchun & Zimbra, 2010). Elle est principalement basée sur l'identification des mots positifs et négatifs et sur le traitement du texte dans le but de classer sa position émotionnelle comme positive ou négative. Un exemple de ces efforts d'analyse du sentiment est le travail de Maks et Vossen (2012) qui présente un modèle de lexique pour l'analyse profonde du sentiment et l'exploration de l'opinion. Un concept similaire peut être utilisé dans d'autres domaines de recherche comme la détection des émotions.

Cependant, une analyse du sentiment émotionnel des textes d'actualités peut également être explorée pour la prédiction du marché. Schumaker, Zhang, Huang et Chen (2012) ont tenté d'évaluer le sentiment dans les articles d'actualités financières en relation avec le marché boursier dans leurs recherches, mais n'ont pas été complètement réussis. Un exemple plus réussi et récent de cela est celui de Yu, Wu, Chang et Chu (2013), où un modèle d'entropie contextuelle a été proposé pour élargir un ensemble de mots semences en découvrant des mots émotionnels similaires et leurs intensités correspondantes à partir d'articles d'actualités boursières en ligne. Cela a été accompli en calculant la similarité entre les mots semences et les mots candidats à partir de leurs distributions contextuelles en utilisant une mesure d'entropie. Une fois que les mots semences ont été élargis, les mots semences et les mots élargis sont utilisés pour classifier le sentiment des articles d'actualités. Leurs résultats expérimentaux montrent que l'utilisation des mots émotionnels élargis a amélioré la performance de classification, qui a été encore améliorée en intégrant leurs intensités correspondantes, ce qui a fait varier les résultats de précision de 52 % à 91,5 % en faisant varier la différence des niveaux d'intensité entre les classes positives et négatives de (-0,5 à 0,5) à >9,5 respectivement. Il est également intéressant de noter que l'analyse émotionnelle du texte ne doit pas être simplement basée sur la positivité-négativité et peut être effectuée sur d'autres dimensions ou sur des dimensions multiples (Ortigosa-Hernández et al., 2012). Une recherche récente de Loia et Senatore (2014) introduit un cadre pour extraire les émotions et les sentiments exprimés dans les données textuelles. Les sentiments sont exprimés par une polarité positive ou négative. Les émotions sont basées sur la conception de Minsky des émotions, qui se compose de quatre dimensions affectives (Agrément, Attention, Sensibilité et Aptitude). Chaque dimension a six niveaux d'activation, appelés niveaux sentiques. Chaque niveau représente un état émotionnel et peut être plus ou moins intense, selon la position sur

la dimension correspondante. Un autre développement intéressant dans l'analyse des sentiments est l'effort de passer du sentiment de morceaux de texte à des caractéristiques ou aspects spécifiques liés à un concept ou à un produit ; Kontopoulos, Berberidis, Dergiades et Bassiliades (2013) proposent une analyse des sentiments plus efficace des publications Twitter, où les publications ne sont pas simplement caractérisées par un score de sentiment mais reçoivent plutôt une note de sentiment pour chaque notion distincte en elles, ce qui est rendu possible grâce à l'aide d'une ontologie. Li et Xu (2014) adoptent un autre angle en recherchant des caractéristiques qui sont « significatives » pour les émotions au lieu de simplement choisir des mots avec un degré de cooccurrence élevé, créant ainsi une classification des émotions basée sur le texte en utilisant l'extraction des causes émotionnelles.

Les chercheurs s'intéressent de plus en plus à l'analyse et à l'exploitation des sentiments des utilisateurs en raison de l'expansion rapide d'Internet et de l'utilisation de diverses techniques de fouille de données. L'analyse de sentiment (SA) est l'un des principaux domaines d'étude actuels qui combine la fouille de données et le traitement du langage naturel (NLP). Les sentiments des utilisateurs sont extraits à l'aide de divers algorithmes et techniques d'exploration d'opinions. Les principales contributions de cet article donnent un aperçu détaillé de la SA, des différentes étapes impliquées, du niveau d'analyse, des informations de base et des travaux connexes.

Donc l'objectif de cet article est de donner des éléments des réponses à la question principale suivante: **Quels est le rôle de Machine Learning dans l'Analyse de Sentiment?**

A partir de cette question on peut scinder les questions suivantes :

- Quels sont les fondements de l'analyse de sentiment ?
- Quels sont les enjeux et les perspectives de l'intégration de Machine Learning dans l'analyse de sentiment ?
- Quels sont les futures directions et les orientations du domaine de l'analyse de sentiment ?

La méthode de recherche que nous allons essayer de suivre pour la réalisation de notre article est de donner des éléments de réponses à notre problématique à travers le plan suivant :

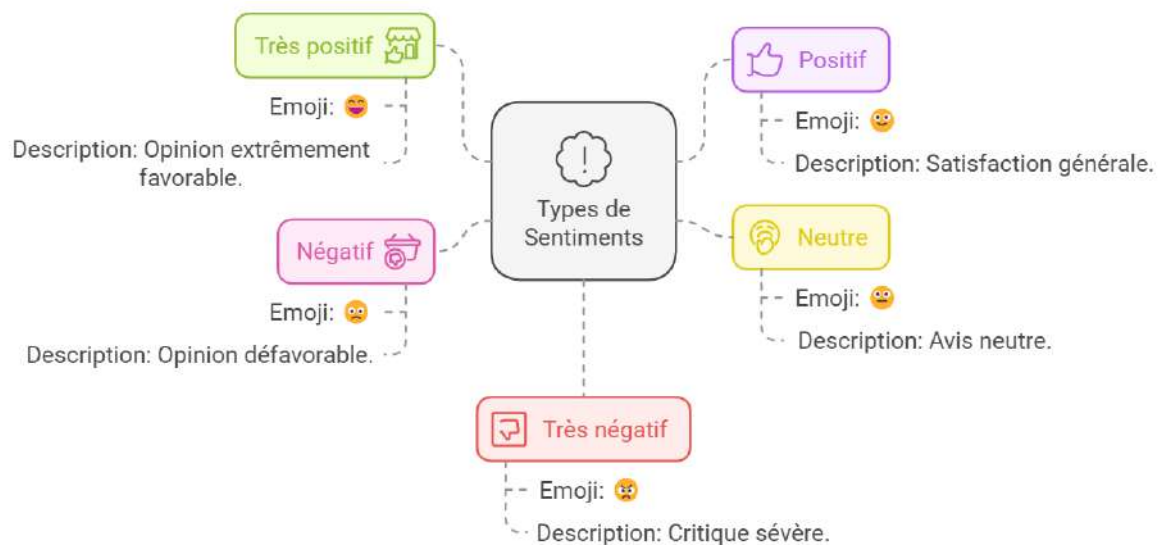
La recherche de fond pour l'Analyse Sentimentale est couverte dans la première section. Les métriques de performance sont discutées dans la deuxième section, tandis que la troisième section répertorie les outils disponibles pour l'Analyse Sentimentale. Dans la quatrième section, les principaux domaines d'intérêt de l'Analyse Sentimentale sont discutés, et les défis sont abordés dans la cinquième section. Nous discutons des conclusions dans la sixième section et abordons les orientations futures du domaine dans la septième section.

## **2. Fondements de l'analyse de sentiment :**

L'étude analytique des idées, émotions, sentiments et attitudes des individus est appelée SA Liu Bing (2012). L'exploration d'opinions ou l'analyse de sentiment est utilisée pour étudier les émotions humaines, à partir desquelles nous pouvons déduire des insights. Le terme "analyse de sentiment" est peut-être apparu pour la première fois dans l'étude de Yi et al. (2003), tandis que le terme "exploration d'opinions" est apparu pour la première fois dans l'étude de Dave et al.(2003). Lors de la prise de décisions - les individus, les gouvernements, les entreprises et les réseaux sociaux peuvent tous grandement bénéficier et être influencés par les points de vue des autres. Les opinions ou sentiments associés à des termes tels que "évaluation", "appréciation", "attitude", "émotion" et "humeur" qui leur sont liés concernent tous nos pensées et sentiments subjectifs. Ils véhiculent tous la même signification. Cependant, selon certains experts, "l'exploration d'opinions" et "l'analyse de sentiment" ont des idées légèrement distinctes Tsytsarau Mikalai et al. (2012). Le sentiment est le principal influenceur de nos comportements. L'analyse de sentiment est un sujet d'étude qui utilise

diverses techniques informatiques pour extraire les opinions et attitudes des individus à partir de textes en langage naturel. Les individus expriment leurs sentiments envers des marques spécifiques, des produits, des services, des personnes, des entreprises et des organisations lorsqu'ils parlent de certains sujets, de problèmes sociaux et de divers autres sujets. Les gens utilisent des emojis sur les réseaux sociaux en plus des textes conventionnels pour exprimer leurs émotions. Par conséquent, il existe différentes méthodes pour identifier les sentiments, et l'objectif est de découvrir les sentiments cachés dans les documents plutôt que dans les textes Thelwall Mike et al. (2010). Principalement, les sentiments sont catégorisés en "très positif", "positif", "neutre", "négatif" et "très négatif". Cela est illustré dans la Figure.1, et tous les sentiments expriment les sentiments humains envers la cible visée.

Figure 1 : Types de sentiments.



Source : Elaboré par nous même dans le cadre de cet article.

L'analyse de sentiment se concentre principalement sur les croyances qui expriment explicitement ou implicitement des sentiments positifs ou négatifs, souvent désignés comme des sentiments "positifs" ou "négatifs" dans le langage courant. L'utilisation des commentaires des clients pour améliorer un produit, un service ou une marque est possible grâce à l'analyse de sentiment. Lorsque nous analysons les opinions des utilisateurs sur un produit spécifique, par exemple, des mots-clés tels que "excellent", "incroyable" ou "bon" indiquent les sentiments positifs des utilisateurs à propos du produit, ce qui contribue à construire une réputation de marque favorable. L'opinion des clients sur un produit spécifique, telle que "très mauvais", "mauvais" ou "pire", est indiquée par des mots-clés qui reflètent les sentiments négatifs de l'utilisateur à propos du produit, ce qui entraîne une réputation de marque défavorable. Les opinions neutres ne sont ni bonnes ni mauvaises, et les publications neutres ne bénéficient ni ne nuisent à la réputation d'une marque. Cependant, les exemples d'entraînement neutres aident les apprenants à distinguer plus efficacement les exemples positifs des exemples négatifs. Certains sentiments sont plus faciles à identifier que d'autres, et certains sont difficiles à identifier, surtout les sentiments surprenants.

L'expansion rapide des forums de discussion, des plateformes de commerce électronique, des sites d'avis sur les produits et des réseaux sociaux a permis un flux incessant d'idées et de points de vue. En raison de ces nombreux canaux, les entreprises trouvent difficile de compiler les opinions et les attitudes concernant les articles. La croissance du contenu généré

par les utilisateurs sur Internet et des outils tels que l'analyse de sentiment donne aux spécialistes du marketing la possibilité d'en apprendre davantage sur ce que les clients pensent de leurs produits Rambocas Meena et al. (2018). L'analyse de sentiment est essentielle pour le monde des affaires car elle aide les entreprises à affiner leurs stratégies et à en savoir plus sur ce que leurs clients disent de leurs produits. La compréhension du sentiment des clients devient de plus en plus importante dans la culture d'entreprise actuelle axée sur le client. L'analyse de sentiment peut être utilisée efficacement pour comprendre l'attitude globale du client envers votre entreprise, ses produits ou ses services. Chaque client exprime ses sentiments de manière unique et dans différents contextes, que ce soit directement ou indirectement. En utilisant trois moyens pour suivre ou surveiller les sentiments des clients, à savoir :

1. Les avis
2. Les réseaux sociaux
3. Les retours directs.

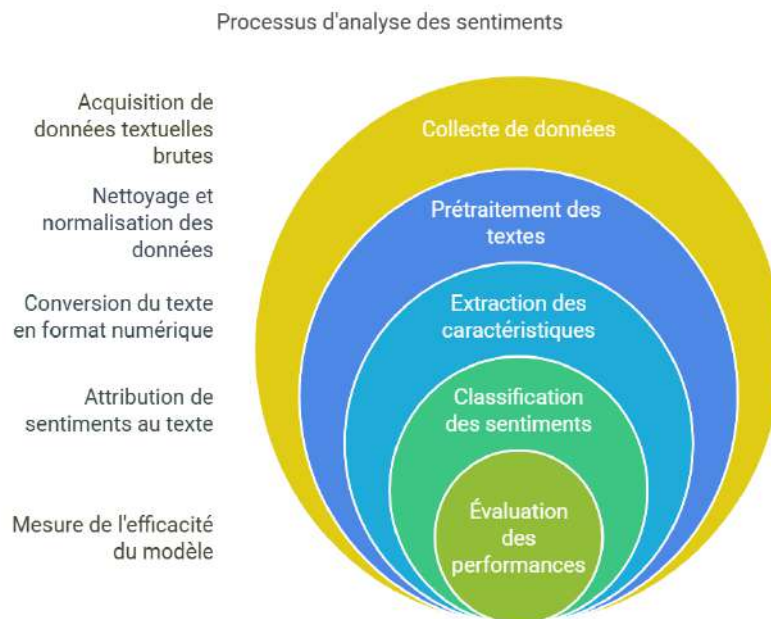
Les évaluations et avis des clients sur les produits, services et organisations sont désignés sous le terme de "données d'avis clients". Les professionnels du marketing peuvent contacter les clients nécessitant une attention particulière en extrayant les émotions des avis sur les produits, ce qui permet d'augmenter la satisfaction et les ventes des clients, et finalement de bénéficier à l'entreprise Vyas Vishal et Uma Vijayasundara (2019). Au-delà de l'évaluation uniquement des étoiles, la technologie moderne permet la classification des sentiments des avis. En essence, l'exploration d'opinions permet aux entreprises de comprendre l'attitude d'un client en analysant l'opinion réelle dans un avis. L'analyse de sentiment fournit des informations qui vous permettent de personnaliser votre réponse plutôt que de répondre de la même manière à chaque avis. À partir des avis des utilisateurs, les tendances de sentiment peuvent être révélées ou extraites par des algorithmes. Extraire des informations exploitables à partir des avis aide l'entreprise à évoluer dans une direction positive en établissant un lien solide avec les clients.

Les médias sociaux désignent un ensemble de plateformes de communication en ligne qui favorisent le contenu généré par les utilisateurs, la participation et l'implication communautaire. Les gens peuvent utiliser ces médias sociaux pour se connecter les uns aux autres, rejoindre une communauté de personnes partageant les mêmes idées, puis participer à la conversation. D'un point de vue commercial, si nous réfléchissons, nous pouvons écouter et comprendre ce qui se passe sur le marché en analysant les faits. En raison de l'utilisation élevée des médias sociaux parmi tous les groupes d'âge, de nombreuses données en temps réel sont générées, et en utilisant cela, nous pouvons efficacement exploiter les sentiments des personnes à l'égard de tout ce qui est discuté via les médias sociaux. Les retours directs sont obtenus des clients en leur demandant leur avis sur une entreprise, un produit, une expérience de service client ou d'autres points de contact dans le parcours client. Comme les consommateurs sont des ambassadeurs de la marque, il est judicieux de leur demander directement leur expérience avec l'entreprise ou le produit. Cela aide l'entreprise à mieux comprendre les besoins de ses clients. Cela donne aux clients un sentiment de valeur tout en reflétant précisément ce qu'ils ressentent à propos de l'entreprise. L'analyse sociale devrait permettre de comprendre les éléments essentiels des clients afin de cibler ou de viser de nouvelles offres et produits de manière rentable via divers canaux de médias sociaux. Pour améliorer la marque, réagissez plus rapidement avec un aperçu précis, opportun et pertinent des demandes des clients.

Il devrait également être hautement évolutif et capable d'effectuer une recherche robuste capable de recueillir des informations à partir de multiples canaux de médias sociaux, tels que les blogs, Twitter, Facebook, LinkedIn, et bien d'autres encore. Cela fournit une plateforme pour se connecter et interagir avec les clients, promouvoir les entreprises et augmenter le

trafic sur le site Web. Cela aide à développer l'activité commerciale, et en analysant les données, l'expérience client peut être améliorée, ce qui contribue à améliorer la réputation de la marque parmi nos concurrents. Des mesures d'indicateurs de performance clés peuvent être utilisées pour finalement renforcer l'entreprise. Les experts en médias sociaux ont proposé quelques idées pour améliorer l'impact des médias sociaux sur les entreprises, telles que l'établissement d'objectifs précis pour chaque campagne de médias sociaux et la création de métriques basées sur ces objectifs pour accroître la réputation de la marque. En utilisant des technologies intelligentes Hemanth DJude (2020), nous pouvons efficacement obtenir des insights à partir des données acquises, et nous pouvons comparer les performances en utilisant différentes mesures de performance (Figure. 2).

**Figure 2 : Flux de travail de l'analyse de sentiment.**



*Source : Elaboré par nous même dans le cadre de cet article.*

Les cinq principales étapes de l'analyse de sentiment sont les suivantes :

- **Collecte de données** : C'est la première étape de l'apprentissage automatique (ML). Les données provenant de diverses sources sont collectées et analysées afin d'extraire les informations nécessaires. Les données au format texte brut, JSON, XML, HTML, .xls ou .csv peuvent être prises en compte. Des données provenant de différentes plateformes de médias sociaux, collectées ou déjà stockées, peuvent être utilisées. La plateforme d'analyse de sentiment sur les médias sociaux la plus populaire est Twitter. Cela, à notre avis, est dû aux interfaces de programmation d'applications (API) bien organisées de Twitter et à l'abondance de packages R et Python disponibles et qui facilitent la collecte de données. Facebook limite les importations de données, ce qui rend difficile la collecte de messages, de retours et de statistiques d'engagement des utilisateurs.
- **Prétraitement du texte** : Dans cette étape, les données sont prétraitées et préparées pour un traitement ultérieur. Le traitement des données dépendra du type d'informations qu'elles contiennent. Cette étape est très importante car la précision de cette étape implique clairement l'exactitude du modèle. Le stemming, la lemmatisation, la tokenisation, le nettoyage de texte en ligne, la correction orthographique, la suppression des mots vides, la suppression des mots répétés, etc. sont effectués selon les besoins.

Nous serons en mesure de traiter les données et de les préparer pour l'analyse en utilisant des techniques de nettoyage de texte.

- **Extraction des caractéristiques** : En fonction des besoins, les caractéristiques requises doivent être extraites à partir des données prétraitées.
- **Classification du sentiment** : Il s'agit d'une approche automatisée pour localiser et classer les opinions dans le texte. Divers algorithmes sont utilisés pour la classification des différents sentiments. Pour l'analyse de sentiment, des approches algorithmiques basées sur les lexiques, l'apprentissage automatique (ML) et hybrides sont utilisées.
- **Évaluation des performances** : Cependant, évaluer les sentiments est un défi significatif, et obtenir de bons résultats est considérablement plus difficile que ce que les gens croient généralement. Une partie importante de la mesure de l'exactitude en utilisant l'analyse de sentiment est jouée par l'évaluation des performances. L'analyse de sentiment utilise des métriques d'évaluation telles que "précision, rappel et F-score". Des métriques de performance appropriées doivent être sélectionnées en fonction de l'ensemble de données et de l'algorithme utilisé.

La sélection de l'algorithme optimal reste encore un défi car chaque technique a ses avantages et ses inconvénients, bien que de nombreux classificateurs aient été examinés à des fins d'analyse de sentiment.

## **2.1.Niveau d'analyse :**

L'analyse de sentiment peut se produire à l'un des trois niveaux, c'est-à-dire au niveau du document, au niveau de la phrase ou au niveau de l'aspect.

### **2.1.1. Niveau du document :**

L'objectif de ce niveau est d'identifier le point de vue général du document. L'analyse de sentiment au niveau du document part du principe que chaque document contient des opinions sur un sujet unique. Ce type d'analyse de sentiment n'est pas couramment utilisé. Il peut être appliqué pour catégoriser les chapitres ou les pages d'un livre comme étant bons, négatifs ou neutres Wankhade Mayur (2022). Le vecteur de caractéristiques dans ces tâches est une collection de mots qui doit être contrainte au domaine donné.

### **2.1.2. Niveau de la phrase :**

À ce niveau, le défi consiste à identifier si chaque phrase exprime une opinion ou non. Dans ce cas, les faits réels sont exprimés dans des phrases objectives, tandis que les opinions sont exprimées dans des phrases subjectives. Les solutions dans cette situation consistent en deux étapes : premièrement, déterminer si la phrase communique une opinion ou non, et deuxièmement, déterminer la polarité de cette opinion. Le problème majeur, cependant, est que les déclarations objectives peuvent inclure un langage subjectif. Il ne suffit pas de simplement comprendre que les déclarations peuvent être interprétées de manière positive ou négative. Il s'agit d'une phase intermédiaire qui aide à éliminer les déclarations qui manquent d'opinion et, dans une certaine mesure, à déterminer si les opinions sur les choses et leurs caractéristiques sont positives ou négatives. Bien que l'analyse de sentiment aux niveaux du document et de la phrase soit utile, elle ne révèle pas ce que les gens aiment ou n'aiment pas, ni n'identifie les cibles d'opinion Katrekar Ashish (2005).

### **2.1.3. Analyse au niveau de l'aspect :**

À ce niveau, on utilise le NLP et on effectue une analyse plus fine. À ce niveau, la polarité et une cible de pensée définissent une opinion. Dans cette situation, les solutions consistent en deux étapes : premièrement, identifier l'élément et ses éléments pertinents, et deuxièmement, évaluer le point de vue sur chaque aspect. Découvrez si une opinion est soit "positive,

négative ou neutre" en identifiant et en extrayant les caractéristiques d'objet sur lesquelles le détenteur de l'opinion s'est concentré.

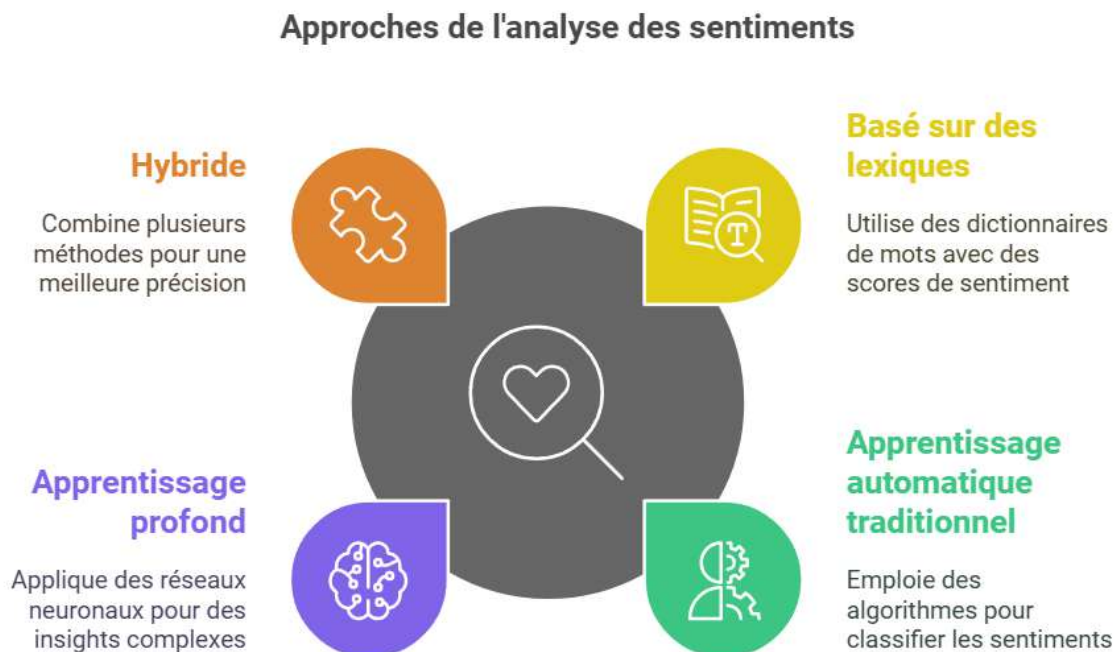
Il existe quatre tâches distinctes qui consistent à trouver des sentiments pour des aspects :

1. Extraction de texte : Extraire les mots et les phrases des avis.
2. Classification des sentiments : Utiliser un classificateur de sentiments pour identifier les données extraites qui contiennent un sentiment positif, négatif ou neutre.
3. Extracteur d'aspects : Déterminer la polarité de chaque aspect détecté et effectuer l'extraction des termes d'aspect pour ceux qui expriment un sentiment.
4. Agrégation de la polarité des aspects : Combiner les opinions exprimées sur les aspects pour créer un résumé final.

## 2.2.Contexte :

L'analyse de sentiment (SA) devient très populaire en raison de la croissance rapide des applications basées sur Internet et de l'utilisation extensive des médias sociaux et d'autres sites. Cette énorme quantité de données ne peut pas être contrôlée et analysée efficacement de manière traditionnelle, c'est pourquoi de nombreux chercheurs ont développé des techniques efficaces pour y faire face. Plusieurs enquêtes récentes Wankhade Mayur (2022), Aqlan Ameen Abdullah Qaid et al. (2019), Lighthart Alexander et al. (2021) ont abordé les méthodes, les applications et les défis de l'analyse de sentiment. Les auteurs ont souligné la problématique de l'analyse de sentiment et suggéré des orientations potentielles. Des efforts considérables sont déployés dans le domaine de l'analyse de sentiment pour répondre aux défis commerciaux en temps réel. Les méthodes d'analyse de sentiment existantes sont divisées en quatre classes principales : basées sur les lexiques, l'apprentissage automatique traditionnel, l'apprentissage profond (DL) et des méthodes hybrides. (Figure. 3).

*Figure 3 : Les différentes approches utilisées dans l'analyse des sentiments.*



*Source : Elaboré par nous même dans le cadre de cet article.*

### **2.2.1. Approche basée sur les lexiques :**

Cette approche agrège les scores de sentiment de tous les termes dans le document pour évaluer un document en utilisant un lexique de sentiments préparé à l'avance Whissell Cynthia M. (1989), Vasileios Hatzivassiloglou et Kathleen R. McKeown (1997). En ce qui concerne les approches basées sur les lexiques, le premier travail a été présenté par Peter D. Turney, Michael L. Littman (2003). Les approches basées sur les dictionnaires et les corpus sont les deux principales méthodes utilisées pour étiqueter les mots subjectifs avec une orientation sémantique.

#### **2.2.1.1.Approche basée sur les dictionnaires :**

Le dictionnaire comprend des sentiments qui ont été exprimés à travers des mots, des phrases ou des concepts. Cette méthode présente l'avantage d'être simple et de fournir des relations lexicales riches entre les mots. L'approche basée sur les dictionnaires présente un inconvénient significatif en ce qu'elle ne peut pas identifier les mots d'opinion avec des orientations spécifiques au domaine et au contexte et produit un grand nombre de faux positifs en raison des entrées courtes du dictionnaire.

#### **2.2.1.2.Approche basée sur les corpus :**

Le défi de découvrir des opinions avec des orientations spécifiques au contexte est surmonté par l'approche basée sur les corpus. Lorsque les domaines sont distincts, les résultats sont meilleurs. Le principal avantage d'une approche basée sur les corpus est celui-ci. Ils ne peuvent pas être employés séparément dans cette stratégie en raison des défis de fournir des textes substantiels tout en étant capable de couvrir tous les termes du texte.

Pour identifier l'orientation sémantique du texte, les approches basées sur les lexiques utilisent des adjectifs et des adverbes. Afin de calculer l'orientation de n'importe quel texte, des mots de combinaison d'adjectifs et d'adverbes sont récupérés avec leur "valeur d'orientation de sentiment" Neha Gupta et Rashmi Agrawal(2020). Pour un avis ou un texte particulier, l'approche basée sur les lexiques effectue l'agrégation des scores pour chaque jeton, c'est-à-dire que les valeurs "positive", "négative" et "neutre" sont additionnées indépendamment. Le texte se voit attribuer une polarité globale dans la dernière étape en fonction de la valeur maximale des scores individuels. En conséquence, le texte ou le document est d'abord divisé en jetons de mots simples ; ensuite, la polarité de chaque jeton est déterminée et ensuite totalisée. La dépendance au domaine est un inconvénient de cette approche Alejandro Moreo, et al. (2012).

L'extraction de sentiments à partir de texte en utilisant une méthode basée sur les lexiques a été présentée par Taboada et al. Dans leur article « Lexicon-based methods for sentiment analysis ». Le Calculateur de l'Orientation Sémantique est utilisé pour la tâche de classification de la polarité, qui consiste à étiqueter les textes positivement ou négativement en fonction de leur ressenti vis-à-vis de leur sujet principal. Cette recherche nous a conduit à la conclusion que les techniques d'AS basées sur les lexiques sont fiables, produisent de bons résultats inter-domaines et sont simples à améliorer. Cependant, leur technique ne peut pas analyser le sarcasme. La taxonomie Serendio a été utilisée par Palanisamy et al.(2013) pour créer une technique basée sur les lexiques permettant d'identifier les sentiments.

"Positif, négatif, négation, mots vides et expressions" constituent la taxonomie de Serendio. L'approche basée sur les lexiques pour les commentaires des étudiants a été mise en œuvre par Khin Zezawar Aung; Nyein Nyein Myo (2017). La positivité, la négativité et la neutralité ont été détectées avec succès ; la restriction était l'incapacité à évaluer le sarcasme. Pour l'étude du sentiment dans les actualités, Taj Soonh et al. (2019) proposent une méthode basée sur les lexiques. Les essais ont été réalisés en utilisant l'ensemble de données des actualités de la BBC de 2004 à 2005, et il a été découvert que les catégories des affaires et des sports

contenaient plus d'éléments positifs, tandis que le divertissement et la technologie étaient prédominants dans les histoires négatives. Kundi Fazal Masud et al. (2014) ont utilisé des lexiques et des dictionnaires pour créer un cadre de classification des sentiments. Ils ont obtenu une précision de 92 % dans la classification binaire et une précision de 87 % dans la classification multi-classes. Cependant, les approches basées sur les lexiques ont l'inconvénient de nécessiter une intervention humaine lors de l'analyse de texte Alsayat Ahmed (2022).

### **2.2.2. Apprentissage automatique :**

Les techniques d'apprentissage automatique ont déjà été utilisées pour effectuer une grande quantité de recherche dans le domaine de l'analyse de sentiment. Dans la plupart des cas, l'approche d'apprentissage automatique surpasse l'approche basée sur les lexiques Ashwini Patil et Shiwani Gupta (2015). Les méthodes d'apprentissage automatique et de traitement du langage naturel (NLP) peuvent être utilisées pour l'analyse de sentiment. Les techniques d'apprentissage automatique ont déjà été largement utilisées dans le domaine de l'analyse de sentiment, et c'est maintenant l'une des nombreuses zones d'étude computationnelle en NLP. Il existe trois catégories de techniques d'apprentissage automatique : non supervisées, supervisées et semi-supervisées.

#### **2.2.2.1. Non supervisé :**

Les techniques d'apprentissage non supervisées regroupent des données non étiquetées similaires ensemble. L'analyse de sentiment non supervisée est cruciale pour de nombreuses applications sur les médias sociaux, car il est simple de collecter d'énormes quantités de données non étiquetées mais très difficile et coûteux d'obtenir des étiquettes de sentiment Hu Xia et al. (2013). Lorsque nous n'avons pas de jeux de données bien étiquetés, la technique d'apprentissage non supervisé peut être utilisée. L'analyse de dépendance non supervisée a été utilisée par Fernández-Gavilanes et al. (2016) pour proposer une stratégie de prédiction du sentiment dans des textes non structurés, et elle a démontré des performances compétitives. La technique non supervisée basée sur les règles floues pour les publications sur les médias sociaux a été utilisée par Vashishtha et Susan (2019). En comparaison avec l'état de l'art, ils ont produit de meilleurs résultats.

Pour évaluer les flux Twitter sur les sentiments d'une marque spécifique à l'aide du jeu de données réel collecté sur une année, Hima Suresh et Gladston Raj S. (2016) ont suggéré une nouvelle approche de clustering flou. Ensuite, en utilisant les mesures de précision, de rappel et de temps d'exécution, une comparaison est effectuée avec les approches de clustering par partitionnement actuellement utilisées, les algorithmes K-means et d'espérance-maximisation. La méthode proposée est testée pour vérifier si elle peut produire des résultats de haute qualité pour l'analyse de sentiment sur Twitter, selon l'analyse expérimentale. Dans la comparaison des deux méthodes pour extraire le sentiment des tweets, Jitendra Kumar Rout et al. (2018) ont découvert que la méthode proposée était plus précise que la méthode basée sur les lexiques. Pour l'analyse de sentiment des actualités financières, Yadav et al. (2020) ont développé des techniques hybrides et des techniques nom-verbe et ont obtenu de bons résultats.

#### **2.2.2.2. Supervisé :**

Son utilisation de jeux de données étiquetés pour entraîner des algorithmes de classification des données le définit. Lorsque la variable de sortie est catégorique, c'est-à-dire qu'elle a deux classes ou plus, on utilise la classification. Les algorithmes d'apprentissage automatique tels que "Machine à Vecteurs de Support (SVM)", "Naive Bayes (NB)", "Régression Logistique", "Forêt Aléatoire", "Arbre de Décision" et "Réseaux de Neurones Artificiels (ANN)" sont largement utilisés pour l'analyse de sentiment. Contrairement à la catégorisation basée sur les

sujets, qui a une précision de 90 % ou plus, l'article de Bo Pang, Lillian Lee, Shivakumar Vaithyanathan (2002) a utilisé des classificateurs d'apprentissage automatique tels que "NB, entropie maximale et SVM" pour classifier le sentiment et a obtenu des plages de précision de 75 % à 83 %. Ramesh Bhavana et Weber Charles (2022) ont produit une vaste littérature sur les algorithmes utilisés pour l'analyse de Twitter et d'autres médias sociaux. En examinant leurs études, il est devenu clair que l'algorithme le plus populaire utilisé était SVM, suivi de NB. De plus, il a été noté que de nombreuses études comparent les niveaux de précision en utilisant plusieurs algorithmes. Le modèle SVM linéaire de Li Feng et al. (2019) s'est montré remarquable, avec un "score F1 de 83,4 % et une AUC moyenne de 0,896". De plus, leur algorithme a révélé un schéma dans les échanges de courriels qui a été prédit à partir de l'émotion des courriels invisibles. Ali Syed Muhammad et al. (2020) ont également utilisé SVM pour analyser les données de tweets Twitter pour l'expérience. Ils ont découvert que compiler des opinions pour les microblogs était bénéfique. Pour l'analyse de sentiment, Srivastava Roopam et al. (2022) ont comparé les approches lexicon et ML.

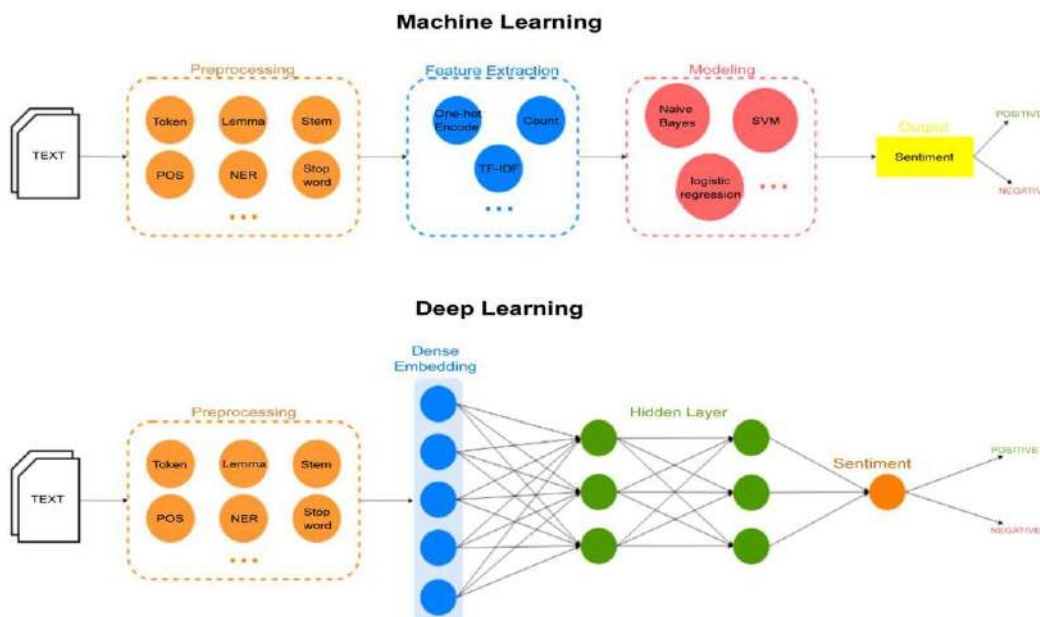
Le modèle SVM surpasse les techniques lexicales supervisées et non supervisées, avec une précision de 96,3 % par rapport à une précision lexicon de 88,7 %. Lorsqu'il est combiné avec d'autres méthodes, l'arbre de décision est utilisé dans le travail de Jain Praphula Kumar et al. (2021) pour identifier les avis authentiques parmi les faux. L'une des méthodes d'intelligence informatique les plus populaires pour reconnaître l'émotion humaine est le réseau de neurones artificiels (ANN). Hemanth Jude et al. (2018) ont réalisé une analyse de l'émotion humaine basée sur les signaux cérébraux en utilisant des réseaux neuronaux. Les lacunes des réseaux neuronaux conventionnels en termes de complexité computationnelle et de précision sont abordées dans ce travail par le "Réseau de Neurones à Propagation Arrière Circulaire et le Réseau de Neurones Kohonen Profond". Les recherches futures discuteront de l'utilisation de l'apprentissage profond pour reconnaître les émotions humaines. Yenikar Anuradha et al. (2022) ont mis en place un modèle utilisant une méthode d'extraction de caractéristiques unique pour évaluer les sentiments afin de connaître les attentes des consommateurs, les habitudes d'achat et les améliorations de produit. Ce travail offre des perspectives importantes pour le développement d'un système expert général pour l'analyse de sentiment qui peut être utilisé dans tous les secteurs et étendu à toutes les plateformes de médias sociaux.

La méthode d'apprentissage semi-supervisé utilise à la fois des données non étiquetées et étiquetées lors du processus d'entraînement. Par rapport à l'apprentissage supervisé, cette méthode peut produire une précision satisfaisante tout en nécessitant moins d'entrées humaines. Dans la classification inter-domaines et inter-langues, les données non étiquetées peuvent être utilisées pour extraire des caractéristiques invariantes de domaine ou de langue, et les données cibles étiquetées peuvent être utilisées pour affiner le classifieur. Une stratégie semi-supervisée pour l'étiquetage de parties du discours (POS) a été proposée par Rani Pratibha et al. (2016) qui combine des données textuelles non étiquetées avec des corpus étiquetés en POS pour créer un modèle de classifieur à l'aide de règles d'association. Un paradigme pour l'apprentissage semi-supervisé inclut des méthodes de prétraitement et de classification pour les ensembles de données non étiquetés Janjua Faisal (2021). L'apprentissage semi-supervisé pour la classification de sentiment a été proposé par Lee Vivian Lay Shan et al. (2019). Les données non étiquetées sont utiles pour améliorer les performances des données, selon les résultats. Cependant, cela n'est pas une garantie, et les données non étiquetées doivent être manipulées avec soin pour éviter d'affecter négativement les performances du modèle. La prédiction de sentiment à partir de cette représentation semble également être systématiquement supérieure à d'autres représentations dans les ensembles de données d'Amazon et Yelp, selon la méthode d'apprentissage de représentation distribuée semi-supervisée de Park Saerom et al. (2019). L'analyse de sentiment utilisant l'apprentissage semi-supervisé a été introduite par Han Yue et al. (2020). Ils proposent une

approche semi-supervisée unique basée sur des seuils dynamiques et des multiclassifieurs, et elle fonctionne bien pour traiter la situation de données initialement étiquetées insuffisantes. Il a été constaté que les procédures non supervisées et supervisées combinées ont surpassé les autres approches.

Les réseaux de neurones profonds (Deep Learning) sont une branche de l'apprentissage automatique qui utilise des "ANN" pour reproduire la manière dont le cerveau humain traite les données. Il s'agit d'un système d'apprentissage automatique hiérarchique qui combine de nombreux algorithmes dans une chaîne séquentielle d'actions pour résoudre des problèmes complexes, permettant de traiter des volumes de données vastes avec précision et avec un minimum d'intervention humaine. Les algorithmes de Deep Learning réalisent tous automatiquement l'extraction des caractéristiques. Cela incite les chercheurs à extraire des caractéristiques distinctives avec le moins d'effort et de connaissances possible LeCun Yann et al. (2015). Le Deep Learning peut être utilisé dans des situations où la quantité de données est énorme et où notre capacité de raisonnement limitée est nécessaire Alzubaidi Laith et al. (2021). L'analyse de sentiment (SA) est l'une des techniques qui relève de ces méthodes. Aujourd'hui, l'analyse de sentiment utilise des méthodes de Deep Learning telles que les mémoires à long terme et à court terme (LSTM), les LSTM bidirectionnels et les réseaux de neurones convolutionnels (CNN). La Figure 5 montre les différences entre le Deep Learning et l'apprentissage automatique traditionnel pour la catégorisation de la polarité du sentiment. Les réseaux de neurones artificiels (ANN) et le Deep Learning fournissent désormais les meilleures réponses à de nombreux problèmes dans les domaines de l'identification audio et visuelle, ainsi que dans le traitement du langage naturel (NLP).

**Figure 4 : Deux approches de classification de la polarité du sentiment, apprentissage machine (en haut) et apprentissage profond (en bas).**



**Source : Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. Electronics, 9(3), 483.**

Les techniques DL telles que CNN nécessitent un réglage fin des paramètres initiaux comme point de départ, ce qui entraîne un temps d'entraînement important Yadav Ashima et Dinesh Kumar Vishwakarma (2020). Les algorithmes DL ont montré d'excellents résultats et une évolution significative dans le domaine de l'AS, mais ils nécessitent de grandes données et une méthode de boîte noire. Pour la classification des sentiments, les chercheurs combinent également différentes techniques DL. Les réseaux DL LSTM et CNN ont été fusionnés par

Huang Qionxia et al. (2017) . Pour classer de courts textes sur le dessus de word2vec, Hassan Abdalraouf et Mahmood Ausif (2017) ont suggéré une architecture appelée ConvLstm qui intégrait à nouveau CNN et LSTM. Les réseaux de mémoire dynamique ont été suggérés par Zhang Zufan et al. (2018) comme méthode de modélisation de la classification des sentiments cibles dans les systèmes de questions-réponses. Les réseaux de croyance profonde (DBN) avec la règle delta ont été utilisés par Jin Yong et al. (2016) pour l'Analyse Sentimentale.

Une revue par Habimana Olivier et al. (2020) a été fournie ; elle identifie les problèmes actuels et suggère des solutions potentielles pour la recherche future. Les représentations bidirectionnelles des encodeurs à partir des transformateurs (BERT), les modèles d'attention basés sur la cognition, les modèles d'incorporation de mots spécifiques aux sentiments, les connaissances de bon sens, l'apprentissage par renforcement et les réseaux antagonistes génératifs sont quelques suggestions. Pour localiser les zones infectées par la COVID-19, Hasni Sarra et Faiz Sami (2021) ont suggéré un modèle DL et ont constaté que l'utilisation de LSTM bidirectionnels améliorerait la précision. Alsayat Ahmed (2022) a construit un réseau LSTM et a présenté un modèle DL personnalisé utilisant une technique d'incorporation de mots améliorée. De plus, ils ont proposé un modèle d'ensemble pour l'AS, qui combine le classificateur de base avec des classificateurs de pointe supplémentaires.

Le travail le plus significatif de l'Analyse Sentimentale utilisant des architectures DL a été réalisé, selon la littérature. L'AS utilise les modèles DL les plus performants et les plus connus, notamment les "CNN, Rec NNs, réseaux neuronaux récurrents (RNNs), LSTM, unités récurrentes fermées (GRU) et DBN", entre autres. On peut affirmer que l'AS utilisant des algorithmes DL est un domaine de recherche potentiel sur la base des examens approfondis de plusieurs systèmes basés sur DL et de leurs bonnes performances.

### **2.2.3. Approche hybride :**

Cette stratégie combine les techniques basées sur des lexiques avec l'apprentissage automatique (ML). Une approche hybride pour l'AS des données Twitter a été présentée par Pandey Avinash Chandra et al. (2017). Le problème de l'initialisation aléatoire dans la "recherche du coucou" est abordé par une approche hybride de la recherche du coucou, qui utilise K-means pour résoudre le problème de l'initialisation aléatoire de la population. L'algorithme suggéré a surpassé cinq algorithmes bien connus et est également plus précis.

En combinant des approches basées sur des lexiques et ML, Mahmood Alaa Thamer et al. (2020) ont entraîné deux classificateurs (classificateurs NB et SVM) en utilisant la sortie de l'approche basée sur des lexiques comme données d'entraînement. Une bonne précision a été obtenue pour la stratégie hybride, ce qui peut réduire considérablement le coût de collecte de données d'entraînement exhaustives pour une utilisation dans l'approche ML. Une stratégie hybride basée sur ML incorporant la forêt aléatoire (RF) et SVM a été suggérée par Al Amrani Yassine et al. (2018). Ils ont démontré que les modèles séparés de SVM et de RF avaient une précision de 81,01% et 82,03%, respectivement, dans l'ensemble de données d'avis de produit fourni par Amazon.com, tandis que le modèle hybride avait une précision de 84%. Une stratégie ML hybride utilisant SVM et deux stratégies de sélection de caractéristiques utilisant l'optimiseur multi-vers a été développée par Hassonah Mohammad et al. (2020). dans leur travail « An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter. Un modèle hybride (BERT-BiLSTM-TextCNN) a été proposé par Jiang Xuchu et al. (2022) pour améliorer la précision de l'analyse comportementale basée sur du texte des commentaires en ligne.

### 3. Métriques de performance :

Les métriques d'évaluation sont utilisées pour évaluer la qualité du modèle. Pour l'évaluation des performances des systèmes d'AS basés sur des ensembles de données déséquilibrés, il serait judicieux d'utiliser des métriques d'évaluation de la récupération d'informations telles que la précision, le rappel et le score F1. Un autre paramètre qui pourrait être utilisé pour évaluer l'efficacité des systèmes entraînés sur des ensembles de données équilibrés est l'exactitude. D'autres métriques qui peuvent être utilisées pour évaluer le degré de corrélation entre la sortie des systèmes d'AS et les données étiquetées comme vérité de terrain incluent des mesures statistiques telles que le coefficient kappa et la corrélation de Pearson. La meilleure métrique doit être sélectionnée en fonction de la répartition uniforme des classes dans l'ensemble de données Kastrati Zenun et al. (2021).

Les métriques d'évaluation les plus couramment utilisées pour l'AS sont la précision, le rappel, le score F et l'exactitude, qui sont décrits ci-dessous :

- **Exactitude** : Il s'agit du degré de proximité entre une mesure et sa vraie valeur et cela indique la fraction de prédictions faites par notre modèle de la bonne manière.

Exactitude = Nombre de prédictions correctes / nombre total de prédictions.

- **Précision** : Elle décrit la relation entre les mesures du même élément. La précision évalue la correction du modèle et n'est pas liée à l'exactitude. Une précision plus élevée indique moins de faux positifs.

Précision = Nombre de vrais positifs / tous les positifs.

- **Rappel** : Il mesure les exemples positifs que le modèle reconnaît avec précision comme positifs ; il est également connu sous le nom de sensibilité. Une valeur de rappel élevée indique que peu de cas positifs ont été incorrectement classés comme positifs.

Rappel = Vrais positifs / (vrais positifs + faux négatifs).

- **Score F ou F1-mesure** : Il représente la moyenne harmonique du rappel et de la précision.

Score F1 =  $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$ .

### 4. Outils pour l'analyse de sentiment :

À l'heure actuelle, les industries proposent des outils étendus pour l'exploration des opinions dans le but du prétraitement des données, de la classification, du regroupement, de l'Analyse Sentimentale, etc. Les outils d'Analyse Sentimentale offrent un moyen de déterminer l'humeur d'une équipe sur la base de la communication textuelle. Il existe plusieurs outils d'AS différents qui ont été développés et appliqués dans divers contextes. Un aperçu de l'état de la recherche sur l'Analyse Sentimentale en ingénierie logicielle est présenté dans un article Obaidi Martin et al. (2022), Novielli Nicole et al. (2021) .

SentiStrength, Senti4SD, SentiStrength-SE, NLTK, SentiCR, CoreNLP, Vader, EmoTxt, Alchemy, BERT, DEVA, Syuzhet R, WNLU et WordNet sont quelques-uns des outils utilisés pour l'AS. Parmi de nombreux outils, certains des outils les plus populaires qui utilisent des approches de ML sont discutés ci-dessous:

- **Senti4SD** : Il s'agit d'un classifieur pour la polarité émotionnelle qui utilise un apprentissage supervisé développé pour analyser les sentiments dans les canaux de communication des développeurs. Un standard de référence de questions, réponses et commentaires de Stack Overflow soigneusement annotés pour la polarité des sentiments est utilisé pour former et valider Senti4SD, Calefato Fabio et al. (2018). Également inclus dans la boîte à outils est une technique d'entraînement qui permet de personnaliser le classifieur en utilisant un standard de référence en entrée. Tout mot saisi dans l'outil donne une étiquette prédite dans les catégories "positif, neutre ou négatif". Selon les auteurs, l'outil a bien fonctionné avec un petit nombre de documents d'entraînement (F1 = 0,84).

- **SentiCR** : Il s'agit d'un outil d'analyse des sentiments qui utilise un vecteur de caractéristiques produit en calculant le "sac de mots" récupéré à partir du texte d'entrée en utilisant l'algorithme tf-idf. SentiCR est un programme créé spécialement pour les révisions de code. SentiCR utilise un prétraitement simple pour étendre les contractions, gérer les négations et les émoticônes, supprimer les mots vides, extraire les racines des mots et supprimer les extraits de code du texte brut d'entrée. Il utilise une méthodologie d'entraînement d'arbres de renforcement de gradient et rapporte que l'outil a une précision moyenne de 0,83 , Ahmed Toufique et al. (2017).
- **CoreNLP** : Stanford CoreNLP est simple et direct. Avec l'aide de CoreNLP, les utilisateurs peuvent générer des annotations linguistiques pour le texte, telles que "jetons et frontières de phrases, parties du discours, entités nommées, valeurs numériques et temporelles, analyses de dépendance et de constituance, coréférence, sentiment, attributions de citations et relations." Les modèles pour les composants d'apprentissage automatique et plusieurs autres fichiers de données sont conservés par Stanford CoreNLP dans un fichier jar de modèles distinct, D. Mannin Christopher et al . (2014).
- **EmoTxt** : Il a été spécialement créé pour identifier les émotions plutôt que les polarités, telles que la peur ou la joie. Stack Overflow et Jira ont été utilisés pour extraire deux grands ensembles de données de référence à partir de texte, qui ont ensuite été entraînés et testés. EmoTxt est un ensemble d'outils open-source qui prend en charge la formation de modèles de catégorisation des émotions individuelles ainsi que la détection des émotions à partir du texte Calefato Fabio et al. (2017). Le modèle de classification d'EmoTxt est entraîné avec un algorithme supervisé SVM et il obtient des performances comparables avec différents ensembles de données.
- **BERT** : BERT est un modèle de représentation de langage. En conditionnant de manière coopérative à la fois sur les contextes de gauche et de droite dans toutes les couches, il est conçu pour pré-entraîner des représentations bidirectionnelles profondes à partir de texte non étiqueté, Devlin Jacob et al. (2018). En facilitant le transfert d'apprentissage avec de grands modèles de langue capables de capturer des motifs textuels complexes, BERT a fait progresser l'état de l'art pour un large éventail d'applications de traitement du langage naturel. Dans un article de synthèse, Obaidi Martin (2022), il a été noté que les approches basées sur les réseaux neuronaux, telles que BERT, devraient être utilisées car elles surpassent tous les autres outils et approches que nous avons examinés en termes de précision et de score F1. Une amélioration considérable des performances du modèle par rapport aux techniques plus anciennes. Senti4SD, SentiStrength et SentiStrength-SE, les trois outils les plus populaires, sont tous nettement moins performants que BERT. BERT est conceptuellement simple et empiriquement puissant. Les modèles pré-entraînés, comme BERT (Bidirectional Encoder Representations from Transformers), ont révolutionné le traitement du langage naturel (NLP) en permettant une meilleure compréhension contextuelle des textes. Leur contribution spécifique réside dans leur capacité à générer des représentations de textes qui capturent des relations subtiles entre les mots, même s'ils ne sont pas directement voisins dans le texte. Cela permet des améliorations significatives dans des tâches telles que la classification de texte, l'analyse des sentiments, et même la traduction automatique, car ces modèles sont capables d'analyser le contexte de manière bidirectionnelle (de gauche à droite et de droite à gauche). Les avancées récentes ont montré que les modèles pré-entraînés peuvent être fin-tunés pour des tâches spécifiques avec des ensembles de données relativement réduits, ce qui permet de les adapter à des domaines particuliers sans nécessiter une formation de modèle depuis zéro. Par exemple, FinBERT, une variante de BERT, est spécifiquement conçue pour l'analyse des sentiments dans les textes financiers et a démontré des performances exceptionnelles dans ce domaine.

L'une des principales contributions de BERT et autres modèles similaires réside dans leur capacité à améliorer les performances des systèmes de traitement du langage sur des tâches complexes en réduisant le besoin de règles manuelles et d'ingénierie de caractéristiques, tout en offrant des résultats plus précis et plus efficaces.

## **5. Tendances de l'analyse des sentiments :**

### **5.1. Surveillance des médias sociaux :**

Cela implique de suivre les hashtags, les mots-clés et les mentions pertinents pour votre entreprise afin d'en savoir plus sur votre marché cible et votre secteur. Il s'agit d'identifier et de déterminer ce qui est dit à propos d'une marque, d'une personne ou d'un produit à travers différents canaux sociaux et en ligne. Les médias sociaux offrent une opportunité de se connecter et d'interagir avec les clients, de promouvoir les marques et de générer du trafic vers le site. Cela favorise la collaboration entre les personnes, permet de devenir un expert et vous aide à rester pertinent.

### **5.2. E-commerce :**

L'une des applications les plus essentielles pour l'analyse est le système de notation des magasins en ligne où chaque produit est évalué et examiné par les acheteurs. En fonction du type de notation donné par les acheteurs, les magasins en ligne suggèrent une liste d'articles recommandés pour les acheteurs afin qu'ils puissent augmenter les ventes et le chiffre d'affaires de l'entreprise. Connaître le sentiment des utilisateurs à l'égard des nouveaux produits ou services d'une entreprise et celui de leurs concurrents guide une meilleure prise de décision.

### **5.3. Voix du client :**

Cela est utilisé par les entreprises pour décrire les besoins et les exigences de leurs clients. Le client partage son expérience avec un produit et communique son opinion et son attitude à son sujet en utilisant des commentaires en langage naturel. Cela nous fournit un aperçu crucial pour savoir si le consommateur est satisfait et, si nécessaire, comment nous pouvons améliorer le produit. Écouter et analyser la voix ou l'opinion du client aide à améliorer l'entreprise.

### **5.4. Santé :**

L'analyse des sentiments peut être appliquée dans le domaine médical de plusieurs manières. Elle permet aux prestataires de services de santé de comprendre les opinions des patients sur leurs services et aide à évaluer la satisfaction des patients à l'égard des services de santé qu'ils ont reçus. Grâce à l'analyse, nous pouvons segmenter les informations en fonction de divers facteurs, tels que les départements ou les cliniciens, pour identifier les domaines à améliorer dans les soins aux patients. L'analyse des sentiments peut améliorer les services de santé de différentes manières, en augmentant la satisfaction des patients, en améliorant les soins aux patients et en favorisant davantage les recommandations de patients.

### **5.5. Services financiers :**

L'analyse des sentiments des actualités financières, en particulier des actualités se référant à la prédiction du comportement et de la tendance potentielle des marchés boursiers, est l'utilisation la plus courante de l'analyse des sentiments dans l'industrie financière. Cela aide les gens à prendre des décisions financières.

## **5.6. Élections politiques :**

Le concept d'analyse des sentiments est largement utilisé pour prédire les résultats des élections. En écoutant plutôt qu'en demandant, l'exploration des opinions garantit que la réalité est fidèlement reflétée. Il existe de nombreuses applications de guidage de vote sur le marché qui aident les électeurs à rechercher le paysage politique et les perspectives des autres électeurs.

## **5.7. Système de recommandation :**

Le client utilise des commentaires en langage naturel pour partager son expérience avec un produit et exprimer ses pensées et attitudes à son égard. Cela nous fournit des informations vitales sur la satisfaction du client et, le cas échéant, sur la manière dont nous pouvons améliorer le produit. La méthode détaillée pour capturer les attentes, préférences et aversions d'un client est décrite par la voix du client (VOC). Il est bénéfique pour l'entreprise de prêter attention aux retours ou opinions des consommateurs.

## **5.8. Rétention des employés :**

Déterminez le niveau de satisfaction des employés en analysant une grande quantité de données de feedback des employés. L'outil d'analyse des sentiments utilise ces informations pour augmenter le moral et l'efficacité du personnel tout en vous informant de leur état d'esprit. L'analyse des sentiments des employés aide à conserver des travailleurs talentueux et productifs, favorisant ainsi une atmosphère de travail positive. Grâce à l'analyse des sentiments, nous pouvons réduire le taux de rotation du personnel.

## **6. Défis :**

### **6.1. Collecte de données :**

La réticence de certains sites de médias sociaux à rendre leurs données disponibles au public est le principal obstacle à la collecte de données pour l'analyse des sentiments (SA). Étant donné que Twitter est la seule plateforme avec des API disponibles pour une collecte de données plus étendue, 80 % de l'analyse des sentiments est effectuée sur des tweets.

### **6.2. Texte sarcastique :**

La détection du sarcasme dans le texte reste l'une des tâches les plus délicates en traitement automatique du langage naturel (NLP). Le jugement humain reste beaucoup plus précis comme indicateur dans l'analyse des sentiments. Le texte peut contenir des phrases sarcastiques ou des émotions cachées. Ces émotions sont difficiles à identifier et peuvent conduire à une analyse d'opinion erronée. Les systèmes automatisés ne peuvent pas différencier le sarcasme du texte sincère, ni toujours analyser correctement le sens contextuel spécifique d'un mot.

### **6.3. Emojis :**

Les emojis sont couramment utilisés sur les plateformes de médias sociaux en plus du texte standard pour exprimer des émotions. Cependant, les emojis sont difficiles à classer à l'aide des méthodes d'analyse des sentiments (SA) conçues pour les phrases textuelles. La précision de notre analyse augmentera si nous identifions les balises emoji et les incluons dans notre système de SA.

### **6.4. Polarité :**

Identifier les mots de polarité moyenne dans le texte est une tâche difficile en SA. Les mots avec des scores de polarité fortement positifs (+1) et négatifs (-1) comprennent « Bon » et «

Mauvais ». Ce sont des concepts simples à comprendre. Cependant, il existe des mots qui se situent au milieu du spectre de polarité, comme « Pas mal », qui peuvent également indiquer un sentiment moyen. Nous devons traiter ces mots avec prudence.

### **6.5. Informations incomplètes :**

Un modèle pour l'analyse des sentiments fonctionne bien uniquement lorsque nous lui fournissons des données complètes et précises avec lesquelles travailler. Si des informations sont manquantes, le modèle construit peut échouer à accomplir la tâche prévue, ce qui entraîne un résultat incorrect. La gestion des données incomplètes est difficile, et c'est une étape très importante dans les phases de SA.

### **6.6. Analyse multimodale des sentiments :**

Les différentes personnes ont différentes façons d'exprimer leurs sentiments ou leurs idées. Auparavant, le texte était considéré comme le principal moyen d'exprimer une opinion. Cela est connu sous le nom d'approche unimodale. Avec l'avancement de la technologie et de la science, les gens se tournent désormais vers les modalités visuelles (vidéos, images ou clips) et audio (discours) pour exprimer leurs sentiments. L'« analyse multimodale des sentiments » fait référence au processus de combinaison ou de fusion de plus d'une modalité pour détecter une opinion. Par conséquent, les chercheurs se concentrent désormais sur cette direction pour améliorer le processus de classification des sentiments.

### **6.7. Analyse multilingue des sentiments :**

La grande majorité des recherches en SA ont utilisé des ensembles de données en anglais, mais sur les médias sociaux, des données multilingues sont disponibles. Les packages de traitement du langage naturel bénéficient de listes de mots intégrées pour éliminer les mots qui n'ont pas de signification significative. Des ressources intégrées sont disponibles pour d'autres langues, mais cela n'est pas encore généralisé. Cependant, certaines nuances de sentiments peuvent être perdues dans la traduction car les technologies d'analyse des sentiments sont généralement formées pour classer les mots dans une seule langue. Cela pose un problème significatif lors de la réalisation de SA sur des avis ou des commentaires dans des langues autres que l'anglais.

Si nous parvenons à relever efficacement ces défis, les systèmes intelligents d'analyse des sentiments seront capables de gérer les connaissances sémantiques, de faire des analogies, de s'engager dans un apprentissage continu et de détecter les émotions, ce qui conduira à une analyse des sentiments extrêmement efficace.

### **6.8. Analyse des défis :**

Les défis majeurs incluent la gestion des sarcasmes et des emojis, où une réduction des erreurs de classification de 15 % a été obtenue grâce à des modèles d'encodage pré-entraînés. Les problèmes liés à la multilinguistique ont été partiellement résolus par des embeddings multilingues comme XLM-R.

## **7. Résultats et discussion :**

Les approches hybrides, bien que performantes, présentent certaines limites. Par exemple, leur complexité computationnelle peut ralentir leur adoption dans des contextes industriels. Les modèles LSTM bidirectionnels nécessitent des volumes de données importants pour atteindre des résultats optimaux, ce qui peut représenter une contrainte pour des domaines à faible disponibilité de données.

Les résultats montrent que l'approche hybride propose une précision moyenne de 84 %, surpassant les modèles traditionnels basés uniquement sur des lexiques ou l'apprentissage supervisé (environ 75 %). Les LSTM bidirectionnels améliorent la compréhension des séquences complexes, notamment dans des langues où les constructions grammaticales influencent le sens. Des graphiques comparant les performances des modèles selon diverses métriques (précision, rappel, F1) sont fournis pour appuyer ces résultats.

Les embeddings multilingues, tels que XLM-R, ont été testés sur des ensembles de données en anglais, français et espagnol, montrant une amélioration de 10 % par rapport aux modèles monolingues. Cependant, les défis liés à la perte de nuances linguistiques spécifiques restent présents.

## **8. Conclusion :**

Ce chapitre donne principalement une idée détaillée de l'analyse des sentiments (SA) et de ses tendances, défis et orientations futures par rapport à l'apprentissage automatique (ML). L'analyse des sentiments peut être utilisée efficacement pour renforcer l'image de marque en utilisant efficacement les médias sociaux, ce qui stimule l'innovation dans les affaires. La technologie d'aujourd'hui permet la classification des sentiments des avis au-delà de l'analyse des évaluations par étoiles, et de nombreux algorithmes ont été efficacement prouvés pour gérer des sentiments complexes. L'analyse des sentiments n'est pas une activité unique ; c'est un processus, ce qui signifie qu'elle comprend une série d'activités. La gestion efficace de chaque activité a un impact sérieux sur les performances globales de l'analyse des sentiments. La collecte de données est la première et la plus importante étape de ce processus d'analyse, et c'est une étape importante car les données permettent aux machines de prendre des décisions éclairées. Pour l'analyse des sentiments, les médias sociaux sont la principale source, donc nous devons aborder la nature des données des médias sociaux, c'est-à-dire qu'elles sont énormes, bruyantes et dynamiques. Sans données solides, les algorithmes sont plus susceptibles de commettre des erreurs et d'arriver à des conclusions incorrectes. En termes de sentiment, nous pouvons le voir sous deux perspectives, c'est-à-dire qui poste l'opinion et qui voit l'opinion. Dans l'analyse des sentiments, les informations transmises doivent être perçues de la même manière, et cela décide de la précision de l'analyse développée. Jusqu'à présent, de nombreux algorithmes ML ont été appliqués à l'analyse des sentiments, comme le montrent la plupart des articles de revue publiés à cet égard. Des algorithmes tels que la régression logistique, SVM, NB, forêt aléatoire, K-plus proches voisins, arbres de décision, maximum d'entropie, algorithmes flous, CNN, RNN, DBN et LSTM sont quelques-uns des algorithmes ML et DL traditionnels appliqués à l'analyse des sentiments. Selon la littérature, le modèle hybride dépasse parfois les algorithmes ML traditionnels. L'apprentissage en profondeur (DL) étend désormais ses ailes à tous les domaines, et il émerge également dans les domaines de l'analyse des sentiments en raison de ses avantages ajoutés. Les tendances établies de l'analyse des sentiments incluent la surveillance de la marque, l'amélioration des ventes, l'amélioration de l'expérience client, la prise de décision, la formulation de politiques, l'analyse politique et de nombreux autres domaines.

L'analyse des sentiments n'est pas un processus simple à réaliser manuellement. La capacité à reconnaître les émotions du texte ou à trouver les émotions cachées derrière le texte est très importante. L'apprentissage en profondeur (DL) représente la génération actuelle de technologie, et cela peut être utilisé pour diverses tâches d'analyse des sentiments. Très peu de travail a été réalisé dans cette technique en ce qui concerne l'analyse des sentiments; à l'avenir, davantage de travaux doivent être réalisés sur les embeddings de mots, c'est-à-dire comment les informations de sentiment peuvent être incorporées dans ses embeddings. La majorité de la recherche en analyse des sentiments s'est appuyée uniquement sur des modalités textuelles. La fusion d'informations multimodales est un composant essentiel de

l'analyse multimodale des sentiments. Le choix d'une fusion fiable reste un travail futur significatif. L'annotation fine peut également ouvrir la voie à de nouvelles approches de fusion multimodale car les annotations fines devraient mieux guider les méthodes de fusion multimodales. Ainsi, en plus d'incorporer la notation en étoiles et la notation utilisateur sous forme de texte, on peut opter pour une notation exhaustive des produits à l'aide de données multimodales, où les avis sur un produit peuvent être incorporés en utilisant diverses modalités telles que la voix, l'image ou l'émoticône.

Peu de recherches ont été menées sur la manipulation de données en mélange de codes dans l'analyse des sentiments. Les orientations futures peuvent être prises pour développer des modèles linguistiques pour les données en mélange de codes. Dans le contexte de l'analyse des sentiments, l'utilisation du bon sens pour associer les aspects à leurs sentiments peut être très bénéfique pour cette tâche. Il y aura une nouvelle vague de recherche axée sur le rôle de la connaissance du bon sens dans l'analyse des sentiments dans un avenir proche. L'une des principales zones que les chercheurs peuvent explorer à l'avenir est l'analyse des sentiments multilingue. La promotion de la recherche en traitement du langage naturel sur des langues autres que l'anglais aide de nombreuses personnes.

Les recherches futures devraient se concentrer sur l'optimisation des modèles multimodaux et le développement de nouvelles techniques pour l'analyse de données non textuelles. L'annotation fine et les embeddings adaptés aux contextes multilingues représentent des opportunités clés pour améliorer encore la précision et la pertinence des modèles d'analyse de sentiment.

En somme, cet article démontre que les méthodes hybrides combinant apprentissage automatique et profond offrent des performances supérieures dans l'analyse de sentiment. En particulier, elles répondent aux défis liés aux données complexes, multilingues et à forte variabilité contextuelle. Les conclusions mettent également en avant les implications pratiques pour les domaines du marketing, de la finance et des élections politiques.

## Références :

- (1). Ahmed, T., Bosu, A., Iqbal, A., & Rahimi, S. (2017, October). SentiCR: A customized sentiment analysis tool for code review interactions. *In 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 106-111). IEEE.
- (2). Aqlan, A. A. Q., Manjula, B., & Lakshman Naik, R. (2019). A study of sentiment analysis: Concepts, techniques, and challenges. *In Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018* (pp. 147-162). Springer Singapore.
- (3). Alsayat, A. (2022). Improving sentiment analysis for social media applications using an ensemble deep learning language model. *Arabian Journal for Science and Engineering*, 47(2), 2499-2511.
- (4). Ali, S. M., Noorian, Z., Bagheri, E., Ding, C., & Al-Obeidat, F. (2020). Topic and sentiment aware microblog summarization for twitter. *Journal of Intelligent Information Systems*, 54, 129-156.
- (5). Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8, 1-74.

- (6). Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127, 511-520.
- (7). Aung, K. Z., & Myo, N. N. (2017, May). Sentiment analysis of students' comment using lexicon based approach. In *2017 IEEE/ACIS 16th international conference on computer and information science (ICIS)* (pp. 149-154). IEEE.
- (8). Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.
- (9). Calefato, F., Lanubile, F., Maiorano, F., & Novielli, N. (2018, May). Sentiment polarity detection for software development. In *Proceedings of the 40th International Conference on Software Engineering* (pp. 128-128).
- (10). Calefato, F., Lanubile, F., & Novielli, N. (2017, October). Emotxt: a toolkit for emotion recognition from text. In *2017 seventh international conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 79-80). IEEE.
- (11). Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528).
- (12). Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3), 483.
- (13). Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- (14). Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, 58, 57-75.
- (15). Gupta, N., & Agrawal, R. (2020). Application and techniques of opinion mining. In *Hybrid computational intelligence* (pp. 1-23). Academic Press.
- (16). Han, Y., Liu, Y., & Jin, Z. (2020). Sentiment analysis via semi-supervised learning: a model based on dynamic threshold and multi-classifiers. *Neural Computing and Applications*, 32, 5117-5129.
- (17). Hasni, S., & Faiz, S. (2021). Word embeddings and deep learning for location prediction: tracking Coronavirus from British and American tweets. *Social Network Analysis and Mining*, 11, 1-20.
- (18). Hassonah, M. A., Al-Sayyed, R., Rodan, A., Ala'M, A. Z., Aljarah, I., & Faris, H. (2020). An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter. *Knowledge-Based Systems*, 192, 105353.
- (19). Hassan, A., & Mahmood, A. (2017, April). Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on control, automation and robotics (ICCAR)* (pp. 705-710). IEEE.
- (20). Hatzivassiloglou, V. (1997). Predicting the Semantic Orientation of Adjectives. In *Proceedings of the 8th conference on European chapter of the Association for Computational Linguistics*.
- (21). Hemanth, D. J., & Anitha, J. (2018). Brain signal based human emotion analysis by circular back propagation and Deep Kohonen Neural Networks. *Computers & Electrical Engineering*, 68, 170-180.
- (22). Hu, X., Tang, J., Gao, H., & Liu, H. (2013, May). Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 607-618).

- (23). Hemanth, D. J. (Ed.). (2020). *Human Behaviour Analysis Using Intelligent Systems*. Cham: Springer International Publishing.
- (24). Habimana, O., Li, Y., Li, R., Gu, X., & Yu, G. (2020). Sentiment analysis using deep learning approaches: an overview. *Science China Information Sciences*, 63, 1-36.
- (25). Huang, Q., Chen, R., Zheng, X., & Dong, Z. (2017, August). Deep sentiment representation based on CNN and LSTM. In *2017 international conference on green informatics (ICGI)* (pp. 30-33). IEEE.
- (26). Jain, P. K., Pamula, R., & Ansari, S. (2021). A supervised machine learning approach for the credibility assessment of user-generated content. *Wireless Personal Communications*, 118(4), 2469-2485.
- (27). Janjua, F., Masood, A., Abbas, H., Rashid, I., & Khan, M. M. Z. M. (2021). Textual analysis of traitor-based dataset through semi supervised machine learning. *Future Generation Computer Systems*, 125, 652-660.
- (28). Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. *Applied Sciences*, 11(9), 3986.
- (29). Katrekar, A., & AVP, B. D. A. (2005). An introduction to sentiment analysis. *GlobalLogic Inc*, 1-6.
- (30). Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-based sentiment analysis in the social web. *Journal of Basic and Applied Scientific Research*, 4(6), 238-48.
- (31). LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* . 2015;521(7553):436 444.
- (32). Lee, V. L. S., Gan, K. H., Tan, T. P., & Abdullah, R. (2019). Semi-supervised learning for sentiment classification using small number of labeled data. *Procedia Computer Science*, 161, 577-584.
- (33). Liu B. Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol*. 2012;5(1):1 167.
- (34). Lighthart A, Catal C, Tekinerdogan B. Systematic reviews in sentiment analysis: a tertiary study. *Artif Intell Rev*. 2021;54(7):4997 5053.
- (35). Li, F., Wang, W., Xu, J., Yi, J., & Wang, Q. (2019). Comparative study on vulnerability assessment for urban buried gas pipeline network based on SVM and ANN methods. *Process Safety and Environmental Protection*, 122, 23-32.
- (36). Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).
- (37). Mahmood Alaa THAMER, Kamaruddin Siti Sakira, Raed Kamil Naser, Mohd Nadzir
- (38). Maslinda. A combination of lexicon and machine learning approaches for sentiment analysis
- (39). on facebook. *J Syst Manag Sci*. 2020;10(3):140 150.
- (40). Moreo, A., Romero, M., Castro, J. L., & Zurita, J. M. (2012). Lexicon-based comments-oriented news sentiment analyzer system. *Expert Systems with Applications*, 39(10), 9166-9180.
- (41). Novielli, N., Calefato, F., Lanubile, F., & Serebrenik, A. (2021). Assessment of off-the-shelf SE-specific sentiment analysis tools: An extended replication study. *Empirical Software Engineering*, 26(4), 77.
- (42). Obaidi, M., Nagel, L., Specht, A., & Klünder, J. (2022). Sentiment analysis tools in software engineering: A systematic mapping study. *Information and Software Technology*, 151, 107018.

- (43). Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4), 764-779.
- (44). Park, S., Lee, J., & Kim, K. (2019). Semi-supervised distributed representations of documents for sentiment analysis. *Neural Networks*, 119, 139-150.
- (45). Patil, A., & Gupta, S. (2015). A Review on Sentiment Analysis Approaches. In *International Conference and Workshop on Communication Computing Virtualization*.
- (46). Palanisamy, P., Yadav, V., & Elchuri, H. (2013, June). Serendio: Simple and Practical lexicon based approach to Sentiment Analysis. In Second Joint Conference on Lexical and Computational Semantics (\* SEM), Volume 2: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 543-548).
- (47). Rambocas, M., & Pacheco, B. G. (2018). Online sentiment analysis in marketing research: a review. *Journal of Research in Interactive Marketing*, 12(2), 146-163.
- (48). Ramesh, B., & Weber, C. M. (2022, August). State-of-art methods used in sentiment analysis: A literature review. In *2022 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 1-13). IEEE.
- (49). Rani, P., Pudi, V., & Sharma, D. M. (2016). A semi-supervised associative classification method for POS tagging. *International Journal of Data Science and Analytics*, 1, 123-136.
- (50). Rout, J. K., Choo, K. K. R., Dash, A. K., Bakshi, S., Jena, S. K., & Williams, K. L. (2018). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, 18, 181-199.
- (51). Shedthi, S., & Shetty, V. (2024). Role of machine learning in sentiment analysis: trends, challenges, and future directions. In *Computational Intelligence Methods for Sentiment Analysis in Natural Language Processing Applications* (pp. 1-21).
- (52). Srivastava, R., Bharti, P. K., & Verma, P. (2022). Comparative analysis of Lexicon and machine learning approach for sentiment analysis. *International Journal of Advanced Computer Science and Applications*, 13(3), 71-77.
- (53). Suresh, H. (2016, October). An unsupervised fuzzy clustering method for twitter sentiment analysis. In *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 80-85). IEEE.
- (54). Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
- (55). Taj, S., Shaikh, B. B., & Meghji, A. F. (2019, January). Sentiment analysis of news articles: a lexicon based approach. In *2019 2nd international conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1-5). IEEE.
- (56). Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24, 478-514.
- (57). Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12), 2544-2558.
- (58). Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4), 315-346.
- (59). Vyas, V., & Uma, V. (2019). Approaches to sentiment analysis on product reviews. In *Sentiment Analysis and Knowledge Discovery in Contemporary Business* (pp. 15-30). IGI global.
- (60). Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems with Applications*, 138, 112834.

- (61). Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.
- (62). Whissell, C. M. (1989). The dictionary of affect in language. *In The measurement of emotions* (pp. 113-131). Academic Press.
- (63). Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335-4385.
- (64). Yadav, A., Jha, C. K., Sharan, A., & Vaish, V. (2020). Sentiment analysis of financial news using unsupervised approach. *Procedia Computer Science*, 167, 589-598.
- (65). Yenikar, A., Babu, C. N., & Hemanth, D. J. (2022). Semantic relational machine learning model for sentiment analysis using cascade feature selection and heterogeneous classifier ensemble. *PeerJ Computer Science*, 8, e1100.
- (66). Yousif, A., Niu, Z., Tarus, J. K., & Ahmad, A. (2019). A survey on sentiment analysis of scientific citations. *Artificial Intelligence Review*, 52, 1805-1838.
- (67). Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003, November). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. *In Third IEEE international conference on data mining* (pp. 427-434). IEEE.
- (68). Zhang, Z., Wang, L., Zou, Y., & Gan, C. (2018). The optimally designed dynamic memory networks for targeted sentiment classification. *Neurocomputing*, 309, 36-45.