

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 Arréridj
Faculté des Mathématiques et d'Informatique
Département d'informatique



MEMOIRE

Présenté en vue de l'obtention du diplôme

Master en informatique

Spécialité : Ajoutez votre spécialité ici

THEME

La Classification Automatique du Feedback Client et son impact sur la Prise de Décision Stratégique dans le E-commerce et le Marketing Digital

Présenté par :
DAOUD Nadjem eddine

Soutenu publiquement le : 14/06/2025
Devant le jury composé de:
Président : Dr.FARES Nour El Houda
Examineur : Dr.BENABID Sonia
Encadreur :Dr.BELAYADI Yahia

2024/2025

Dédicace

Chère famille,

C'est avec une immense gratitude que je vous dédie ce travail. Votre soutien indéfectible a été le pilier de mon parcours académique. Chaque encouragement, chaque sacrifice, chaque mot de réconfort a été une source d'inspiration et de force pour surmonter les défis. Merci d'avoir toujours cru en moi, d'avoir été présents à chaque étape, et d'avoir fait de mes rêves les vôtres. Cette réussite est aussi la vôtre.

À tous mes professeurs,

Je ne saurais citer chacun de vos noms tant vous avez été nombreux à jalonner mon parcours. Cependant, je tiens à vous exprimer ma plus sincère reconnaissance. Chacun d'entre vous a contribué, à sa manière, à forger mes connaissances, à stimuler ma curiosité et à développer ma pensée critique. Votre passion pour l'enseignement et votre dévouement ont été exemplaires. Merci pour votre patience, votre guidance et pour toutes les leçons, académiques et humaines, que vous m'avez transmises. Vous avez semé les graines de ce que je suis aujourd'hui.

Remerciement

Je tiens à exprimer ma profonde gratitude à ma famille pour leur soutien inconditionnel tout au long de ce parcours académique. Un immense merci également à tous mes professeurs dont l'enseignement a été précieux. Merci à ma famille pour son soutien indéfectible et à tous mes professeurs pour leur précieux enseignement.

Résumé

Cette recherche étudie l'impact de la classification automatique des retours clients sur la prise de décision stratégique et opérationnelle dans le e-commerce et le marketing digital. Face au volume croissant des commentaires, l'analyse manuelle devient inefficace. L'étude propose donc d'automatiser cette tâche grâce aux techniques de Traitement du Langage Naturel (TALN).

L'objectif est de développer un système capable de classifier les commentaires clients en temps réel, selon leur tonalité (positive, négative, neutre) ou leur contenu (frustration, satisfaction, description). Cela permet de transformer des données brutes en informations exploitables, utiles pour la stratégie d'entreprise.

Le cadre théorique insiste sur le rôle clé de l'information dans la réduction de l'incertitude. La classification automatique rend les textes non structurés quantifiables grâce à des techniques comme le prétraitement du texte (nettoyage, lemmatisation, vectorisation) et des modèles de classification (lexiques, apprentissage automatique, réseaux neuronaux, modèles hybrides).

L'étude utilise des données issues de commentaires Amazon et privilégie le modèle VADER, bien adapté aux textes courts et informels. Ce modèle combine un lexique validé humainement et des règles grammaticales pour une détection fine des sentiments, avec des performances remarquables

Enfin, la classification automatique permet une meilleure compréhension des clients, révèle tendances et signaux faibles, et segmente les utilisateurs selon leurs sentiments. Ces analyses influencent les décisions stratégiques (produits, marketing, ventes) et opérationnelles

Text minig, opinion mining, text analysis, classification des textes

Abstract

This research looks at how using automated tools to sort and understand customer feedback can help companies make better decisions in online shopping (e-commerce) and digital marketing. Since there are so many customer reviews, it's hard for people to read and understand them all. To solve this, the study uses Natural Language Processing (NLP) to automatically analyze and sort comments by emotion, like frustration, satisfaction, or just a description. It also checks whether the feedback is positive, negative, or neutral.

The study explains that turning messy text into clear information helps companies make smarter choices. While customer reviews are full of useful ideas, they're often hard to understand because of their large number and complexity. Automatic text analysis helps make this easier by turning those words into useful data.

The research also goes over some key NLP steps, like cleaning the text, removing unimportant words, and changing words into their base forms. It talks about ways to turn text into numbers using methods like Bag-of-Words, and it covers different ways to classify text — using dictionaries, machine learning, deep learning, and combinations of these.

The study uses real data from Amazon reviews, and it focuses on a tool called VADER, which is really good at understanding short, casual messages like those on social media. VADER can handle slang, emojis, and even punctuation to figure out how someone feels in a sentence. It performs very well, even better than some human reviewers.

Finally, the research shows that automatic feedback classification helps companies better understand their customers — their needs, complaints, and preferences. This helps in making smart business choices, like creating new products, improving customer service, and managing online reputation.

يدرس هذه البحث تأثير التصنيف الآلي لتعليقات العملاء على اتخاذ القرارات الاستراتيجية والتشغيلية في مجال التجارة الإلكترونية والتسويق الرقمي. ونظرًا للعدد المتزايد من التعليقات، أصبحت عملية التحليل اليدوي غير فعالة، ولهذا تقترح الدراسة استخدام تقنيات معالجة اللغة الطبيعية (NLP) لأتمتة هذه المهمة.

الهدف هو تطوير نظام قادر على تصنيف تعليقات العملاء حسب نغمتها (إيجابية، سلبية، محايدة) أو محتواها (إحباط، رضا، وصف). وهذا يساعد على تحويل البيانات الخام إلى معلومات قابلة للاستفادة منها في اتخاذ القرارات.

يعتمد الإطار النظري على أهمية المعلومات في تقليل الشك أثناء اتخاذ القرارات، ويستخدم تقنيات مثل تنظيف النص، واستخراج الجذور، وتحويل النص إلى بيانات رقمية، بالإضافة إلى نماذج تصنيف تعتمد على القواميس، والتعلم الآلي.

تعتمد الدراسة على بيانات من تعليقات موقع أمازون، وتستخدم نموذجًا يُسمى VADER، وهو فعال جدًا في تحليل النصوص القصيرة مثل تلك الموجودة في وسائل التواصل الاجتماعي. يتميز هذا النموذج بدقة عالية.

في النهاية، يساهم هذا التصنيف الآلي في فهم أفضل للعملاء، واكتشاف الاتجاهات الجديدة، وتقسيم المستخدمين حسب مشاعرهم. وتؤثر هذه التحليلات بشكل مباشر على القرارات الاستراتيجية مثل تطوير المنتجات والتسويق، وأيضًا على القرارات التشغيلية مثل خدمة العملاء وتحسين العمليات وإدارة السمعة الإلكترونية.

تصنيف النصوص – تحليل النصوص -

Table des matières

Dédicace.....	2
Remerciement	3
Résumé.....	4
Abstract.....	5
Introduction Général	9
Contexte	9
Problématique:	10
Objectifs.....	10
Chapitre 1 Cadre Conceptuel et Revue de Littérature	11
1.1. Fondements de la Prise de Décision Stratégique	11
1.1.1. Définition et importance de la prise de décision stratégique dans le contexte des entreprises	11
1.1.2. Les différents niveaux de décision (stratégique, tactique, opérationnel)	12
1.1.3. Le rôle de l'information dans la prise de décision stratégique	13
1.2. Le E-commerce et le Marketing Digital: Enjeux et Tendances Actuelles	14
1.2.1. Évolution et caractéristiques du e-commerce.	14
1.2.2. Les spécificités du marketing digital et son importance	15
1.2.3. Les défis liés à la gestion de la relation client en ligne.....	15
1.3. Le Feedback Client en Ligne: Source d'Information Stratégiqu	16
1.4. La Classification Automatique du Texte: Concepts Fondamentaux et Applications.....	17
1.4.1. Définition et objectifs de la classification automatique de texte.....	17
1.4.2. Présentation détaillée des concepts clés du Traitement du Langage Naturel.....	17
1.4.3. Aperçu des différentes techniques de classification textuelle.....	19
1.4.4. Applications de la classification automatique de texte dans divers domaines	20
Chapitre 2 Méthodologie de recherche :	22
2.1 conception de la recherche :	22
2.1.1. les source de données :.....	22
2.1.2.Methode de collecte de données:	22
2.1.3 Description de l'échantillon de donnée	22
2.2. Méthode d'analyse de données:	23
2.2.1 Description détaillée des techniques de classification utilisé.....	23
2.2.2. présentation du Materiels et logiciels utilisé	25

Chapitre 3 Résultats de l'Analyse et Discussion	27
3.1. Présentation des Résultats de la Classification Automatique	27
3.1.1. Exemples concrets de commentaires classifiés et des catégories attribuées.	27
3.2. Analyse de l'Impact de la Classification sur la Compréhension des Clients.....	27
3.2.1. Comment la classification révèle les besoins, les attentes et les problèmes des clients.....	27
3.2.2. Identification des tendances émergentes et des signaux faibles :.....	28
3.2.3. Segmentation des clients basée sur l'analyse des commentaires.....	28
3.3. Influence de l'Analyse des Tonalités et des Thèmes sur les Décisions Stratégiques :	29
3.4. Influence sur les Décisions Opérationnelles	31
3.5 Discussion et résultat	32
3.5.1 pourquoi l'approche basée sur le lexique:.....	32
3.5.2 analyse des résultats en comparaison avec les notation en étoiles:.....	33
Conclusion Générale	35
References.....	36

Introduction Générale

La prise de décision est une compétence fondamentale qui régit nos vies au quotidien, qu'il s'agisse de choix personnels mineurs ou de décisions aux conséquences majeures. Dans presque tous les domaines d'activité humaine, la capacité à prendre des décisions éclairées et efficaces est un facteur clé de succès et de progrès. Qu'il s'agisse de la gestion d'une entreprise, de l'élaboration de politiques publiques, de l'avancement de la recherche scientifique ou même de nos interactions sociales, chaque action et chaque résultat découle d'une série de décisions. Dans le monde des affaires, et plus particulièrement dans les secteurs dynamiques du e-commerce et du marketing digital, la prise de décision est soumise à une pression constante. Les entreprises évoluent dans des environnements complexes, caractérisés par une concurrence accrue, des attentes clients en constante évolution et une quantité exponentielle de données. Dans ce contexte, la capacité à transformer ces données brutes en informations exploitables pour prendre des décisions stratégiques et opérationnelles judicieuses est devenue un impératif. L'une des sources d'information les plus précieuses pour les entreprises réside dans les retours de leurs clients. Leurs commentaires, qu'ils soient positifs ou négatifs, constituent une mine d'or d'insights sur leurs besoins, leurs préférences, leurs frustrations et leurs attentes. Cependant, le volume croissant de ces données rend leur analyse manuelle fastidieuse et souvent impossible. C'est là qu'intervient la classification automatique des commentaires clients, une approche qui permet d'extraire, d'organiser et d'interpréter efficacement ces informations pour éclairer la prise de décision.

Contexte

Avec le développement rapide du commerce en ligne, les clients laissent chaque jour de nombreux commentaires sur les produits et services qu'ils achètent. Ces avis sont très utiles pour les entreprises, car ils permettent d'améliorer les produits, de mieux comprendre les besoins des clients et d'adapter les stratégies marketing. Les commentaires sont devenus un outil important : ils influencent la réputation d'une marque, les décisions d'achat d'autres clients, et montrent ce qui fonctionne bien ou mal dans un produit. Mais comme leur nombre augmente chaque jour, il devient très difficile de les lire et de les analyser manuellement. Les méthodes classiques ne suffisent plus, surtout quand il faut comprendre des messages complexes, ambigus ou ironiques. C'est pour cela que l'utilisation du

traitement automatique du langage (NLP) est aujourd'hui une solution importante pour analyser et classer ces commentaires plus rapidement et plus efficacement.

Problématique:

Le défi de gérer beaucoup d'avis, c'est que les grandes entreprises, comme les petites boutiques en ligne, reçoivent des milliers de commentaires. Il est impossible pour des humains de lire et d'analyser tous ces messages pour en tirer des informations utiles. Il est difficile de savoir ce que les clients préfèrent quand il y a trop d'informations.

C'est pourquoi nous voulons automatiser cette tâche. Grâce à des outils informatiques, nous pouvons analyser ces commentaires rapidement et facilement, nous allons voir quels outils et techniques sont les plus efficaces pour trier automatiquement les commentaires. Cela permet aux entreprises de mieux comprendre leurs clients sans passer des heures à lire des textes. Nous détaillerons aussi les critères pour bien comprendre la "tonalité" des retours clients (s'ils sont positifs, négatifs, ou neutres). Enfin, nous montrerons comment les résultats de ces analyses influencent les décisions importantes (stratégiques) et les choix de tous les jours (opérationnels) des entreprises.

Objectifs

Notre objectif est de développer un logiciel capable de classer un nouveau commentaire mis en entrée en temps réel ou de segmenter une grande quantité de texte à la fois, et de donner le pourcentage des trois sentiments : frustration, satisfaction, ou description. Pour cela nous nous appuyerons sur des techniques de traitement du langage naturel (TALN). Nous visons également l'affichage des résultats de classification, permettant une visualisation claire des tendances de sentiment et des thèmes émergents. La classification des avis, ou "opinion mining", correspond au processus d'analyse de gros volumes de textes en vue de déterminer si ces derniers expriment un sentiment positif, négatif ou neutre. Le présent rapport a pour objectif de détailler le rôle et l'impact de la classification automatique du feedback client sur la prise de décision stratégique dans le e-commerce et le marketing digital.

Chapitre 1 Cadre Conceptuel et Revue de Littérature

Ce chapitre établit les fondements théoriques et synthétise la recherche existante pertinente à la prise de décision, au e-commerce, au marketing digital, au retours client en ligne et à la classification automatique de texte.

1.1. Fondements de la Prise de Décision Stratégique

1.1.1. Définition et importance de la prise de décision stratégique dans le contexte des entreprises

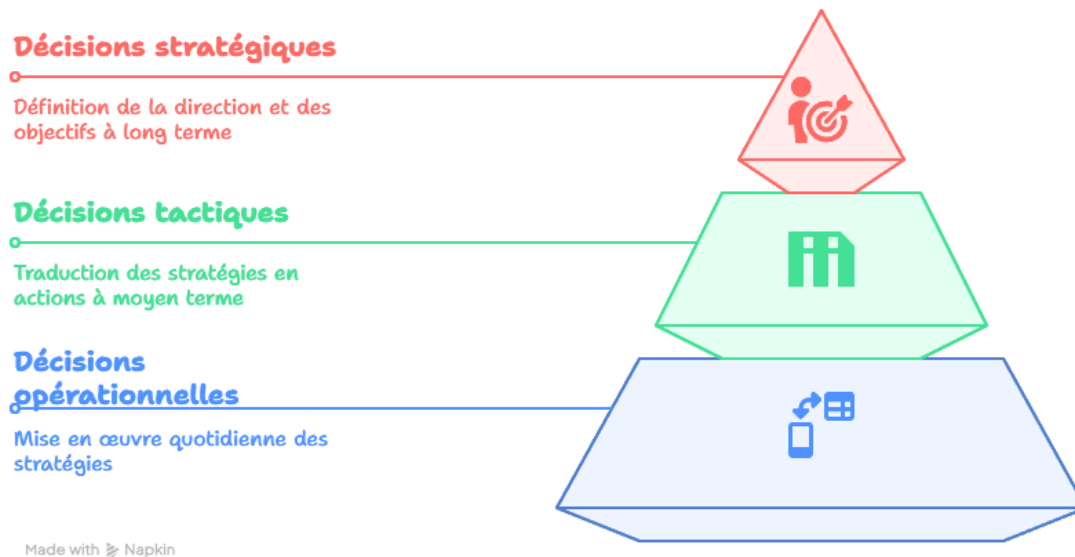
Dans un environnement concurrentiel, la prise de décision est un élément clé de la stratégie d'entreprise. Elle se définit comme l'ensemble des manœuvres permettant à une entreprise de se démarquer. Ces décisions se caractérisent par une orientation à long terme afin d'atteindre ses objectifs. L'importance de ces décisions réside dans leur capacité à offrir un contrôle sur l'évolution de l'entreprise, à assurer la cohérence entre ces différentes activités, et à permettre l'anticipation des opportunités et des menaces.

[1] [2]

1.1.2. Les différents niveaux de décision (stratégique, tactique, opérationnel)

Les décisions se manifestent à différents niveaux, chacun ayant des caractéristiques distinctives

Hiérarchie des décisions organisationnelles



Les décisions stratégiques représente le niveau le plus élevé, définissant la direction, la vision, la mission et les objectifs à long terme de l'entreprise [1]

les décisions tactiques, niveau intermédiaire, traduit les choix stratégiques en actions concrètes à moyen terme au sein de différents services. Les responsables intermédiaires élaborent les plans tactiques à moyen terme à partir des plans stratégiques. Ces plans reflètent les décisions relatives à la contribution de leur service fonctionnel à l'atteinte des objectifs organisationnels. [1]

Enfin, **les décisions opérationnelles** est le niveau le plus bas, axé sur la mise en œuvre quotidienne des stratégies, touchent les activités courantes de l'entreprise et assurent son bon fonctionnement au quotidien. La nature hiérarchique de ces décisions implique un flux d'informations et une influence en cascade. Les décisions stratégiques établissent la direction générale, ce qui contraint et guide ensuite les choix tactiques et opérationnels.

Les gestionnaires de premier niveau élaborent les plans opérationnels, qui entraînent des décisions routinières concernant la production quotidienne ou hebdomadaire de biens ou de services.

[1]

Niveau de Décision	Planification dans le temps	Portée	Nature des Décisions	Exemples
Stratégique	Long terme	Globale de l'entreprise	Définition de la vision, objectifs majeurs, allocation de ressources	Ouverture de marchés étrangers, Lancement de nouveaux produits majeurs
Tactique	Moyen terme	Fonctionnelle/ Départementale	Planification d'actions pour atteindre les objectifs stratégiques	Campagnes marketing spécifiques, Budgétisation annuelle
Opérationnel	Court terme	Quotidienne/Spécifique	Exécution des tâches, optimisation des processus	Gestion du service client, Optimisation des stocks

Tableau 1.1: Comparaison des Niveaux de Décision (Stratégique, Tactique, Opérationnel).

1.1.3. Le rôle de l'information dans la prise de décision stratégique

L'information est le point d'appui du processus décisionnel, étant la source de toute décision à prendre et touchant au pouvoir de la stratégie et au fonctionnement de l'entreprise. Son objectif principal est de réduire l'incertitude, permettant aux décideurs de mieux comprendre les contraintes environnementales, d'identifier des solutions et d'anticiper les conséquences. Pour être utile, l'information doit être pertinente, compréhensible, fiable, accessible et obtenue avec un effort et un temps raisonnables.¹⁰ Dans le contexte de la prise de décision stratégique, la valeur de l'information ne réside pas seulement dans sa présence, mais dans sa capacité à être exploitée. [1] [2]

1.2. Le E-commerce et le Marketing Digital: Enjeux et Tendances Actuelles

1.2.1. Évolution et caractéristiques du e-commerce.



Au début des années 1990 et la commercialisation d'Internet, le e-commerce a connu une évolution significative, rendant la technologie informatique ouverte et la connectivité abordables. Les consommateurs s'attendent de plus en plus à pouvoir effectuer leurs achats dans le confort de leur foyer, peu importe l'heure et le moment de la semaine ou de l'année. Convaincre les clients d'acheter des produits en ligne nécessite cependant une approche différente. Les caractéristiques du e-commerce font qu'aucun vendeur n'est présent, le client n'est pas influencé par l'ambiance, et l'environnement est entièrement virtuel. Il faut alors donner confiance à l'internaute.

[3] [1]

1.2.2. Les spécificités du marketing digital et son importance

L'usage d'Internet et des autres technologies numériques à des fins commerciales a donné naissance à de nombreux termes. Celui de « marketing digital » est lui-même assez récent. Dans le passé, on désignait ces pratiques sous les noms de « e-marketing », « Internet marketing » ou encore « webmarketing », ce dernier étant toujours utilisé aujourd'hui. On peut simplement définir le marketing digital comme l'atteinte d'objectifs propres au marketing (image, préférence, engagement des clients, taux de rétention, ventes ou part de marché, etc.) grâce à l'usage de l'outil informatique. Ces diverses technologies permettent d'atteindre des objectifs de fidéliser la clientèle actuelle, de développer l'e-réputation d'une entreprise. Cela signifie que les entreprises doivent non seulement collecter le feedback client, mais aussi l'écouter activement.

La classification automatique est la technologie qui permet cette écoute à grande échelle et en temps réel, ainsi que des connaissances exploitables. [1] [4] [5]

1.2.3. Les défis liés à la gestion de la relation client en ligne.

La gestion de la relation client, connue sous l'acronyme de CRM (Customer Relationship Management) en anglais, combine les technologies et les stratégies commerciales pour offrir aux clients les produits et les services qu'ils attendent ou qu'ils sont prêts à payer. La gestion de la relation client est la capacité à identifier, à acquérir et à fidéliser les meilleurs clients dans l'optique d'augmenter le chiffre d'affaires et les bénéfices. Le terme de gestion de la relation client est devenu le fédérateur de nombreux fournisseurs de solutions informatiques. La gestion de la relation client en ligne est confrontée à plusieurs défis, notamment le maintien d'interactions client authentiques à l'ère numérique, l'établissement de stratégies de communication numérique exceptionnelles et la gestion du volume croissant de données client. L'intégration de l'analyse du Big Data dans la gestion de la relation client (GRC) est complexe et gourmande en ressources, nécessitant une gouvernance des données robuste, une intégration des systèmes et des considérations éthiques. Les outils de data mining jouent un rôle crucial dans l'exploitation du Big Data et de la GRC, en automatisant les tâches et en offrant des expériences personnalisées. [6]

1.3. Le Feedback Client en Ligne: Source d'Information Stratégique

Les feedback Client en ligne



Made with Napkin

Le feedback client en ligne se présente sous diverses formes : notes chiffrées (par exemple, de 1 à 5 étoiles), commentaires écrits, avis et interactions sur les médias sociaux, les forums et les blogs. Des plateformes comme Amazon intègrent les avis et les prennent même en compte dans leurs algorithmes de classement. La diversité des formes et des plateformes de feedback en ligne implique qu'un système d'analyse complet ne peut pas se limiter à un seul type de données ou une seule source.

C'est pourquoi les outils sociaux tels que Twitter, Facebook et les plateformes de blog une mine de données inestimable pour comprendre les préférences, le comportement d'achat et les besoins, permettant ainsi des décisions stratégiques éclairées. Ces plateformes permettent aux entreprises de recueillir et d'analyser les avis et commentaires des clients. Ces retours ont une valeur unique et renforcent la capacité d'une entreprise à comprendre ses clients en surveillant les informations échangées. Des chercheurs ont étudié plusieurs aspects afin de mieux les analyser. Une grande partie de ces recherches montre qu'elles peuvent jouer un rôle important dans le rapprochement du consommateur vers une décision d'achat positive et proposer des offres ciblées pour comprendre les préférences, le comportement d'achat et les besoins, permettant ainsi des décisions stratégiques éclairées. Il contribue à réduire l'incertitude et le risque d'achat pour les consommateurs, influençant leurs décisions d'achat,

améliorer leurs stratégies de gestion de la relation client.

Cependant, la collecte, la gestion et l'analyse du feedback client à grande échelle présentent des défis significatifs. Le volume est un problème majeur, avec des millions d'interactions client à traiter quotidiennement sur de multiples plateformes. La complexité réside dans l'interprétation d'expressions de sentiment très contextuelles, ambiguës et spécifiques à un domaine. La qualité et la précision des données sont cruciales, car l'efficacité de l'analyse du Big Data dépend de la fiabilité des données analysées [6][7] [8]

1.4. La Classification Automatique du Texte: Concepts Fondamentaux et Applications

1.4.1. Définition et objectifs de la classification automatique de texte

La classification automatique de texte (ou catégorisation de texte, TC) est une tâche fondamentale dans l'exploration de texte et le traitement du langage naturel (TALN). Son objectif est d'attribuer des catégories ou des étiquettes prédéfinies à des documents textuels. Cela permet une récupération d'informations, un filtrage de contenu, une identification de sujets et une analyse efficaces, rationalisant les processus et améliorant la productivité tout en soutenant la prise de décision. L'objectif principal de l'automatisation de la classification de texte est de transformer des données textuelles non structurées et volumineuses en informations structurées et exploitables. Cette transformation est essentielle pour la prise de décision stratégique, car elle convertit le feedback qualitatif en informations quantifiables qui peuvent être agrégées, analysées pour des tendances et comparées, permettant ainsi des stratégies axées sur les données à grande échelle. [9] [10]

1.4.2. Présentation détaillée des concepts clés du Traitement du Langage Naturel

Le traitement du langage naturel (NLP) repose sur plusieurs concepts clés pour permettre aux machines de comprendre et de traiter le langage humain.

1- Le Nettoyage de Base (Prétraitement du Texte) prétraitement des données textuelles brutes. Cette étape consiste à nettoyer le texte en le transformant dans un format plus facilement compréhensible par les machines.

2- Suppression des Mots Vides (Stop-word Removal) qui n'apportent pas de signification au texte ou qui sont des mots qui n'apportent pas beaucoup d'informations dans les textes. Les supprimer permet de réduire la dimensionnalité des données et de se focaliser sur les termes clés, améliorant ainsi l'efficacité

et la pertinence de l'analyse.

ex :

'a', 'about', 'above', 'across', 'after', 'afterwards', 'again', 'against', 'all', 'almost', 'alone', 'along', 'already', 'also', 'although', 'always', 'am', 'among', 'amongst', 'amount', 'an', 'and', 'another', 'any', 'anyhow', 'anyone', 'anything', 'anyway', 'anywhere', 'are', 'around', 'as', 'at', 'back', 'be', 'became', 'because', 'become', 'becomes', 'becoming', 'been', 'before', 'beforehand', 'behind', 'being', 'below', 'beside', 'besides'

3- Lemmatisation La lemmatisation est une technique qui vise à obtenir la forme canonique (le lemme) d'un mot à partir de ses différentes flexions. Par exemple, pour un verbe, elle réduit le mot à son infinitif (ex: "montrer", "montreras", "montraient" deviennent "montrer")..

4- Racinisation (Stemming) Le processus de racinisation, ou "stemming", vise à réduire les mots à leur racine, c'est-à-dire une forme tronquée du mot. Par exemple, "manger", "mangent", "mangé" peuvent être racinés à "mang". Il s'agit de la suppression des flexions et des suffixes..

5- Vectorisation et Sac de Mots (Bag-of-Words) La vectorisation est une méthode fondamentale utilisée dans le traitement du langage naturel pour transformer le texte en un format numérique (vecteur ou matrice) que les algorithmes peuvent comprendre et traiter. Le "Bag-of-Words" (BoW) est une technique de représentation textuelle qui convertit un texte en un vecteur de comptage d'occurrences de chaque mot dans le document, en ignorant l'ordre des mots et la structure linguistique du texte. Chaque terme est traité comme un élément dans un vecteur, où la valeur représente sa fréquence d'apparition.
[10]

'company', 'took', 'money', 'sent', 'email', 'telling', 'product', 'shipped', 'week', 'half', 'later', 'received', 'email', 'telling', 'sorry', 'actually', 'items', 'received', 'email', 'telling', 'shipped', 'mistake', 'finally', 'got', 'money', 'went', 'company', 'buy', 'product', 'work', 'phone', 'depicts', 'sent', 'numerous', 'emails', 'company', 'actually', 'find', 'phone', 'number'

Tableau 1.3: Concepts Fondamentaux du Traitement du Langage Naturel (NLP)

Concept NLP	Définition/Objectif	Importance pour l'analyse de feedback
Tokenisation	Découper le texte en unités	Prépare le texte pour l'analyse
Lemmatisation	Réduire les mots à leur forme de base	Normalise le vocabulaire
Suppression des mots vides	Éliminer les mots courants non significatifs	Réduit le bruit
Vectorisation	Transformer le texte en format numérique	Permet aux algorithmes de traiter le langage
<i>(Ex: Bag-of-Words, TF-IDF, Word Embeddings, Contextualized Embeddings)</i>		

Ce tableau décrit les techniques du NLP en définissant clairement chaque concept

1.4.3. Aperçu des différentes techniques de classification textuelle

Les techniques de classification des textes peuvent être classées en trois grandes catégories : les approches basées sur les lexiques, les approches basées sur l'apprentissage automatique et les approches hybrides.

1. Approches basées sur les lexiques :

Ces méthodes reposent sur des listes prédéfinies de mots (lexiques) associés à un sentiment positif ou négatif. Chaque mot du texte est comparé au lexique, et un score de sentiment est calculé en fonction de la présence et de l'intensité de ces mots. Bien que les approches basées sur les lexiques soient simples et faciles à interpréter, elles peuvent rencontrer des difficultés avec le contexte et le langage nuancé.

2. Approches basées sur l'apprentissage automatique :

Ces techniques impliquent l'entraînement de modèles d'apprentissage automatique sur des ensembles de données étiquetés pour classifier les sentiments. Parmi les algorithmes populaires figurent Naive Bayes, les machines à vecteurs de support (SVM) et, plus récemment, les modèles d'apprentissage profond tels que les réseaux neuronaux récurrents (RNN) et les transformateurs. Les approches basées sur l'apprentissage automatique permettent de capturer des schémas complexes et le contexte, mais nécessitent de grandes quantités de données annotées ainsi que des ressources computationnelles importantes.

3. Approches hybrides :

Combinant les forces des méthodes basées sur les lexiques et de celles basées sur l'apprentissage automatique, les approches hybrides visent à améliorer la précision et la robustesse de l'analyse des sentiments. Ces méthodes utilisent les lexiques pour un premier score de sentiment, puis des modèles d'apprentissage automatique pour affiner et contextualiser ces scores. [11]

1.4.4. Applications de la classification automatique de texte dans divers domaines

La classification des textes est devenue essentielle dans de nombreux secteurs et applications, notamment :

Marketing : Les entreprises utilisent cette technique pour évaluer l'opinion publique sur leurs produits et services, surveiller la réputation de leur marque et adapter leurs stratégies marketing en fonction des ressentis des consommateurs.

Surveillance des réseaux sociaux : Les outils d'analyse des sentiments permettent de suivre les plateformes sociales afin de comprendre la réaction du public face à des événements, campagnes ou changements de politiques publiques, offrant ainsi des informations précieuses pour les relations publiques et la gestion des crises.

Analyse des retours clients : Les entreprises exploitent cette approche pour interpréter les avis des clients, ce qui leur permet d'améliorer la satisfaction et la fidélité en répondant rapidement aux ressentis négatifs.

Marchés financiers : Les investisseurs et analystes se basent sur l'analyse des sentiments pour anticiper les tendances et évolutions du marché, en s'appuyant sur les sentiments exprimés dans les articles de presse, les réseaux sociaux et les rapports financiers.

Santé : Cette technique est utilisée pour comprendre les retours des patients, améliorer les services de santé et surveiller la santé mentale à travers l'analyse des réseaux sociaux et des forums en ligne.

Chapitre 02 Méthodologie de recherche :

2.1 conception de la recherche :

2.1.1. les source de données :

La recherche de datasets de commentaires clients pour des plateformes comme Amazon s'est avérée plus complexe que prévu. Ces données sont souvent la propriété des entreprises et ne sont pas toujours faciles à trouver publiquement. Cependant, après des efforts soutenus et des recherches sur diverses sources en ligne, j'ai finalement réussi à localiser les ensembles de données pertinents qui me sont nécessaires pour mon projet. Cette étape, bien que difficile, a été cruciale pour la suite de mon travail.

Ces données sont présent ici pour des raison éducatif [21]

2.1.2. Methode de collecte de données:

Plusieurs méthodes peuvent être employées pour collecter ces données. L'utilisation d'**Interfaces de Programmation d'Applications (API)**, lorsqu'elles sont mises à disposition par les plateformes (par exemple, l'API de Twitter), est une méthode privilégiée car elle permet de recueillir des données de manière structurée et souvent en temps réel, concernant les interactions des utilisateurs et le contenu qu'ils publient.

En l'absence d'API ou lorsque celles-ci sont limitées, le **web scraping** (ou moissonnage web) peut être utilisé

2.1.3 Description de l'échantillon de donnée

Cet échantillon de données textuelles est un avis client Amazon, caractérisé par son contenu subjectif et non structuré rédigé en langage naturel par un utilisateur. Il s'agit d'un retour d'expérience brut sur un produit ou un service, exprimant un sentiment global (qu'il soit positif, négatif ou neutre) et détaillant des aspects de l'interaction client, de l'achat ou de l'utilisation du produit. Sa spécificité réside dans sa capacité à offrir une perspective directe et authentique du client, même si elle peut contenir le bruit typique des données générées par les utilisateurs.

2.2. Méthode d'analyse de données:

2.2.1 Description détaillée des techniques de classification utilisé

VADER est un modèle simple basé sur des règles pour l'analyse générale des sentiments. Il est spécifiquement adapté à l'analyse des sentiments dans les textes des médias sociaux (comme Twitter et Facebook), qui posent des défis particuliers en raison de leur brièveté, de leur sparsité contextuelle et de l'utilisation de conventions linguistiques abrégées comme les acronymes, les initialismes, les émoticônes et l'argot. Malgré son optimisation pour les médias sociaux, VADER généralise favorablement à d'autres domaines

Le modèle VADER se compose de deux éléments principaux :

1. Un lexique de sentiments validé par des humains : Ce lexique est une liste de caractéristiques lexicales (mots, émoticônes, acronymes, argot) associées à des mesures d'intensité de sentiment (valence). La création et la validation de ce lexique ont utilisé une approche centrée sur l'humain et la méthode de la "sagesse des foules" (wisdom-of-the-crowd). Des évaluateurs humains indépendants ont noté l'intensité du sentiment de plus de 9 000 caractéristiques lexicales potentielles sur une échelle de -4 (extrêmement négatif) à +4 (extrêmement positif), avec un point neutre à 0. Le lexique final contient un peu plus de 7 500 caractéristiques lexicales avec des scores de valence validés. Contrairement à d'autres lexiques comme LIWC ou GI, le lexique VADER est spécifiquement accordé aux expressions de sentiment dans les médias sociaux. Contrairement à certains, il associe des scores de valence plutôt qu'une simple classification binaire (positif/négatif).

2. Un ensemble de cinq règles générales basées sur des conventions grammaticales et syntaxiques : Ces règles sont basées sur l'observation de la manière dont les humains expriment et accentuent l'intensité du sentiment dans le texte. Elles vont au-delà d'un simple modèle "sac-de-mots" et intègrent des relations sensibles à l'ordre des mots. Les règles incluent la prise en compte de :

- La ponctuation (notamment le point d'exclamation) qui augmente l'intensité sans modifier la polarité.
- La capitalisation (utilisation de MAJUSCULES pour souligner un mot lié au sentiment) qui augmente l'intensité.
- Les modificateurs de degré (intensificateurs, adverbes) qui augmentent ou diminuent l'intensité du

sentiment.

- La conjonction contrastive "but" qui signale un changement de polarité, le sentiment après "but" étant dominant.
- La négation (identifiée par l'examen du tri-gramme précédant un terme de sentiment) qui inverse la polarité. Ces règles ont été identifiées par analyse qualitative puis quantifiées expérimentalement en utilisant des tweets manipulés et des évaluations humaines, démontrant leur impact statistiquement significatif sur l'intensité perçue du sentiment.

Les performances de VADER ont été évaluées et comparées à celles de onze autres références, y compris d'autres lexiques (LIWC, ANEW, General Inquirer, SentiWordNet, SenticNet, Hu-Liu04, WSD) et des techniques d'apprentissage automatique (Naive Bayes, Maximum Entropy, SVM). Toutes les comparaisons de lexiques ont utilisé le modèle basé sur des règles de VADER pour être justes.

- Corrélation avec la vérité terrain : Sur un corpus de plus de 4 000 tweets annotés par 20 évaluateurs humains, la corrélation de VADER ($r = 0,881$) avec la vérité terrain (la moyenne des notes humaines agrégées) est aussi bonne que celle des évaluateurs humains individuels ($r = 0,888$).
- Classification en 3 classes (positif/négatif/neutre) : Pour la classification des tweets en trois classes (avec des seuils de $-0,05$ et $+0,05$ pour le neutre sur une échelle normalisée de -1 à 1), VADER atteint un score F1 de $0,96$, ce qui surpasse même les évaluateurs humains individuels ($F1 = 0,84$). VADER a également montré des performances de classification fortes sur d'autres domaines (critiques de films, critiques de produits, articles d'opinion du NY Times), avec des scores F1 de $0,95$, $0,91$ et $0,92$ respectivement.
- Comparaison avec l'apprentissage automatique : Les modèles d'apprentissage automatique entraînés sur des domaines spécifiques n'ont pas surpassé de manière significative VADER. Dans trois cas sur quatre testés, VADER a obtenu des performances équivalentes ou supérieures dans d'autres domaines par rapport aux méthodes ML dans le domaine pour lequel elles étaient entraînées. VADER est également plus efficace sur le plan computationnel et plus rapide que certains modèles ML complexes, surtout sur de grands corpus.

VADER montre une performance remarquable avec un score F1 de $0,96$ sur les tweets, surpassant même les évaluateurs humains individuels. Il démontre également une capacité favorable à généraliser à d'autres types de texte, tout en étant rapide, interprétable et ne nécessitant pas de données d'entraînement

étendue

2.2.2. présentation du Matériels et logiciels utilisé

Pour ce projet, un **ordinateur portable** avec les spécifications suivantes a été utilisé :

- **Processeur** : Intel Core i5-8250U
- **Mémoire vive (RAM)** : 8 Go
- **Stockage** : 256 Go SSD

Logiciels

Plusieurs outils logiciels essentiels ont été employés pour le développement et l'analyse :

VS Code (Visual Studio Code)

VS Code est un éditeur de code source développé par Microsoft. Il est reconnu pour sa légèreté, sa puissance, et sa popularité, notamment pour le développement Python et l'analyse de données, grâce à son riche écosystème d'extensions.

Avantages :

- **Polyvalence** : Supporte une multitude de langages de programmation et dispose d'extensions utiles.
- **Performance** : Offre une expérience rapide et fluide.
- **Écosystème d'extensions riche** : Des milliers d'extensions sont disponibles, permettant l'intégration de fonctionnalités avancées comme Jupyter Notebook.
- **Interface intuitive** : Facile à prendre en main et à utiliser.
- **Gratuit et Open Source** : Accessible à tous sans coût.

Python

Python est un langage de programmation de haut niveau, interprété et polyvalent. Il est apprécié pour sa syntaxe simple, sa lisibilité et sa vaste collection de bibliothèques, ce qui en fait un choix privilégié pour l'automatisation, l'intelligence artificielle, l'apprentissage automatique et l'analyse de données.

Avantages :

- **Simplicité et clarté** : Sa syntaxe intuitive le rend facile à apprendre et à utiliser.
- **Polyvalence** : Utilisé dans de nombreux domaines tels que la science des données (avec

Pandas/NumPy), l'IA (avec TensorFlow/PyTorch) et l'automatisation de scripts.

- **Grande communauté et écosystème** : Bénéficie d'une communauté très active et d'un immense ensemble de bibliothèques et de frameworks (via PyPI et Anaconda).
- **Productivité** : Permet un développement rapide grâce à sa syntaxe concise et ses bibliothèques riches.

Jupyter Notebook (intégré dans VS Code)

Nous utilisons Jupyter Notebook directement au sein de VS Code, combinant ainsi l'interactivité des notebooks avec la puissance de notre IDE. Ce n'est pas une application web séparée, mais une fonctionnalité intégrée à notre environnement de travail.

Pourquoi l'utiliser : Cette intégration nous offre un environnement unifié où nous pouvons alterner fluidement entre l'écriture de scripts Python traditionnels et l'expérimentation interactive. Jupyter Notebook est indispensable pour l'exploration des données, la prototypage rapide et la documentation collaborative. Il nous permet d'exécuter du code cellule par cellule, de visualiser les résultats et les graphiques instantanément, et de documenter nos analyses avec du texte narratif, tout cela sans quitter VS Code. Cela fluidifie notre workflow et nous permet de maintenir nos données et calculs sur nos machines locales, offrant un contrôle total et une performance optimisée.

Chapitre 3: Résultats de l'Analyse et Discussion

Ce chapitre présentera les résultats de la classification automatique, analysera leur impact sur la compréhension des clients et discutera de leur influence sur les décisions stratégiques et opérationnelles.

3.1. Présentation des Résultats de la Classification Automatique

3.1.1. Exemples concrets de commentaires classifiés et des catégories attribuées.

First of all, the company took my money and sent me an email telling me the product was shipped. A week and a half later I received another email telling me that they are sorry, but they don't actually have any of these items, and if I received an email telling me it has shipped, it was a mistake. When I finally got my money back, I went through another company to buy the product and it won't work with my phone, even though it depicts that it will. I have sent numerous emails to the company - I can't actually find a phone number on their website - and I still have not gotten any kind of response. What kind of customer service is that? No one will help me with this problem. My advice - don't waste your money!

le score attribué pour chaque classe

```
{'neg': 0.062, 'neu': 0.88, 'pos': 0.059, 'compound': -0.1107}
```

donc il est négatif

3.2. Analyse de l'Impact de la Classification sur la Compréhension des Clients

La classification automatique des sentiments, telle que décrite par le modèle basé sur le lexique proposé dans le document, permet de "cerner le sentiment et l'état émotionnel d'un utilisateur en ligne" en classifiant les textes. L'objectif principal est de prédire la "polarité des sentiments d'un utilisateur, à savoir : positif, négatif et neutre, à partir d'avis en ligne"

3.2.1. Comment la classification révèle les besoins, les attentes et les problèmes des clients

En classifiant un avis comme positif, négatif ou neutre, le système évalue "comment les consommateurs réagissent aux produits". Une réaction positive suggère que les besoins et attentes sont probablement satisfaits et que les problèmes sont absents. Une réaction négative indique le contraire.

Plus spécifiquement, l'analyse des avis négatifs est particulièrement révélatrice une "opinion négative sur un produit aide le vendeur en le rendant conscient des défauts et des lacunes existantes dans

le produit". Cette prise de conscience pousse ensuite le vendeur à "évaluer et améliorer les standards du produit". L'analyse de la tonalité et du contenu des avis négatifs permet donc d'identifier directement les problèmes spécifiques rencontrés par les clients et, par extension, les besoins ou attentes qui n'ont pas été satisfaits.

L'ensemble des avis, qu'ils soient positifs ou négatifs, "peuvent fournir une connaissance des opinions des clients sur les produits". En analysant ces opinions classifiées, les entreprises peuvent comprendre si la perception globale est alignée avec leurs objectifs et identifier les points forts (souvent mis en évidence par les avis positifs) et les points faibles (souvent mis en évidence par les avis négatifs). Le modèle proposé calcule une "valeur de sentiment" (sentival) qui est positive (>0), négative (<0) ou nulle ($=0$), permettant de prédire si l'utilisateur est un "optimiste (avec une polarité positive), un pessimiste (avec une polarité négative) ou un ambivalent (avec une polarité neutre)". Cette classification simple mais efficace révèle l'attitude générale des utilisateurs.

3.2.2. Identification des tendances émergentes et des signaux faibles :

L'analyse d'un "immense dépôt de données textuelles" provenant du web. L'application de la classification des sentiments sur ce vaste ensemble de données permet d'identifier des patterns récurrents. Par exemple, une augmentation soudaine des avis négatifs mentionnant un aspect spécifique d'un produit, ou une mention de plus en plus fréquente d'un point positif particulier par les "avis positifs" qui guident d'autres clients, peuvent être interprétées comme l'identification de signaux faibles ou de tendances émergentes concernant la réception d'un produit ou service. L'analyse continue de ces avis classifiés permet de suivre l'évolution du sentiment client.

3.2.3. Segmentation des clients basée sur l'analyse des commentaires

la "polarité des sentiments d'un utilisateur" ou le "niveau de polarité des sentiments d'un utilisateur comme positif, négatif ou neutre". Cette classification de chaque utilisateur ou commentaire selon sa polarité sentimentale (optimiste, pessimiste, ambivalent) constitue une forme de segmentation simple basée sur l'attitude émotionnelle exprimée dans les avis.

Il existe d'autres méthodes plus complexes de segmentation client (basée sur des données démographiques, l'historique d'achat, la fréquence des commentaires, etc.) en utilisant les résultats de l'analyse de sentiments. La segmentation abordée se limite à la classification de l'individu ou du commentaire en catégories de sentiment.

3.3. Influence de l'Analyse des Tonalités et des Thèmes sur les Décisions Stratégiques

:

L'analyse des sentiments basée sur les avis en ligne a une influence directe et indirecte sur les décisions stratégiques des entreprises, principalement en les aidant à mieux comprendre le marché et la performance de leurs produits/services.

les mécanismes par lesquels cette analyse peut influencer les décisions stratégiques :

Développement de nouveaux produits ou amélioration de services :

L'analyse des avis négatifs rend le vendeur "conscient des défauts et des lacunes existantes dans le produit". Cette information est cruciale pour "évaluer et améliorer les standards du produit". Cela implique que l'identification récurrente de problèmes spécifiques par la classification négative peut orienter les efforts de recherche et développement ou les plans d'amélioration des services. L'analyse aide à "répondre aux demandes et aux besoins des consommateurs". [13] [14]

Ajustements marketing :

Les avis positifs servent de "guide qui pousse d'autres clients à prendre des décisions d'achat", tandis que les avis négatifs aident un client "à repenser son choix avant d'acheter un produit". Comprendre la polarité générale et les raisons derrière le sentiment peut aider les équipes marketing à ajuster les messages promotionnels, à mettre en avant les points forts soulignés par les avis positifs, ou à répondre proactivement aux préoccupations soulevées par les avis négatifs. L'analyse des sentiments permet d'évaluer "comment les consommateurs réagissent aux produits", information fondamentale pour des campagnes marketing efficaces.

Augmentation des ventes et croissance :

L'objectif global de l'analyse des avis pour les entreprises est d'augmenter leurs ventes et de favoriser la croissance. L'analyse aide à "transformer toute la méthodologie de trading" en offrant une "nouvelle façon de gérer les plaintes et les avis", ce qui est censé impacter positivement les résultats commerciaux. Analyse de l'impact sur le positionnement de la marque, la fidélisation client, et la veille concurrentielle :

Positionnement de la marque :

L'analyse des sentiments révèle la perception des clients sur la marque et ses produits. Une prédominance d'avis positifs renforce un positionnement favorable, tandis que des avis négatifs fréquents

peuvent nuire à la réputation de la marque. La classification permet de quantifier cette perception et d'identifier les thèmes associés, aidant ainsi à affiner ou à défendre le positionnement de la marque.

Fidélisation client :

En identifiant et en répondant aux "défauts et lacunes" signalés dans les avis négatifs, les entreprises peuvent améliorer leurs produits et services pour mieux "répondre aux demandes et aux besoins des consommateurs". Cette amélioration, guidée par le feedback client classifié, est essentielle pour améliorer la satisfaction et potentiellement la fidélisation des clients existants. Le document ne mesure pas la fidélisation directement, mais décrit le processus qui y contribue.

Veille concurrentielle :

Le document se concentre principalement sur l'analyse des avis concernant *les produits de l'entreprise en question*. Il ne mentionne pas explicitement l'application de cette méthodologie à l'analyse des avis sur les produits concurrents pour la veille concurrentielle. Cependant, le principe de classification des sentiments et l'analyse du "dépôt de données textuelles" pourraient logiquement être appliqués aux avis concurrents pour obtenir des informations sur leurs points forts et faibles perçus par les clients, bien que cela ne soit pas décrit dans le document fourni.

En résumé, le modèle de classification basé sur le lexique est un outil fondamental pour comprendre la polarité des sentiments des clients. Cette compréhension permet d'identifier les problèmes (défauts et lacunes), d'évaluer les réactions des consommateurs, et aide les entreprises dans leur objectif d'augmenter les ventes et la croissance en améliorant les produits et en répondant aux besoins.

3.4. Influence sur les Décisions Opérationnelles

Voici comment l'analyse des commentaires, notamment via la classification des sentiments (positif, négatif, neutre) peut influencer les décisions opérationnelles

[15] [16]

Amélioration du service client :

Les avis en ligne offrent une "nouvelle façon de gérer les plaintes et les avis" et leur analyse permet de "rencontrer les demandes et les besoins des consommateurs"

Identifier les avis négatifs (pessimistes) signale les "défauts et les lacunes existantes dans le produit", permettant au vendeur d'en prendre conscience.

Optimisation des processus :

Une "opinion négative sur un produit" aide le vendeur en le rendant "conscient des défauts et des lacunes existantes dans le produit". Cela "pousse les entreprises à évaluer et améliorer les standards du produit". L'analyse de sentiment, en identifiant la polarité, met en évidence les domaines nécessitant potentiellement une révision des processus sous-jacents (qualité, production, etc.).

Gestion de la réputation en ligne :

Les avis en ligne sont un facteur majeur qui "influence les comportements d'achat et les habitudes d'achat" des consommateurs. Donc un avis positif "prouve être un guide qui pousse d'autres clients à prendre des décisions d'achat". Mais un avis négatif "aide un client à repenser son choix avant d'acheter un produit". La classification des avis client "détecte les avis en ligne et les manipule pour évaluer comment les consommateurs réagissent aux produits en prédisant les sentiments des utilisateurs", ce qui est essentiel pour mesurer et suivre la réputation. La classification en positif, négatif ou neutre donne une mesure directe de cette réaction. [16] [18] [19]

[20]

3.5 Discussion et résultat

3.5.1 pourquoi l'approche basée sur le lexique:

Les techniques d'analyse des sentiments ont considérablement évolué au fil des ans et peuvent être classées en trois grandes catégories : les approches basées sur les lexiques, les approches basées sur l'apprentissage automatique et les approches basées sur l'apprentissage profond.

1. **Approches basées sur les lexiques :**

- o **Aperçu** : Ces méthodes reposent sur des listes prédéfinies de mots porteurs de sentiment (lexiques) et leurs scores associés. Elles calculent le sentiment global d'un texte en agrégeant les scores individuels des mots présents.
- o **Avantages** : Simplicité, interprétabilité et facilité de mise en œuvre. Elles ne nécessitent pas de données étiquetées pour l'entraînement.
- o **Inconvénients** : Capacité limitée à gérer le contexte, le sarcasme et le langage nuancé. Leur performance dépend fortement de la qualité et de la couverture du lexique.
- o **Exemples** : SentiWordNet, AFINN, VADER.

2. **Approches basées sur l'apprentissage automatique :**

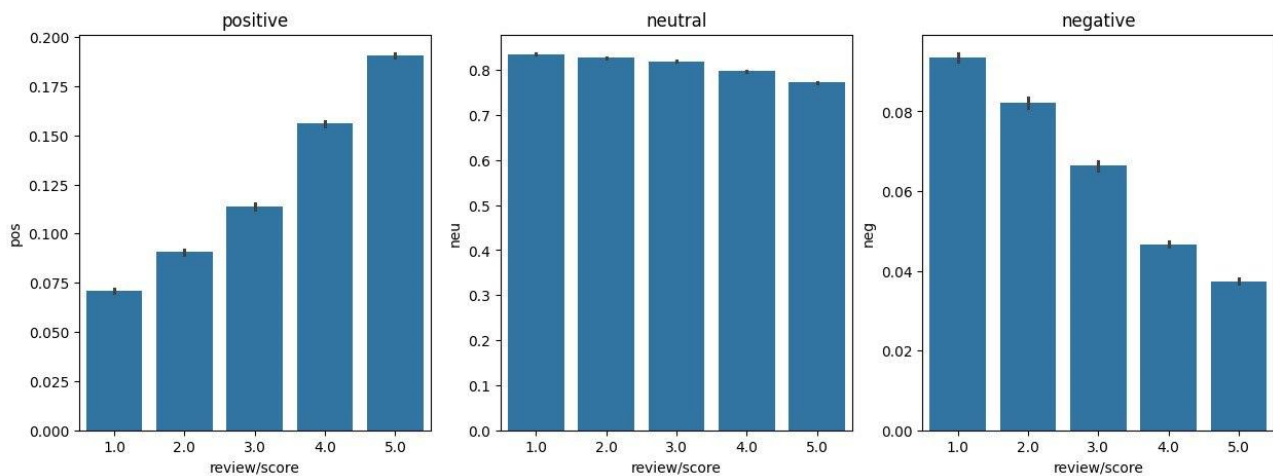
- o **Aperçu** : Ces méthodes impliquent l'entraînement d'algorithmes sur des ensembles de données étiquetées pour classifier les sentiments d'un texte. Parmi les algorithmes courants, on trouve Naive Bayes, les machines à vecteurs de support (SVM) et la régression logistique.
- o **Avantages** : Capacité à apprendre à partir des données et à capturer des schémas complexes. Elles offrent une précision plus élevée que les méthodes basées sur les lexiques.
- o **Inconvénients** : Nécessitent une quantité importante de données étiquetées pour l'entraînement. Elles peuvent ne pas bien se généraliser aux données inconnues. L'interprétation des modèles peut être difficile.
- o **Exemples** : Classificateurs Naive Bayes, SVM, forêts aléatoires.

3. **Approches basées sur l'apprentissage profond :**

- o **Aperçu** : Ces techniques exploitent les réseaux neuronaux tels que les réseaux neuronaux récurrents (RNN), les mémoires à long court terme (LSTM) et les transformateurs pour modéliser le sentiment d'un texte.

- o **Avantages** : Haute précision et capacité à capturer le contexte ainsi que les dépendances à long terme dans un texte. Elles permettent de gérer les nuances complexes du langage et les expressions idiomatiques.
- o **Inconvénients** : Exigent de vastes ensembles de données et des ressources computationnelles importantes pour l'entraînement. Les modèles complexes peuvent être difficiles à interpréter.
 - o **Exemples** : RNN, LSTM, BERT, GPT-3

Review	Classification
"I am absolutely thrilled with my new noise-cancelling headphones. The sound quality is crisp, the battery life is impressive, and they are incredibly comfortable to wear for long periods. The shipping was also surprisingly fast. I highly recommend this product to anyone in the market for new headphones!"	Positive
"The product arrived later than the estimated delivery date. When I finally received it, the color was slightly different from what was shown on the website, and the material feels cheaper than I anticipated. Overall, a disappointing experience."	Negative
"The package arrived on time and the product is as described. It functions as expected."	Neutral
"I ordered a medium, but it fits a bit more snugly than other brands I'm used to. The quality seems okay for the price."	Neutral
"The website was easy to navigate and the checkout process was straightforward. The product itself is decent, but not outstanding."	Neutral



3.5.2 analyse des résultats en comparaison avec les notation en étoiles:

"Plus la note est haute, plus le commentaire est positif."

Détail : Regardez le graphique de gauche ("positive"). On voit une augmentation constante de la hauteur des barres de gauche à droite (de la note 1 à la note 5). La barre pour la note 1 est très basse (peu de

positif), tandis que celle pour la note 5 est la plus haute (beaucoup de positif). Cela veut dire que lorsque les gens donnent une bonne note (comme 4 ou 5 étoiles), ils utilisent généralement des mots comme "excellent", "super", "j'adore", "parfait", etc. C'est une relation directe et forte.

"Plus la note est basse, plus le commentaire est négatif."

Détail : Regardez le graphique de droite ("negative"). Ici, c'est l'inverse : la hauteur des barres diminue constamment de gauche à droite. La barre pour la note 1 est très haute (beaucoup de négatif), tandis que celle pour la note 5 est très basse (très peu de négatif). Donc, quand les gens sont mécontents et donnent une mauvaise note (comme 1 ou 2 étoiles), ils ont tendance à utiliser des mots comme "horrible", "décevant", "problème", "mauvais", etc. C'est aussi une relation directe et forte, mais dans le sens opposé.

"Les commentaires sont souvent neutres, peu importe la note."

Détail : Regardez le graphique du milieu ("neutral"). Contrairement aux deux autres, les barres ici sont toutes très hautes et assez stables. Elles ne changent pas beaucoup de hauteur, qu'il s'agisse d'une note de 1 ou de 5. La proportion de "neutre" reste autour de 0,80 (ou 80%).

Qu'est-ce que ça signifie ? Cela veut dire que même dans un avis très positif ou très négatif, il y a une grande partie du texte qui n'est ni l'un ni l'autre. C'est du contenu descriptif, factuel, ou informatif. Par exemple :

"Le produit X a été livré en 3 jours." (neutre)

"Le bouton est situé sur le côté droit." (neutre)

"J'ai acheté ce téléphone pour ma fille." (neutre) Même si l'avis global est "Ce téléphone est génial, ma fille l'adore !" ou "Ce téléphone est nul, la batterie ne tient pas !", ces phrases neutres composent une partie importante du texte. Cela montre que les avis ne sont pas que des jugements émotifs, mais aussi des descriptions objectives.

Conclusion Générale

Ce projet, consacré à la classification de commentaires e-commerce via Python, a été une excellente mise en pratique des techniques de résolution de problèmes en informatique et d'analyse de sentiment. Le processus, qui a couvert le nettoyage des données textuelles, leur prétraitement, et l'application de modèles de machine learning, a permis de développer un classificateur capable de distinguer les avis positifs, négatifs et neutres avec une précision notable.

Sur le plan pratique, la valeur d'un tel système pour une plateforme e-commerce est considérable. Il constitue un outil puissant pour identifier rapidement les problèmes récurrents liés aux produits ou services, un levier pour améliorer l'expérience utilisateur en comprenant finement les attentes des clients, et enfin, une solution efficace pour automatiser à grande échelle le traitement des retours.

Finalement, ce travail constitue une base solide pour des développements futurs. Il ouvre la voie à des améliorations significatives, telles que l'intégration de modèles de deep learning plus performants comme BERT ou LSTM, l'extension du système pour traiter des commentaires multilingues, ou encore l'enrichissement fonctionnel par l'ajout d'un module de recommandation de produits.

References

- [1] D. Lamaute, J. Laganière, H. Chassé and I. Ben Kadhi, *l'entreprise*, Chenelière Éducation, 2023.
- [2] M. H. Nouhoum and SOW Ousmane , "Le rôle de l'information dans le processus de prise de décision," UNIVERSITE MOULOUD MAMMERI DE TIZI OUZOU , Tizi ousou, 2021.
- [3] S.-C. Chu, , , Lawrence Lunge and Ya van hui, "Evolution of e-commerce Web sites: A conceptual framework and a longitudinal study," *research gate*, 2007.
- [4] R. V. G. d. M. W. F. François Scheid, *Le marketing digital: Développer sa stratégie numérique*, Eyrolles, 2019.
- [5] c. mohamed and s. zebaree, "A Review of Customer Engagement Strategies in Digital Marketing for E-Business Growth," *researchgate*, 2025.
- [6] W. R. V. Kumar, "Customer Relationship Management Concept, Strategy, and Tools Third," *Springer Berlin, Heidelberg*, 2018.
- [7] C. Odionu, B. Bristol-Alagbariya and R. Okon, "Big data analytics for customer relationship management: Enhancing engagement and retention strategies," *researchgate*, 2024.
- [8] M. N. S. & N. C. Adjei, "The influence of C2C communications in online brand communities on customer purchase behavior," *springer*, 2010.
- [9] H. Allam, L. Makubvure, G. Kwadwo and B. Gyamfi, "Text Classification: How Machine Learning Is Revolutionizing Text Categorization," *researchgate*, 2024.
- [10] B. PRIEUR, *Traitement automatique du langage naturel avec Python*, eni.
- [11] N. Gupta and R. Agrawal, Chapter 1 - Application and techniques of opinion mining, *scienceDirect*, 2020.
- [12] S. Mandal and S. Gupta, "A Lexicon-based text classification model to analyse and predict sentiments from online reviews," *IEEE*, 2017.
- [13] J. Q. a, "Mining customer requirements from online reviews: A product improvement perspective," *scienceDirect*, 2016.
- [14] R. Decker and M. Trusov, "Estimating aggregate consumer preferences from online product reviews," *ScienceDirect*, 2010.
- [15] L. O. Berko, "Harnessing Big Data, Machine Learning, and Sentiment Analysis to Optimize Customer Engagement, Loyalty, and Market Positioning," *Researchgate*, 2025.
- [16] Y. h. hu and K. chen, "Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings," *sciencedirect*, 2016.
- [17] a. t. Sameh Al-Natour, "A comparative assessment of sentiment analysis and star ratings for consumer reviews," *sciencedirect*, 2020.
- [18] "What drives consumers to spread electronic word of mouth in online consumer-opinion platforms," *scienceDirect*.
- [19] "What drives consumers to spread electronic word of mouth in online consumer-opinion platforms," *sciencedirect*.
- [20] "Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings - ScienceDirect," *science direct*.
- [21] amazon reviews dataset, stanford edu <https://snap.stanford.edu/data/web-Amazon.html>