

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABBES LAGHROUR KHENCHELA
FACULTE DES SCIENCES ET DE LA TECHNOLOGIE



Département de mathématique et informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de Master (L.M.D)

Spécialité : Informatique

Option : sécurité et Technologies web

THEME

Analyse de la satisfaction vis-à-vis des traitements par fouille d'opinions des patients

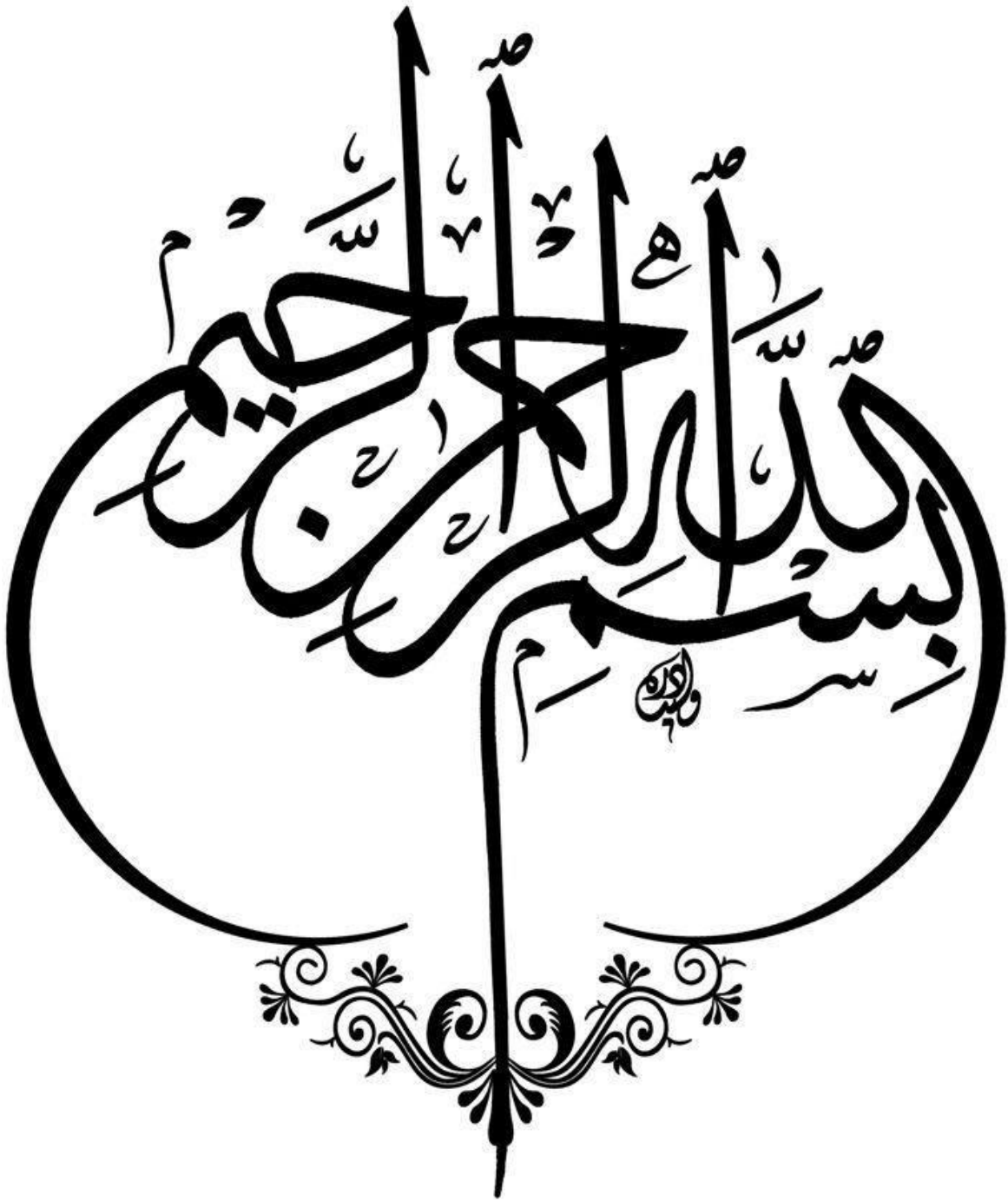
Réalisé par :

Narimane Chafia Senoussi
GASMI Nesrine

Encadré par :

Dr. Hichem Rahab

Année Universitaire : 2022/2023



Remerciement

Tout d'abord, nous remercions Dieu tout-puissant, qui nous a donné des connaissances que nous ne savions pas. Nous le louons beaucoup digne de son grandeur, et beaucoup de bénédictions.

Nous exprimons nos sincères remerciement et notre gratitude au docteur HICHEM RAHAB, qui nous avons été honorés de proposer et d'encadrer notre mémoire. Et il ne nous a épargné aucune information ni aucun conseil. Nous sommes également heureux d'exprimer nos remerciements et notre gratitude aux membres du jury pour avoir accepté l'évaluation de notre mémoire.

A la fin du parcours universitaires, nous exprimons nos remerciements et notre gratitude aux honorables enseignants du département de mathématiques et informatique de l'université ABBES LAGHROUR KHENCHELA.

Nous remercions également tous ceux qui nous ont aidés de près ou de loin à mener à bien ce travail.

Nesrine & Narimane

Hesrine 

*Je suis profondément reconnaissant envers mes parents, mes frères
pour leur soutien constant et les sacrifices qu'ils ont faits pour moi.*

Leur présence et leurs encouragements ont été des éléments indispensables dans ma réussite.

Je suis également reconnaissant envers mes amis, qui ont été le plus beau cadeau.

Leur amitié sincère et leur soutien inébranlable ont été une source de joie et de force tout au long de mon parcours.

Narimane

Je dédie ce travail à :

A ma mère,

A mon père,

A mon frère et mes sœurs,

A tous mes chers amis,

A toute ma famille,

*A tous mes professeurs qui m'ont enseigné au primaire, au collège, au
lycée et*

À l'université

Résumé

La recherche dans le domaine de l'analyse des sentiments médicaux est un domaine de recherche en plein essor qui vise à extraire des informations sur les opinions, les émotions et les attitudes des patients, des professionnels de la santé et du grand public concernant les traitements médicaux, les médicaments, les services de santé et d'autres aspects du domaine médical. Nous nous concentrons sur l'analyse des sentiments médicaux en utilisant deux approches principales ; l'approche lexicale et l'apprentissage automatique. Pour collecter l'ensemble de données, Data_Quality_Cares, nous avons utilisés des questionnaires d'opinion pour le recueil des avis des patients. Les deux approches utilisées sont évaluées en utilisant les deux paramètres l'exactitude (accuracy) et le score F1. Le modèle SVM donne le meilleur résultat.

Mots clés : Sentiments médicaux, analyse des sentiments, extraction d'opinions, apprentissage automatique, SVM, DT, NB. Apprentissage automatique, approche lexical.

Abstract

Research in the field of medical sentiment analysis is a booming field of research that aims to extract information on the opinions, emotions and attitudes of patients, health professionals and the general public regarding medical treatments, medicines, health services and other aspects of the medical field. We focus on the analysis of medical feelings using two main approaches; the lexical approach and machine learning. To collect the dataset, Data_Quality_Cares, we used opinion questionnaires to collect patient opinions. The two approaches used are evaluated using the two parameters; the accuracy and The F1 score. The SVM model gives the best result.

Keywords: Medical sentiments, Sentiment Analysis, opinion mining, Machine Learning, SVM, DT, NB, Lexical Approach.

ملخص

تحليل المشاعر الطبية هو مجال بحث مزدهر يهدف إلى استخلاص معلومات حول آراء ومشاعر واتجاهات المرضى والمهنيين الصحيين بشأن العلاجات الطبية والأدوية والخدمات الصحية وجوانب أخرى في المجال الطبي. نركز على تحليل المشاعر الطبية باستخدام نهجين رئيسيين؛ النهج اللغوي وتعلم الآلة. للحصول على البيانات لتشكيل المدونة (مجموعة البيانات) "Data_Quality_Cares"، تم اللجوء إلى استخدام استبيانات تقييم الرأي لجمع آراء المرضى. تم تقييم النهجين المستخدمين باستخدام معايير التقييم المعروفة باسم معيار الدقة (Accuracy) ومعيار F1_Score. أعطى نموذج SVM أفضل نتيجة.

الكلمات المفتاحية: المشاعر الطبية، تحليل المشاعر، استخراج الآراء، تعلم الآلة، SVM، DT، NB. التعلم الآلي، النهج اللغوي.

Tables des matières

Table des matières

Introduction Générale.....	2
Contexte et objectif	2
Contributions.....	3
Méthodologie	2
Problématique.....	2
Structure du mémoire.....	3
Chapitre I.....	5
1. Introduction.....	6
2. Traitement Automatique de la Langue Naturelle (TALN).....	6
2.1 Définition	7
2.2 Explication de l'analyse des sentiments médicaux	8
2.3 L'importance et l'objectif de l'analyse des sentiments médicaux	9
2.3.1 Améliorer la communication avec les patients	10
2.3.2 Améliorer la performance des employés et services	10
2.4 Les niveaux d'analyse des sentiments.....	10
2.4.1 Niveau du document	10
2.4.2 Niveau de la phrase	11
2.4.3 Niveau des aspects	11
2.5 Types d'analyse de sentiments	11
2.5.1 Analyse fine des sentiments	11
2.5.2 Détection d'émotion.....	12
2.5.3 L'analyse des sentiments basée sur les aspects.....	12
2.6 Sources de donnée de sentiment médicaux.....	12
3. Les approches de l'analyse des sentiments.....	12
3.1 L'approche basée sur l'apprentissage automatique	13
3.2 L'approche basée sur lexicque	14
3.3 Approche hybride.....	16
4. Processus général de l'Analyse des Sentiments	16
5. L'Analyse des Sentiments en langue Arabe.....	16
5.1 La richesse de la langue arabe	17
5.2 Complexité de la langue arabe	18
6. Conclusion	18
Chapitre II.....	19

Tables des matières

<i>Les approches d'analyse des sentiments</i>	19
1. Introduction.....	20
2. L'intelligence artificielle	20
3. Approche basée sur L'apprentissage automatique	21
3.1 Principe de fonctionnement de l'apprentissage automatique	22
3.2 Les types d'apprentissage automatique.....	23
3.2.1 L'apprentissage supervisé.....	23
3.2.2 L'apprentissage non supervisé.....	28
4. Approche lexicale.....	28
4.1 Les types des approches basé sur lexique	29
4.1.1 Approche basée sur le dictionnaire.....	29
4.1.2 Approche basée sur le corpus	29
4.2 Algorithme de l'approche lexicale.....	30
4.3 Comparaison entre les deux approches	30
5. Conclusion	31
Chapitre III	32
<i>Implémentation et expérimentation</i>	32
1. Introduction.....	33
2. L'implémentation.....	33
2.1 Les ressource	33
2.1.1 L'ensemble de données	33
2.1.2 Collecte de données	33
2.1.3 Création de l'ensemble de données Data_Quality_Cares	36
2.1.4 Annotation de données	36
2.1.5 L'agrément entre annotateurs Inter Annotators Agreement (IAA)	37
2.1.6 Description de l'ensemble des données	39
2.2 Le lexique utilisé.....	40
2.2.1 Obtention du lexique.....	40
2.2.2 Enrichissement de lexique	41
3. Architecture de notre application.....	42
3.1 Importation de l'ensemble de données	44
3.2 Prétraitement des données	44
3.2.1 Prétraitement préliminaire	44
3.2.2 Séparation de mots (Tokenisation)	44

Tables des matières

3.2.3 Enracinement (Stemming)	45
3.3.1 Approche basée sur l'apprentissage automatique.....	46
3.3.1.1 Extraction des caractéristiques	46
3.3.1.2 La classification.....	47
3.3.2 Approche lexicale	47
3.3.2.1 Création de dictionnaire	47
3.3.2.2 Règles de classification	47
4. Mesures d'évaluations	47
5. Expérimentation	49
5.1 Approche apprentissage automatique.....	49
5.1.1 La division (Train/Test Split).....	49
5.1.2 L'évaluation de résultats.....	49
5.2. Approche à base de lexique.....	53
5.2.1 Attribution de polarité.....	53
5.2.2 Agrégation des polarités.....	53
5.2.3. L'évaluation de résultats.....	53
6. L'application développée	54
6.1 Fenêtre d'analyse	55
6.2 Fenêtre de résultats	55
6.3 Exemple sur l'utilisation de l'application.....	56
7. Conclusion	58
Conclusion générale.....	60
Annex	62
Certificat de stage.....	62
Bibliographies.....	68

Liste des figures

Table des figures

Figure 1 : Les champs de recherche et applications de TALN.....	7
Figure2 : Schéma présenté les sentiments.....	8
Figure 3 : Les différentes approches d'Analyse des Sentiments.....	13
Figure 4 : Les étapes de l'approche basée sur l'apprentissage automatique	14
Figure 5 : L'organigramme de l'approche basée sur lexicale	15
Figure 6 : Processus général de l'analyse des sentiments.....	16
Figure 7 : La différence de la recherche entre l'arabe et l'anglais	17
Figure 8 : Les approches d'analyse des sentiments	20
Figure 9 : Intelligence artificiel (IA).....	21
Figure 10 : La relation entre apprentissage automatique et l'intelligence artificielle	22
Figure 11 : Les branches de l'apprentissage automatique.....	23
Figure 12 : Exemple d'un SVM	25
Figure 13 : schéma d'un arbre de décision.....	26
Figure 14 : Exemple de prédiction 3 clusters	28
Figure 15 : Réponse de patient sur questionnaire (Exemple 1).....	34
Figure 16 : Réponse de patient sur questionnaire (Exemple 2).....	35
Figure 17 : Réponse de patient sur questionnaire (Exemple 3).....	35
Figure 18 : Présentation de Data_Quality_Cares.....	40
Figure 19 : Les principales étapes de l'analyse du sentiment	43
Figure 20 : Exemple de séparation de mots d'un commentaire	45
Figure 21 : Exemple de racinisation des mots de commentaires.....	45
Figure 22 : Comparaison entre les approches utilisées.	54
Figure 23 : Fenêtre d'analyse	55
Figure 24 : Fenêtre de résultat	56
Figure 25 : Saisir de commentaire	56
Figure 26 : Choix de méthode	57
Figure 27 : Résultat d'analyse.....	57

Liste des tableaux

Liste des tableaux

Tableau 1 : Les avantages et les inconvénients d'approches (Lexique et apprentissage automatique) ..	30
Tableau 2 : Échantillons sur le vote entre les annotateurs.....	37
Tableau 3 : L'agrément entre annotateurs IAA.....	37
Tableau 4 : Statistiques de Data_Quality_Cares.....	39
Tableau 5 : Description de l'ensemble des données (Data_Quality_Cares).....	39
Tableau 6 : Statistiques de lexique télécharger.....	41
Tableau 7 : Statistique du lexique enrichi	41
Tableau 8 : Exemple des mots de lexique	41
Tableau 9: Matrice de confusion	48
Tableau 10 : Matrice de confusion de SVM sur Data_Quality_Cares (sans prétraitement).....	50
Tableau 11 : Performances pour SVM sur Data_Quality_Cares (sans prétraitement).....	50
Tableau 12 : Matrice de confusion de SVM sur Data_Quality_Cares (avec prétraitement)	50
Tableau 13 : Performances pour SVM sur Data_Quality_Cares (avec prétraitement)	50
Tableau 14 : Matrice de confusion de NB sur Data_Quality_Cares (sans prétraitement).....	51
Tableau 15 : Performances pour NB sur Data_Quality_Cares (sans prétraitement).....	51
Tableau 16 : Matrice de confusion de NB sur Data_Quality_Cares (avec prétraitement)	51
Tableau 17 : Performances pour NB sur Data_Quality_Cares (avec prétraitement)	51
Tableau 18 : Matrice de confusion de DT sur Data_Quality_Cares (sans prétraitement).....	52
Tableau 19 : Performances pour DT sur Data_Quality_Cares (sans prétraitement).....	52
Tableau 20 : Matrice de confusion de DT sur Data_Quality_Cares (avec prétraitement)	52
Tableau 21 : Performances pour DT sur Data_Quality_Cares (avec prétraitement)	52
Tableau 22 : Matrice de confusion de Lexique sur Data_Quality_Cares.....	53
Tableau 23 : Performances pour Lexique sur Data_Quality_Cares.....	53
Tableau 24 : Performances des quatre classifieurs	54

Introduction

Générale

Introduction générale

Introduction Générale

Contexte et objectif

La technologie a révolutionné notre monde à une vitesse fulgurante, transformant nos vies de multiples façons. Que ce soit grâce aux smart phones, à l'internet, à l'intelligence artificielle ou à d'autres avancées, la technologie a ouvert de nouvelles possibilités pour connecter les gens à travers le globe et faciliter l'accès à l'information. Une des caractéristiques les plus puissantes de la technologie moderne est qu'elle donne à chacun la possibilité de partager son opinion. Les réseaux sociaux, les blogs et les forums en ligne permettent à tous d'exprimer librement leurs idées et de participer à des débats ouverts.

L'objectif de ce mémoire est d'exploiter les techniques d'analyse des sentiments dans le domaine médical. Dans un contexte où la santé et le bien-être des patients sont d'une importance primordiale, comprendre les émotions et les sentiments associés aux expériences médicales peut jouer un rôle essentiel. L'objectif principal est de développer des techniques et des outils d'analyse des sentiments qui permettent d'évaluer de manière précise et efficace les réactions émotionnelles des patients vis-à-vis de leur expérience médicale. Cela peut inclure l'analyse des données textuelles provenant de différents documents médicaux, tels que les dossiers des patients, les journaux de bord cliniques et les commentaires des patients en ligne. Grâce à cette analyse des sentiments médicaux, il sera possible d'identifier les aspects positifs et négatifs des soins de santé, de détecter les préoccupations des patients et de formuler des recommandations pour améliorer l'expérience globale des patients dans le domaine médical.

Problématique

L'analyse des sentiments médicaux est un domaine de recherche en plein essor qui vise à extraire des informations sur les opinions, les émotions et les attitudes des patients, des professionnels de la santé et du grand public concernant les traitements médicaux, les médicaments, les services de santé et d'autres aspects du domaine médical. Cette analyse des sentiments peut fournir des informations précieuses pour les décideurs de santé, les chercheurs et les praticiens afin d'améliorer la prise de décision clinique, d'optimiser les soins aux patients et de mieux comprendre les préoccupations des patients. Ce travail vient pour soulever un ensemble de problématiques. Quelles sont les meilleures approches et méthodes d'analyse de sentiments médicaux dans le contexte des EPSPs de Khenchela comme échantillon des établissements de la santé en Algérie ? et quelle sont les meilleurs paramètres et prétraitement qui peuvent améliorer les résultats ?

Méthodologie

Pour collecter les données nécessaires, des questionnaires ont été administrés aux participants, leur demandant de fournir des réponses détaillées sur leurs expériences et leurs émotions liées à leurs expériences médicales. La méthodologie utilisée dans cette étude repose sur deux approches de l'analyse des sentiments médicaux ; à savoir, l'approche lexicale et

Introduction générale

l'apprentissage automatique. La première approche utilisée repose sur l'utilisation d'un lexique médical spécifique contenant des mots associés à des sentiments positifs et négatifs. Quand à la deuxième approche, elle consiste à l'entraînement de modèles à partir de données annotées pour prédire les sentiments dans les textes médicaux en utilisant les algorithmes de classification SVM, NB et DT.

Contributions

Dans le cadre de notre mémoire sur l'analyse des sentiments médicaux, nous avons apporté une contribution significative en utilisant deux approches distinctes. Tout d'abord, nous avons constitué « Data_Quality_Cares » un corpus de données spécifiquement collecté dans le domaine médical, ce qui nous a permis d'obtenir des informations précises et pertinentes. Ensuite, nous avons enrichi un lexique, obtenu du site Kaggle, en incorporant des termes d'opinion médicaux spécifiques. Ensuite nous avons adapté les paramètres des algorithmes d'apprentissage automatique existants pour tenir compte de la terminologie médicale. Enfin, nous avons procédé à une évaluation comparative des performances des deux approches, en tenant compte des mesures de précision, de rappel, d'exactitude (accuracy) et de F-mesure. Les résultats obtenus ont fourni des conseils importants pour l'analyse des sentiments médicaux, et ont démontré l'efficacité des approches proposées.

Structure du mémoire

Dans ce mémoire, nous nous concentrons sur l'analyse des sentiments médicaux en utilisant deux approches principales : l'approche lexicale et l'apprentissage automatique. Ces deux approches permettent de traiter les données textuelles provenant de diverses sources, telles que les réseaux sociaux, les forums de discussion en ligne, les journaux médicaux et les dossiers de patients. Dans notre travail nous nous sommes limités des questionnaires d'opinion pour le recueil des avis des patients.

Le premier chapitre de ce mémoire se concentre sur la fouille de sentiments médicaux. Il examine les différentes techniques de prétraitement des données textuelles, l'identification des entités médicales, et les niveaux d'analyse des sentiments. Il présente également les types d'analyse de sentiments, les sources de données de sentiments médicaux, et les approches basées sur l'apprentissage automatique, le lexique et l'approche hybride. Ce chapitre souligne l'importance de l'analyse des sentiments médicaux dans l'amélioration de la communication avec les patients et des services médicaux.

Le deuxième chapitre présente les approches d'analyse des sentiments utilisées dans le domaine médical. Nous examinons en détail l'approche lexicale, qui repose sur l'utilisation de lexiques médicaux spécifiques pour évaluer les sentiments exprimés dans les textes médicaux. Nous explorons également les techniques d'apprentissage automatique, telles que les classificateurs bayésiens, l'arbre de décision et SVM, qui permettent de détecter automatiquement les sentiments médicaux à partir des textes.

Introduction générale

Le troisième chapitre de ce mémoire se concentre sur l'implémentation et la réalisation des approches d'analyse des sentiments médicaux. Nous détaillons les étapes clés de la mise en œuvre, telles que la collecte des données, la création des ensembles de données annotées (notamment Data_Quality_Cares), la sélection des attributs pertinents et l'évaluation des modèles. De plus, nous présentons les résultats expérimentaux obtenus lors de cette mise en pratique. Enfin, nous abordons également l'étape de création de notre application avec une interface graphique conviviale et facile à utiliser.

En conclusion, ce mémoire vise à fournir une analyse approfondie des différentes approches d'analyse des sentiments médicaux, en mettant l'accent sur les approches lexicales et d'apprentissage automatique. Elle offre une compréhension claire des méthodologies utilisées dans ce domaine et des applications potentielles pour améliorer les soins de santé. Les résultats de cette recherche peuvent contribuer à l'avancement de l'analyse des sentiments médicaux et à l'amélioration de la prise de décision dans le domaine médical.

Chapitre I

*Fouille de sentiments
médicaux.*

Chapitre I Fouille de sentiments médicaux

1. Introduction

L'analyse de sentiment médicale est une méthode d'analyse de données qui permet de comprendre les opinions et les perceptions des individus en matière de soins médicaux. Elle peut être utilisée pour extraire des informations pertinentes à partir de diverses sources de données, telles que les réseaux sociaux, les forums de discussion et les commentaires des patients, afin d'identifier les tendances, les besoins non satisfaits et les préoccupations de la communauté médicale. Cette approche peut aider les professionnels de la santé à mieux comprendre les perspectives des patients et à améliorer la qualité des soins médicaux en identifiant les domaines qui nécessitent une attention particulière.

Dans ce chapitre, nous allons explorer les différentes techniques d'analyse de sentiment médicale, en mettant l'accent sur les avantages et les limites de chaque méthode. Nous allons également discuter des différents cas d'utilisation de l'analyse de sentiment médicale dans le domaine de la santé, et examiner comment cette technique peut être appliquée pour améliorer la qualité des soins médicaux.

2. Traitement Automatique de la Langue Naturelle (TALN)

Le traitement automatique du langage naturel est un domaine de recherche qui se situe à l'intersection entre la linguistique et l'informatique. Il utilise diverses techniques telles que l'apprentissage automatique, l'apprentissage profond, et les réseaux de neurones artificiels pour permettre à la machine de comprendre et de générer le langage humain, qu'il soit écrit ou oral. Le TALN est considéré comme l'un des domaines les plus essentiels de l'intelligence artificielle car il est crucial pour améliorer les capacités de communication des machines et des dispositifs. [1]

TALN est un domaine de recherche qui regroupe de nombreuses disciplines aux objectifs et aux méthodes de traitement variés. Il se divise en trois axes principaux, à savoir la sémantique, l'extraction d'informations et la syntaxe, chacun comprenant plusieurs sous-domaines. Certains de ces sous-domaines sont représentés dans **Figure 1**[2].

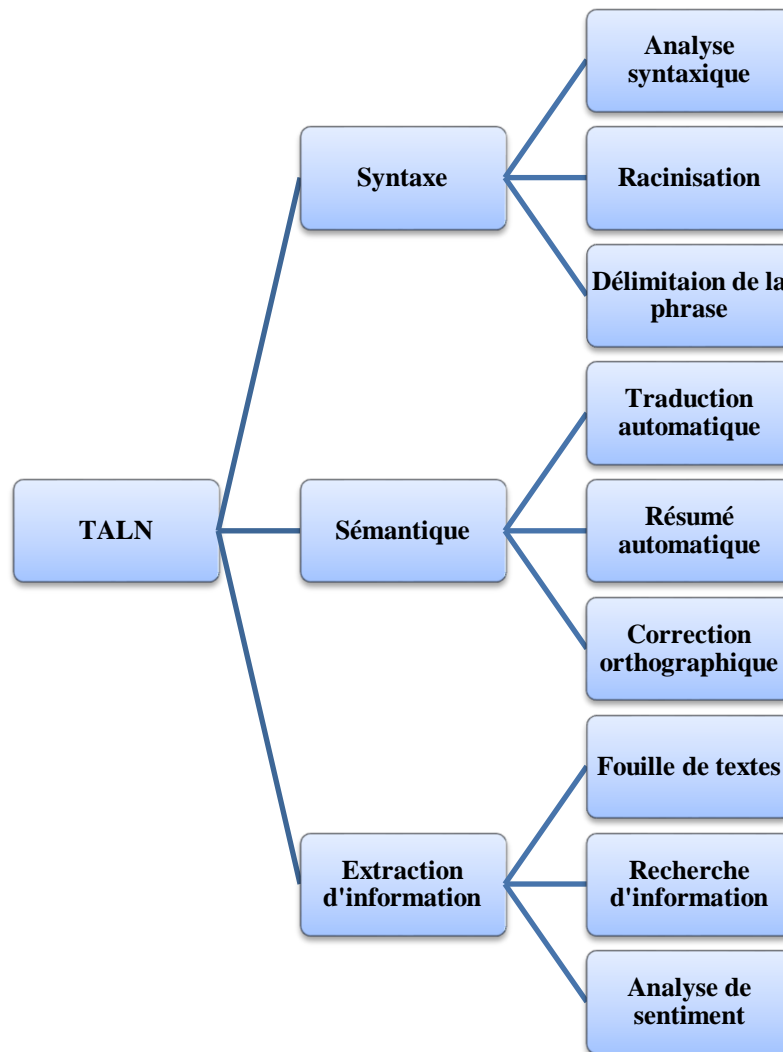


Figure 1 : Les champs de recherche et applications de TALN

2.1 Définition

L'Analyse des sentiments est un domaine d'application du TALN qui se concentre sur l'extraction et la classification des émotions (positives, négatives, neutres, etc.) dans les données textuelles. Elle est souvent associée à la linguistique informatique, notamment le TALN. Grâce à l'utilisation de techniques d'analyse de texte, l'analyse des sentiments permet de comprendre les opinions des patients sur la qualité des soins ou les services, en examinant les conversations et les commentaires en ligne. Cette approche est utile pour recueillir des informations précieuses sur la perception de la clientèle et aider les sociétés médicales à prendre des décisions éclairées pour améliorer leurs produits et services [3].



Figure2 : Schéma présenté les sentiments

2.2 Explication de l'analyse des sentiments médicaux

Dans le domaine médical, l'analyse des sentiments joue un rôle crucial en aidant les prestataires de soins de santé à obtenir un avantage concurrentiel par rapport à leurs concurrents. En comprenant les opinions et les commentaires des patients, les prestataires peuvent améliorer leurs services en fonction de leurs besoins et de leurs attentes.

L'analyse des sentiments des patients fournit des informations précieuses sur le type de traitement dont les patients ont besoin, le niveau de service qu'ils souhaitent recevoir et le type d'établissement médical vers lequel ils se tournent pour leur traitement. En comprenant ces sentiments, les prestataires peuvent adapter leurs services pour mieux répondre aux besoins des patients et offrir une expérience de soins de santé plus satisfaisante.

Grâce à l'analyse des sentiments, les prestataires de soins de santé peuvent identifier les domaines qui nécessitent des améliorations et mettre en œuvre des mesures correctives pour optimiser la satisfaction des patients. Cela peut inclure des changements dans les politiques internes, des formations pour le personnel médical, des améliorations des infrastructures ou des ajustements dans la communication avec les patients.

Chapitre I *Fouille de sentiments médicaux*

En résumé, l'analyse des sentiments dans le domaine médical permet aux prestataires de soins de santé de comprendre les besoins et les attentes de leurs patients, de se différencier de leurs concurrents et d'améliorer en permanence la qualité des soins qu'ils offrent. Cela conduit à une meilleure satisfaction des patients, à une fidélisation accrue et à une réputation positive pour les prestataires de soins de santé.

2.3 L'importance et l'objectif de l'analyse des sentiments médicaux

L'objectif de l'analyse des sentiments est de développer des outils automatisés capables de détecter des informations subjectives pour produire des connaissances structurées et exploitables [4]. L'analyse des sentiments des patients dans le domaine médical est essentielle pour améliorer les résultats cliniques à grande échelle. Les patients expriment des sentiments forts à l'égard des soins médicaux qu'ils reçoivent, ce qui se manifeste par des réactions positives ou négatives lors de leurs interactions avec les prestataires de soins de santé et les hôpitaux.

Les patients ont des sentiments très forts en ce qui concerne les soins médicaux qu'ils reçoivent, et cela se traduit par des réactions positives ou négatives lors de leurs interactions avec les prestataires de soins de santé ou les hôpitaux. C'est pourquoi l'analyse des sentiments dans le domaine de la santé est extrêmement précieuse. Les connaissances acquises grâce à cette analyse permettent aux prestataires de combler les lacunes de communication entre les établissements et les patients, optimisant ainsi l'expérience du patient et améliorant les résultats commerciaux à grande échelle. L'importance de l'analyse des sentiments médicaux est donc expliquée plus en détail dans ce qui suit.

En comprenant les opinions et les commentaires des patients grâce à cette analyse des sentiments, les prestataires de soins de santé peuvent ajuster leurs pratiques et leurs protocoles de traitement afin d'obtenir des résultats cliniques plus efficaces et d'améliorer la qualité des soins prodigués. Cette approche vise à optimiser l'expérience du patient, à réduire les complications et à promouvoir une meilleure santé à grande échelle.

L'analyse des sentiments dans le domaine de la santé revêt une importance particulière, car elle permet aux prestataires de combler les lacunes de communication entre les établissements et les patients. Les informations recueillies grâce à cette analyse permettent d'améliorer l'expérience du patient et d'obtenir des résultats cliniques améliorés à grande échelle, en augmentant la satisfaction des patients et en répondant de manière plus précise à leurs besoins et attentes en matière de soins médicaux.

Chapitre I *Fouille de sentiments médicaux*

2.3.1 Améliorer la communication avec les patients

Grâce à l'analyse des sentiments appliqués aux soins de santé, les prestataires peuvent identifier les éventuelles causes d'insatisfaction des patients et comprendre leur point de vue ainsi que leurs priorités en matière de communication.

Une des forces de l'analyse des sentiments est sa capacité à trier les commentaires des patients en différentes catégories, permettant ainsi une analyse minutieuse des facteurs qui influencent la communication entre les prestataires de soins de santé et les patients, tels que la communication avec les médecins, les infirmières et la réponse du personnel hospitalier.

2.3.2 Améliorer la performance des employés et services

L'utilisation de l'analyse des sentiments dans le secteur de la santé permet aux prestataires de mieux comprendre leurs points forts ainsi que les domaines où ils peuvent rencontrer des difficultés en termes de service aux patients, en particulier en ce qui concerne les stratégies de communication. En classifiant les commentaires des patients en fonction des personnes, des lieux et des processus, les établissements de santé peuvent évaluer quantitativement leur expérience et identifier les domaines nécessitant des améliorations. Par exemple, en analysant des commentaires tels que.

- "مرضة لطيفة ومهتمة" (Positif),
- "خدمة عملاء جيدة ولكن يجب تحسين نظام الحجوزات" (Neutre)
- "انتظرت لساعات طويلة للتحدث مع موظف الاستقبال" (Négatif)

Les prestataires peuvent recueillir les commentaires de manière fréquente et intense afin d'identifier les domaines d'excellence et ceux qui nécessitent des améliorations [5].

2.4 Les niveaux d'analyse des sentiments

Comme énoncé précédemment, l'objectif de l'analyse des sentiments est de créer des outils automatisés capables d'extraire des informations subjectives à partir de textes en langage naturel. Lorsque l'on applique l'analyse des sentiments, la première étape consiste à déterminer l'objet à analyser dans le contexte de l'étude. En général, l'analyse des sentiments peut être abordée à trois niveaux principaux.

2.4.1 Niveau du document

L'objectif de l'analyse des sentiments est de déterminer la polarité d'un texte d'opinion, c'est-à-dire s'il exprime une opinion globale positive, négative ou neutre sur un sujet donné. Par exemple, dans le cadre d'une revue de soins de santé, le système doit déterminer si le texte exprime une opinion globalement positive, négative ou neutre sur les soins reçus. L'hypothèse

Chapitre I *Fouille de sentiments médicaux*

de base est que l'ensemble du texte ne concerne qu'une seule entité (dans cet exemple, les soins de santé) et exprime une seule opinion [4].

2.4.2 Niveau de la phrase

Détermine la polarité de chaque phrase contenue dans un texte. L'hypothèse est que chaque phrase dans le texte exprime une opinion unique sur une entité unique [3].

2.4.3 Niveau des aspects

Ce niveau d'analyse est plus détaillé que les autres car il considère que toute opinion est composée d'un sentiment et d'une cible. Prenons l'exemple de l'avis d'un patient sur la qualité des soins : "Le personnel est très compétent et attentionné, mais les temps d'attente sont trop longs et la nourriture n'est pas très bonne". Cette phrase évalue trois aspects : le personnel (positif), les temps d'attente (négatif) et la nourriture (négatif) [3].

2.5 Types d'analyse de sentiments

Il existe de nombreux types d'analyses de sentiments allant des systèmes qui se concentrent sur la classification de la polarité (positif, négatif, neutre) aux systèmes qui détectent des émotions (en colère, heureux, triste, etc.) ou identifient des intentions (par exemple, intéressé, pas intéressé [1].

2.5.1 Analyse fine des sentiments

Plutôt que de se limiter à des catégories de phrases positives, négatives ou neutres, nous pouvons adopter une échelle de polarité plus fine qui comprend les catégories suivantes.

- Très positive
- Positive
- Neutre
- Négative
- Très négative

Certains systèmes vont même plus loin en distinguant des classifications de polarité en fonction de l'émotion associée au sentiment positif ou négatif. Par exemple, ils peuvent identifier des émotions telles que la colère, la tristesse ou l'inquiétude pour les sentiments négatifs, et le bonheur, l'amour ou l'enthousiasme pour les sentiments positifs [1].

Chapitre I *Fouille de sentiments médicaux*

2.5.2 Détection d'émotion

La détection des émotions est d'identifier des émotions telles que le bonheur, la frustration, la colère, la tristesse, etc. De nombreux systèmes de détection d'émotions utilisent des lexiques de sentiments (des listes d'émotions) ou des algorithmes d'apprentissage automatique avancés pour accomplir cette tâche [1].

2.5.3 L'analyse des sentiments basée sur les aspects

Cette approche dépasse la simple classification d'un texte en tant que positif ou négatif. Elle permet d'analyser le texte afin de détecter différents aspects et d'attribuer un sentiment spécifique à chacun d'entre eux. Cette approche offre des résultats plus détaillés, captivants et précis, car elle examine attentivement les informations contenues dans le texte.

2.6 Sources de donnée de sentiment médicaux

Des exemples de sources qui pourraient être utilisées pour l'analyse des sentiments médicaux, Pour les laboratoires et société pharmaceutique et parapharmaceutique.

- Les commentaires de patients sur les sites web de santé.
- Les publications sur les réseaux sociaux et les forums médicaux.
- Les notes et commentaires des patients dans les dossiers médicaux électroniques.
- Les enquêtes de satisfaction des patients.

En ce qui concerne le sujet des soins, les sources pour l'analyse des sentiments pourraient inclure pour les hôpitaux et les établissements hospitaliers.

- Études de recherche académique.
- Plateformes d'avis et de commentaires en ligne.
- Systèmes de gestion des relations avec les patients.
- Enquêtes et questionnaires.
- Collaboration avec des experts en analyse de données.

3. Les approches de l'analyse des sentiments

En analyse des sentiments, les textes sont souvent classés en termes de positivité ou de négativité, mais il peut également exister des classes multiples incluant la neutralité. Les approches de classification des sentiments peuvent être divisées en trois grandes catégories (l'approche d'apprentissage automatique, l'approche basée sur les lexiques et l'approche

hybride). Chacune de ces approches peut ensuite être subdivisée en sous-catégories comme l'illustre Figure 3[3].

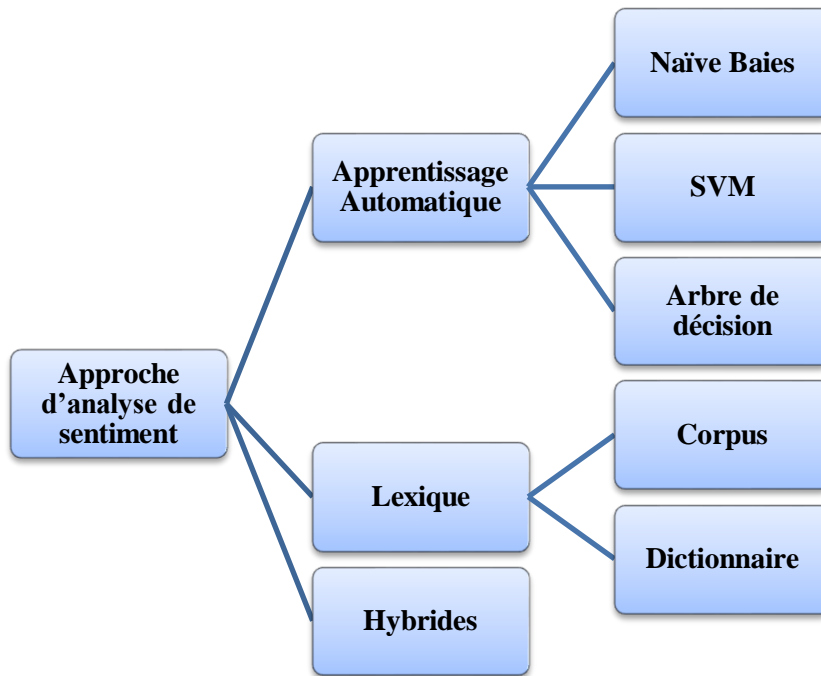


Figure 3 : Les différentes approches d'Analyse des Sentiments

3.1 L'approche basée sur l'apprentissage automatique

L'approche de l'apprentissage automatique permet aux ordinateurs d'agir de manière autonome, sans avoir besoin d'être programmés. Les programmes informatiques exploitent les données disponibles pour identifier des modèles, adapter leur comportement et prendre des décisions intelligentes. Pour la classification des sentiments, cette méthode se fonde sur des techniques d'apprentissage automatique renommées appliquées au texte [1].

Dans la plupart des cas, l'analyse des sentiments est considérée comme un problème de classification dans lequel un texte est présenté à un classificateur qui détermine sa catégorie correspondante, que ce soit positive, négative ou neutre (si l'analyse de polarité est impliquée). Les étapes de l'apprentissage automatique présenté dans la figure au-dessous Figure 4 [5].

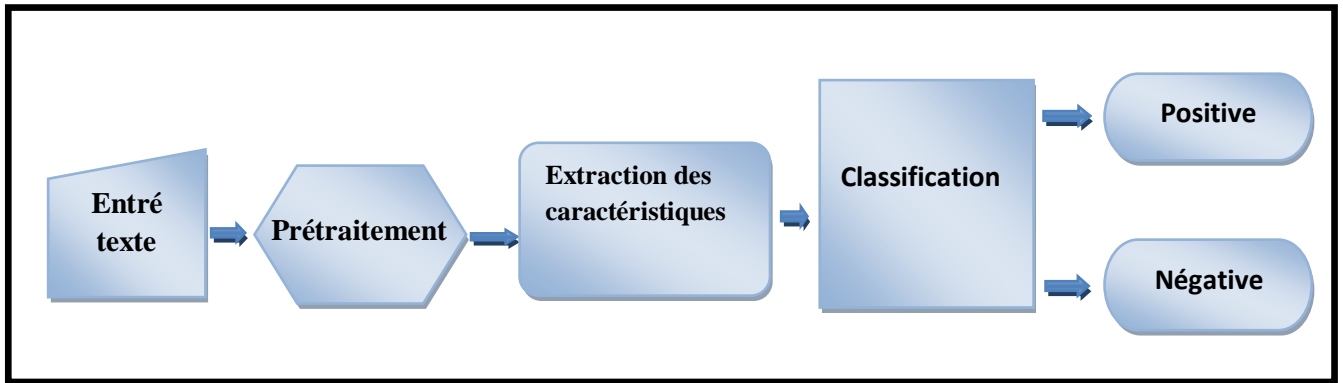


Figure 4 : Les étapes de l'approche basée sur l'apprentissage automatique

3.2 L'approche basée sur lexicque

La méthode du lexicque des sentiments repose sur un ensemble préétabli de termes de sentiments appelé "lexicque des sentiments". Cette approche peut être divisée en deux méthodes l'approche basée sur un dictionnaire et l'approche basée sur un corpus. Les deux méthodes utilisent des techniques statistiques ou sémantiques pour déterminer la polarité des sentiments.

Dans l'approche basée sur le dictionnaire, les termes d'opinion sont identifiés à partir d'une recherche de mots clés, puis le dictionnaire est consulté pour leurs synonymes et antonymes. Dans l'approche basée sur le corpus, une liste de départ de mots d'opinion est utilisée pour identifier d'autres termes d'opinion dans un grand corpus de textes afin de trouver des mots d'opinion avec des orientations spécifiques au contexte [6].

L'approche lexicque (ou l'approche à base de règles) définit un ensemble de règles dans un type de langage de programmation (script) qui identifie la subjectivité, la polarité ou le sujet d'une opinion. Cette approche peut utiliser diverses entrées, telles que.

- Techniques classiques de NLP, telles que la racinisation, tokenisation.
- Autre opération basée sur le lexicque, ils utilisent le dictionnaire des sentiments avec des mots d'opinion et les faire correspondre avec les données pour déterminer la polarité [7].

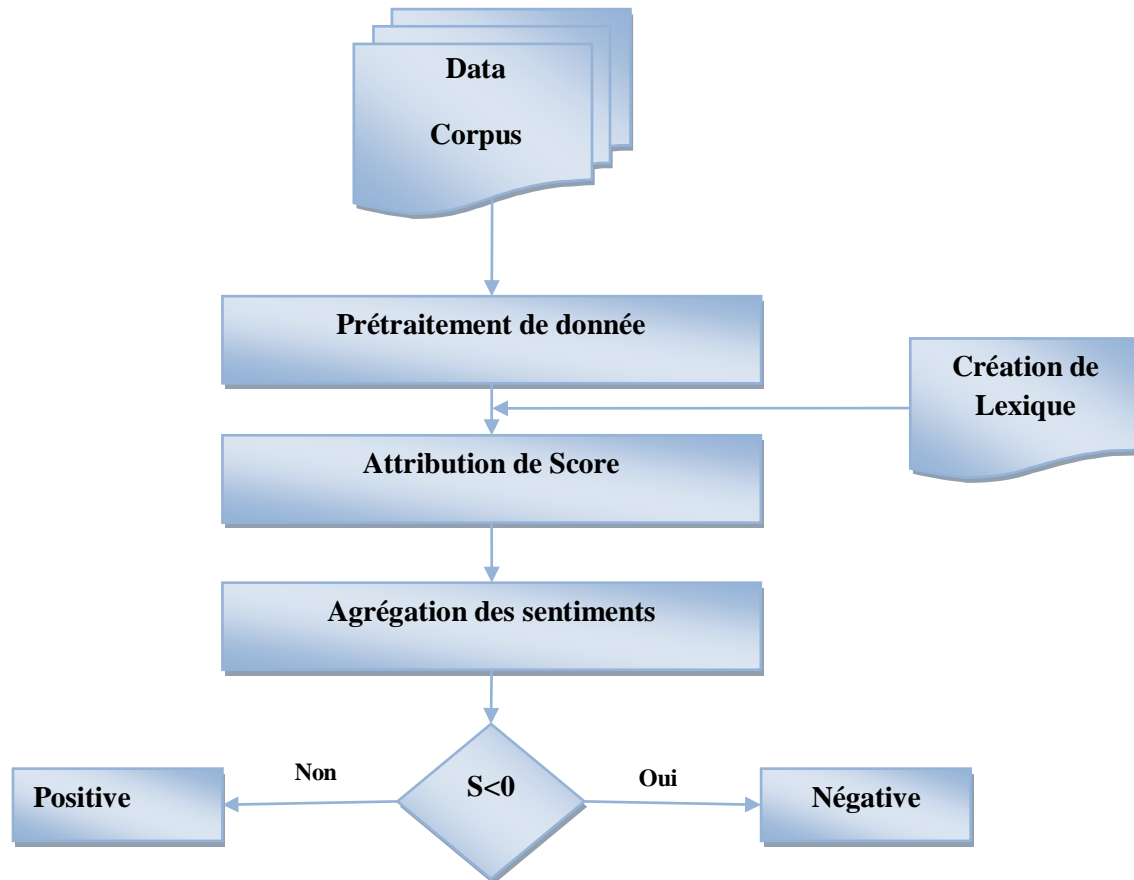


Figure 5 : L'organigramme de l'approche basée sur lexique

L'approche couramment utilisée dans l'analyse des sentiments est basée sur un lexique, comme indiqué dans la revue de littérature. Cette méthode utilise un dictionnaire préétabli de deux catégories de mots pondérés pour identifier la polarité d'un texte. Elle identifie les mots positifs et négatifs dans un texte en utilisant un petit ensemble de mots d'opinion divisé en deux catégories. Les mots dans la première catégorie sont associés à une terminologie plus positive, tandis que les mots de la seconde catégorie sont associés à un sentiment plus négatif.

Pour établir le dictionnaire, les chercheurs commencent généralement par les mots les plus évidents exprimant un sentiment positif, dont le nombre est augmenté par leurs synonymes pour constituer la catégorie positive. Une catégorie de mots négatifs peut être constituée automatiquement à partir des antonymes des mots de la catégorie positive, puis elle est amplifiée par d'autres mots jugés négatifs. Chaque mot est associé à une pondération positive ou négative. En sommant les mots, on obtient une évaluation globale du sentiment dans le texte [19].

3.3 Approche hybride

L'utilisation d'une approche hybride qui combine à la fois l'apprentissage automatique et l'approche basée sur le lexique est considérée comme supérieure à l'utilisation de chacune des approches de manière isolée. Des études ont montré que cette approche hybride permet d'obtenir de meilleurs résultats dans certaines tâches que l'utilisation de l'apprentissage automatique ou de l'approche basée sur le lexique seule [3].

4. Processus général de l'Analyse des Sentiments

Dans cette section nous illustrons l'architecture générale de l'Analyse des Sentiments présenté dans la figure au-dessous Figure 6[3].



Figure 6 : Processus général de l'analyse des sentiments

5. L'Analyse des Sentiments en langue Arabe

La recherche sur l'analyse des sentiments en langue arabe est encore relativement limitée par rapport à celle menée en Anglais. La langue anglaise domine le monde de la science, ce qui a

Chapitre I Fouille de sentiments médicaux

conduit à une plus grande quantité de recherches dans ce domaine. Cependant, récemment, certains chercheurs se sont intéressés à l'application de l'Analyse des Sentiments dans d'autres langues, y compris l'arabe. Il existe encore un écart significatif entre la recherche menée en arabe et en anglais, présenté dans la figure au-dessous Figure 7.

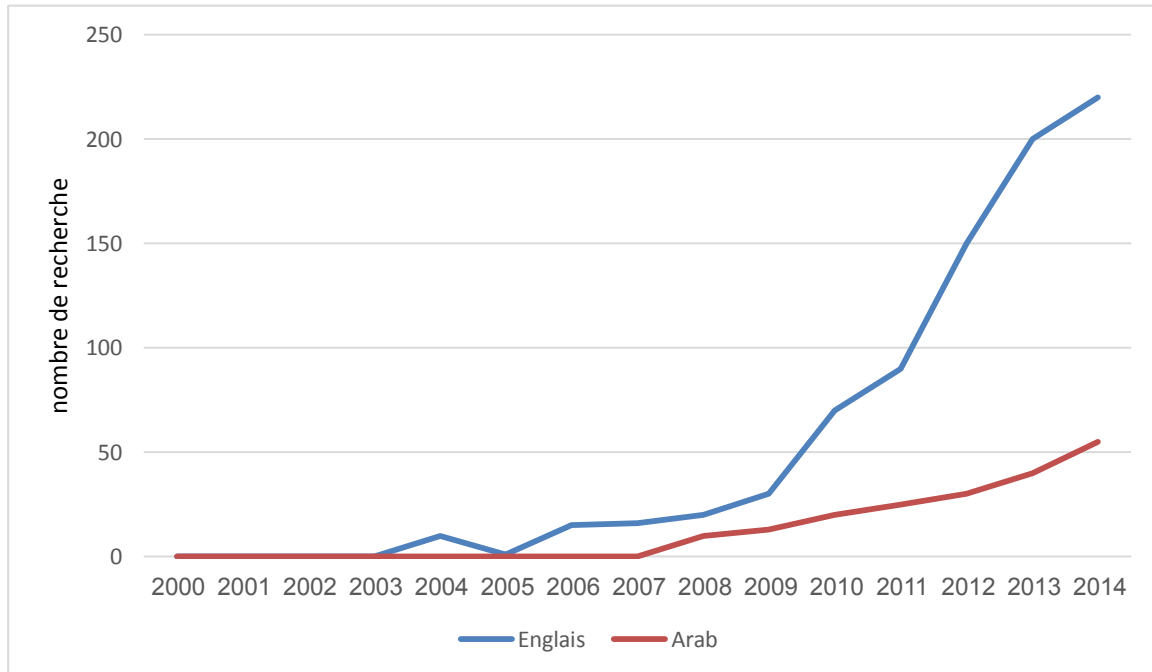


Figure 7 : La différence de la recherche entre l'arabe et l'anglais

Ces données collectées à l'aide de mots clés pertinents dans le champ d'analyse des Sentiments dans les deux langues. Il est clair qu'il existe un grand écart entre le travail réalisé en arabe et l'anglais, cela peut être dû aux limitations des outils ou des ressources du NLP de l'arabe. En outre, il peut révéler que l'arabe nécessite un traitement spécial en raison de sa complexité nature et structure [5].

5.1 La richesse de la langue arabe

La langue arabe est connue pour sa richesse et sa diversité. Elle est écrite au moyen de 28 lettres, représentées notamment par les symboles "أ ؤ ج د", qui constituent son système d'écriture. La langue arabe compte plus de 12 millions de mots, ce qui en fait la langue avec le plus grand nombre de mots. En comparaison, l'anglais en compte environ 600 000, le français 150 000 et le russe 130 000.

Cette abondance lexicale de la langue arabe est remarquable. Par exemple, il existe environ 80 termes différents pour désigner le miel, 200 pour le serpent, 500 pour le lion, et jusqu'à 1000 pour le chameau et l'épée. De plus, il y aurait jusqu'à 4400 termes pour décrire l'idée de malheur.

Chapitre I *Fouille de sentiments médicaux*

Les grammairiens arabes affirment que toutes les racines des mots arabes sont initialement des verbes. Bien que certains prétendent qu'il y ait plus de 12 millions de mots, en réalité, le nombre de racines de mots est estimé à environ 6000. Cette particularité témoigne de la complexité et de la richesse de la structure linguistique de l'arabe [5].

5.2 Complexité de la langue arabe

- **La signification des mots** : peut varier en fonction du contexte, ce qui signifie qu'un mot peut avoir plusieurs sens différents [8].
- **Dans différents contextes** : un mot peut appartenir à plus d'une catégorie lexicale, telle que nom, verbe, adjectif, etc., ce qui entraîne des variations dans sa catégorie grammaticale [8].
- **Synonymes** : Les langues possèdent de nombreux mots considérés comme des synonymes, c'est-à-dire des mots ayant des sens similaires. En utilisant des outils d'analyse morphologique sur un corpus de texte donné, les chercheurs peuvent déterminer les synonymes d'un mot, leur fréquence respective et identifier le synonyme le plus couramment utilisé [8].
- **La forme verbale selon son cas** : Certains mots arabes peuvent changer de forme en fonction de leur cas grammatical (nominatif, accusatif ou génitif) [8].
- **Tachkil** : présent également, contribue à modifier le sens du mot d'une phrase à l'autre. Veuillez traduire en arabe.

6. Conclusion

L'analyse des sentiments revêt une importance cruciale dans le domaine des soins de santé, permettant de mesurer la satisfaction des patients et d'améliorer la qualité des soins. Les acteurs du secteur de la santé peuvent exploiter cette analyse pour identifier les problèmes et les préoccupations des patients, et apporter les ajustements nécessaires afin d'améliorer leurs produits et services. Dans le prochain chapitre, nous explorerons l'utilisation des méthodes d'analyse des sentiments, notamment l'apprentissage automatique et les méthodes lexicales, dans le contexte des soins de santé. Nous examinerons les différents types d'approches et d'algorithmes associés, ainsi que leurs avantages et inconvénients.

Chapitre II

*Les approches d'analyse
des sentiments*

Chapitre II Les approches d'analyse des sentiments

1. Introduction

Ce chapitre explore les approches d'analyse des sentiments dans le domaine médical. Nous examinerons deux approches couramment utilisées : l'apprentissage automatique et l'approche lexicale. L'analyse des sentiments vise à évaluer les réactions émotionnelles des patients, ce qui est essentiel pour comprendre leur bien-être et adapter les soins médicaux en conséquence. L'apprentissage automatique permet aux systèmes informatiques d'apprendre à partir de données afin de détecter et d'évaluer les sentiments exprimés. D'autre part, l'approche lexicale utilise des lexiques spécifiques pour évaluer les sentiments associés aux textes médicaux. En explorant ces approches, nous découvrirons comment elles peuvent contribuer à une meilleure compréhension des émotions des patients et à une amélioration globale des soins de santé.

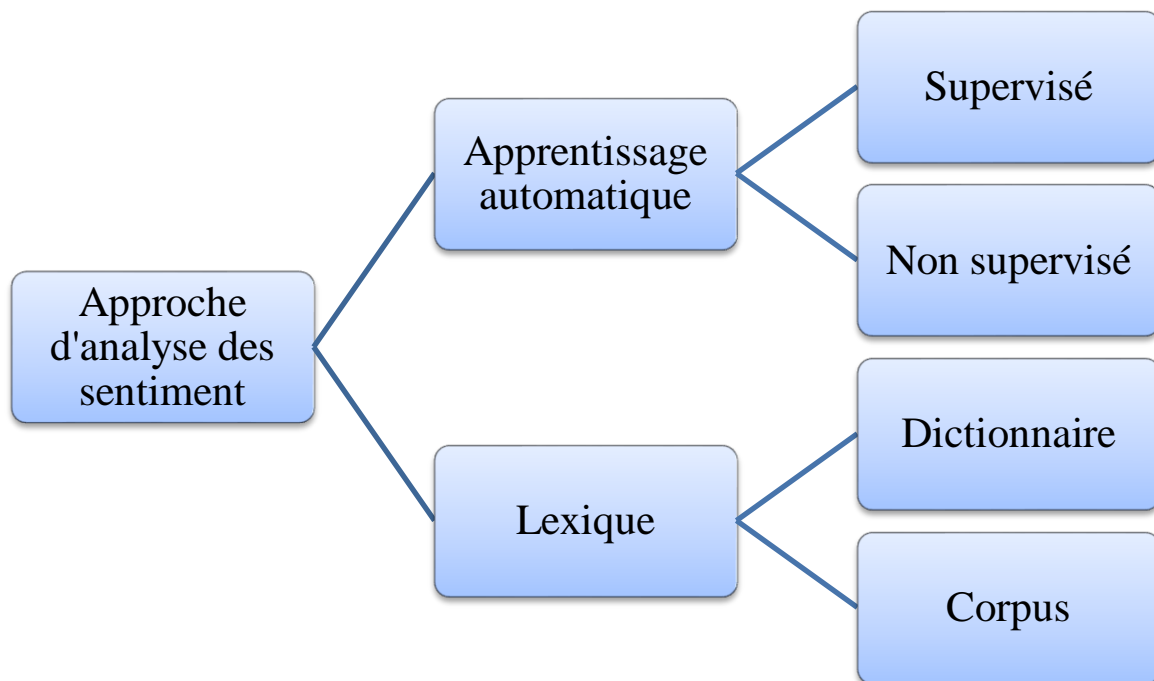


Figure 8 : Les approches d'analyse des sentiments

2. L'intelligence artificielle

L'intelligence artificielle (IA) est un terme largement utilisé ces dernières années (voir Figure 9). Elle se réfère à la capacité d'une machine à agir de manière autonome, en reproduisant des activités liées au comportement humain, sans être explicitement programmée. L'apprentissage automatique, une discipline de l'IA, vise à développer des programmes informatiques capables de s'améliorer automatiquement grâce à l'expérience acquise. En d'autres termes, l'IA permet

Chapitre II Les approches d'analyse des sentiments

aux machines d'apprendre par elles-mêmes et de s'adapter pour accomplir des tâches de manière plus efficace.



Figure 9 : Intelligence artificiel (IA)

L'intelligence artificielle (IA) est un domaine vaste où nous cherchons à reproduire le comportement humain afin de rendre les machines capables d'accomplir diverses tâches telles que la résolution de problèmes, la représentation des connaissances, la reconnaissance vocale, et bien d'autres encore. L'idée fondamentale est de transférer les connaissances dans les machines.

Grâce à ces avancées, de nouveaux systèmes sophistiqués a émergé, capables de modifier leur comportement en se basant uniquement sur leurs données d'entraînement, sans nécessiter de modifications de leur code. Ces avancées dans les techniques d'apprentissage machine ont ouvert de nouvelles perspectives pour l'IA [9].

3. Approche basée sur L'apprentissage automatique

L'apprentissage automatique est un domaine spécifique au sein de l'intelligence artificielle (IA) qui se focalise sur la conception de systèmes capables d'apprendre ou d'améliorer leurs performances en fonction des données qu'ils utilisent. L'intelligence artificielle, quant à elle, est un terme plus général qui englobe les systèmes ou les machines conçus pour imiter l'intelligence humaine. Bien que les termes "apprentissage automatique" et "intelligence artificielle" soient souvent utilisés de manière interchangeable, ils ne signifient pas exactement la même chose voir Figure 10. Il est important de noter que, bien que tout apprentissage automatique repose sur l'intelligence artificielle, cette dernière englobe un domaine beaucoup plus vaste que l'apprentissage automatique lui-même [11].

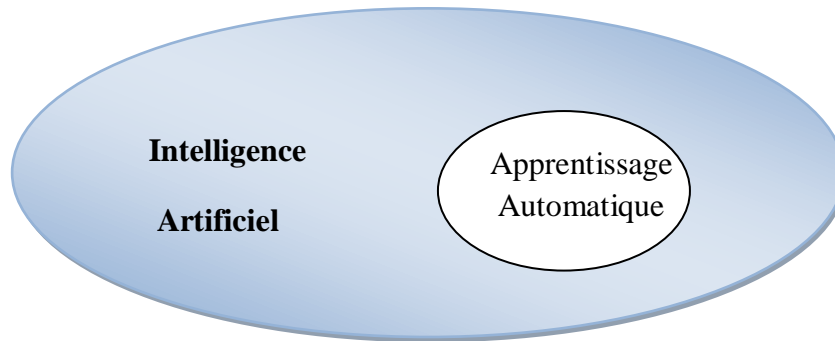


Figure 10 : La relation entre apprentissage automatique et l'intelligence artificielle

L'apprentissage automatique est une branche de l'intelligence artificielle qui confère à un système la capacité de comprendre grâce à divers algorithmes. Son principe repose sur l'idée d'enseigner aux algorithmes à partir de données, leur permettant ainsi de faire des prédictions et d'apprendre à résoudre des tâches spécifiques sans qu'il soit nécessaire de programmer explicitement la solution de la tâche [12].

3.1 Principe de fonctionnement de l'apprentissage automatique

Les algorithmes d'apprentissage automatique se réfèrent à l'apprentissage d'une fonction cible (f) qui associe de manière optimale les variables d'entrée (X) à une variable de sortie (Y) : $Y = f(X)$. Il s'agit d'une tâche d'apprentissage générale où notre objectif est de faire des prédictions futures (Y) à partir de nouveaux exemples de variables d'entrée (X). Nous n'avons pas de connaissance préalable sur la forme ou l'apparence de la fonction (f). Si nous la connaissions, nous pourrions l'utiliser directement sans avoir besoin de l'apprendre à partir de données en utilisant des algorithmes d'apprentissage automatique.

Le type d'apprentissage automatique le plus couramment utilisé consiste à apprendre la correspondance $Y = f(X)$ pour effectuer des prédictions de Y pour de nouvelles valeurs de X . Cela est connu sous le nom de modélisation prédictive ou d'analyse prédictive, et notre objectif est de réaliser les prédictions les plus précises possibles.

En principe, l'apprentissage automatique fonctionne de manière similaire à l'apprentissage humain. Par exemple, tout comme un enfant apprend à reconnaître certains objets sur des images, un ordinateur peut également "apprendre" à identifier des objets ou à distinguer des personnes. Pour ce faire, un logiciel d'apprentissage est d'abord alimenté en données et entraîné.

Chapitre II Les approches d'analyse des sentiments

Dans le domaine médical, par exemple, on peut utiliser l'apprentissage automatique pour aider à diagnostiquer des maladies. En alimentant le logiciel avec des données médicales, telles que des images de scanner ou des résultats de tests, et en les associant aux diagnostics correspondants, l'algorithme peut apprendre à reconnaître les caractéristiques des différentes maladies.

Le logiciel d'apprentissage reçoit constamment des retours d'information de la part des experts médicaux, qui sont utilisés par l'algorithme pour ajuster et optimiser le modèle. À chaque nouvel ensemble de données, le modèle s'améliore et devient capable de distinguer plus précisément les différentes conditions médicales, ce qui peut contribuer à des diagnostics plus précis et à de meilleurs résultats pour les patients [12].

3.2 Les types d'apprentissage automatique

On distingue différents types d'algorithmes Machine Learning. Généralement, ils peuvent être répartis en deux catégories : supervisés et non supervisés.

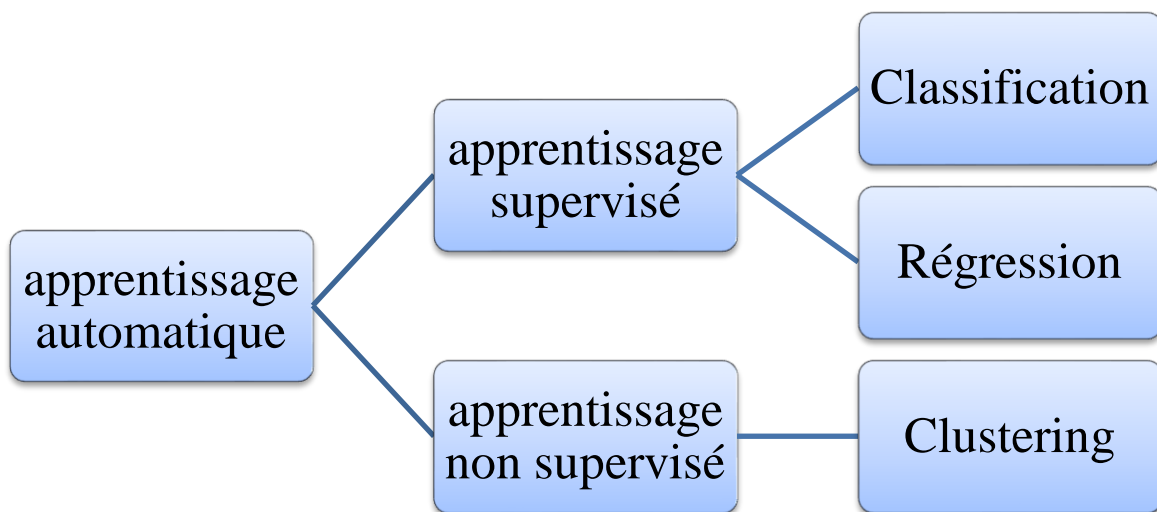


Figure 11 : Les branches de l'apprentissage automatique

3.2.1 L'apprentissage supervisé

L'apprentissage supervisé est une méthode d'apprentissage automatique qui génère automatiquement des règles à partir de données d'apprentissage déjà validées. Son objectif est d'inférer un modèle de prédiction à partir d'un ensemble d'apprentissage composé de paires {observation, étiquette}, où chaque étiquette dépend de l'observation correspondante. Dans ce

Chapitre II Les approches d'analyse des sentiments

processus, un oracle fournit les étiquettes pour chaque donnée d'apprentissage. L'apprentissage supervisé est couramment utilisé pour résoudre des problèmes de classification, de régression. Cette approche permet de prédire des événements futurs avec une probabilité statistique en se basant sur des modèles de prédiction créés à partir d'événements déjà observés [13] [14].

3.2.1.1 Les algorithmes de classification

Lorsque l'ensemble des valeurs de sortie est discret, les méthodes de classification sont utilisées pour attribuer une classe (étiquette ou label) à chaque valeur d'entrée. Ces techniques de classification peuvent être basées sur des hypothèses probabilistes (comme le naïf bayésien), des concepts de proximité (comme les k plus proches voisins) ou des explorations dans des espaces d'hypothèses (comme les arbres de décision). Le choix de la technique appropriée est crucial, afin de sélectionner la méthode la mieux adaptée capable de séparer de manière optimale les données d'apprentissage [13].

Dans notre mémoire, nous nous sommes basés sur trois algorithmes de classification : Naïve Bayes, arbre de décision et SVM (Support Vector Machine).

- **SVM** : La machine à vecteurs de support (SVM) est un algorithme de classification qui se base sur un hyperplan séparateur pour discriminer les données. En utilisant un ensemble de données d'entraînement étiquetées (apprentissage supervisé), l'algorithme génère un hyperplan optimal qui permet de catégoriser de nouvelles données [14].

Imaginons qu'il y ait deux classes de données distinctes. Dans le cadre de l'algorithme des machines à vecteurs de support (SVM), un hyperplan ou une frontière est recherché pour séparer au mieux ces deux classes de données. L'objectif est de trouver l'hyperplan qui maximise la distance entre les deux classes, ce que l'on appelle la marge. Bien qu'il puisse exister plusieurs façons de séparer les deux classes, l'algorithme SVM cherche spécifiquement l'hyperplan qui maximise cette marge ou espace entre les classes, créant ainsi une frontière claire et distincte. La Figure 12 présente un exemple classification avec SVM.

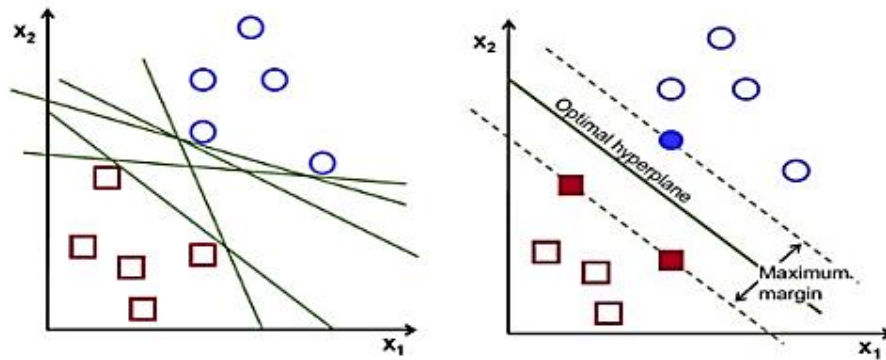


Figure 12 : Exemple d'un SVM

- **Arbre de décision** : est un algorithme qui représente visuellement les données sous forme de branches pour démontrer les résultats potentiels de différentes actions. Il classe et prédit les variables de réponse en se basant sur les décisions prises dans le passé. Cette approche graphique s'est avérée efficace, car les résultats des arbres de décision sont facilement explicables. Les décisions prises et leurs effets probables sur un résultat final sont clairement visibles, même lorsque les ensembles de données d'entrée sont incomplets.

Cependant, les arbres de décision peuvent devenir difficiles à interpréter lorsqu'ils sont utilisés avec de grandes quantités de données et des variables complexes. Par conséquent, ils sont souvent utilisés pour prendre des décisions à faible enjeu, telles que l'anticipation des fluctuations des taux d'intérêt ou les réactions du marché suite à une modification importante d'un produit d'une entreprise [5].

Les arbres de décision sont des méthodes largement utilisées en raison de leur efficacité et de leur simplicité dans le domaine de la classification supervisée. Ils offrent une représentation graphique du modèle qui est facilement interprétable.

Le modèle d'arbre de décision se compose d'un nœud racine, de nœuds intermédiaires, de branches et de feuilles. Le nœud racine est le point d'entrée de l'arbre, tandis que les feuilles représentent les valeurs de classe à prédire. Les branches correspondent aux résultats des tests effectués à chaque nœud. Pour effectuer une classification, l'arbre est parcouru de la racine aux feuilles en suivant une série de tests à chaque niveau de l'arbre. La théorie de Shannon est utilisée pour partitionner de nombreux arbres de décision. La Figure 13 présente une petite classification illustrée par un arbre de décision.

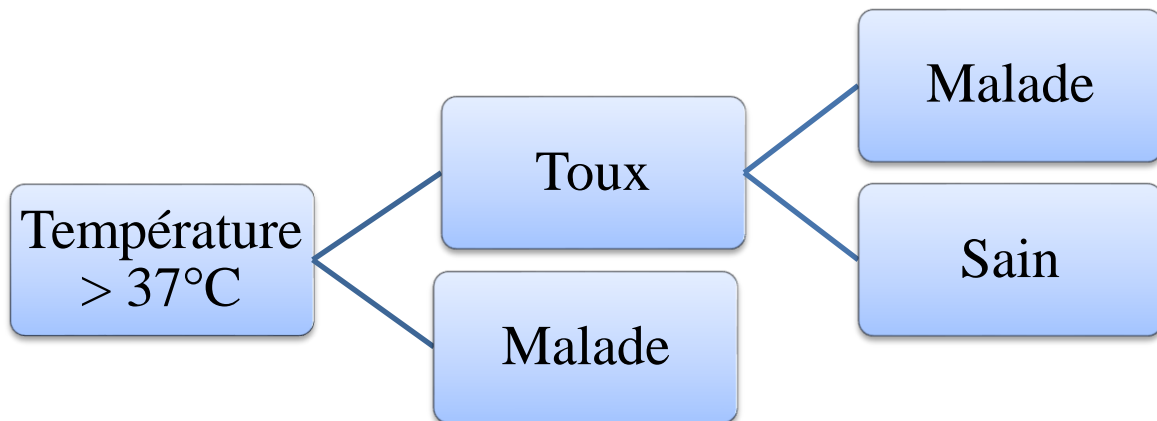


Figure 13 : schéma d'un arbre de décision

- **Naïve Bayes** : est une méthode d'apprentissage automatique qui se base sur le calcul de probabilités en utilisant le concept de l'approche bayésienne. En utilisant le théorème de Bayes, l'algorithme de naïve bayes combine la probabilité a priori et la probabilité conditionnelle dans une formule permettant de calculer la probabilité de chaque classification possible [16].

L'algorithme utilise le théorème de Bayes et suppose que tous les attributs sont indépendants les uns des autres, étant donné la valeur de la variable de classe. Bien que cette hypothèse d'indépendance conditionnelle soit rarement vérifiée dans les applications du monde réel, l'algorithme naïve bayes est réputé pour sa performance et sa capacité à apprendre rapidement dans différents problèmes de classification supervisée. Cette "naïveté" permet à l'algorithme de construire facilement des classifications à partir d'un ensemble de données sans avoir besoin de recourir à des schémas complexes d'estimation itérative des paramètres [17].

3.2.1.2 La régression

Les méthodes de régression sont largement utilisées en apprentissage automatique pour estimer des valeurs continues en sortie. Elles jouent un rôle essentiel dans la modélisation et l'analyse des données, et sont appliquées dans divers domaines, y compris le domaine médical. Par exemple, un modèle de régression pourrait être utilisé pour prédire le taux de croissance d'une tumeur en fonction de variables telles que la taille initiale de la tumeur, l'âge du patient et d'autres facteurs pertinents. Cela permettrait aux médecins de mieux comprendre l'évolution de la tumeur et de prendre des décisions éclairées concernant le traitement et le suivi du patient. En exploitant les données d'entrée, le modèle de régression peut fournir des estimations précises et des prévisions basées sur les caractéristiques observées dans les données médicales [18] [13].

Chapitre II Les approches d'analyse des sentiments

Les algorithmes.

- Régression linéaire.
- Régression logistique.
- Cart.
- K-Nearest Neighbors (KNN).

3.2.1.3 Comparaison des algorithmes d'apprentissage supervisé

Nous avons comparé les performances des algorithmes d'apprentissage supervisé SVM (Support Vector Machines), DT (Decision Trees) et NB (Naive Bayes) pour l'analyse des sentiments médicaux. Chaque algorithme présente ses avantages et ses inconvénients spécifiques qui sont présentés dans le Tableau 1[5].

Algorithmes	Avantages	Inconvénients
SVM	<ul style="list-style-type: none">-Sa grande précision de prédiction-Ils peuvent être plus efficaces car ils utilisent un sous-ensemble de points d'entraînements.	<ul style="list-style-type: none">-Moins efficace sur les jeux de données contenant des valeurs aberrantes et du bruit.
DT	<ul style="list-style-type: none">-Il est simple à comprendre car il suit le même processus que celui que suit l'homme lorsqu'il prend une décision dans la vie réelle.-Il peut être très utile pour résoudre les problèmes liés à la prise de décision.	<ul style="list-style-type: none">-L'arbre de décision contient beaucoup de couches, ce qui le rend complexe.-Il peut présenter un problème de surajustement, qui peut être résolu à l'aide de l'algorithme forêt d'arbres décisionnels.
NB	<ul style="list-style-type: none">- est l'un des algorithmes de ML les plus rapides et les plus simples pour prédire une classe d'ensembles de données.-Il peut être utilisé pour les classifications binaires et multi-classes.-Il donne de bons résultats dans les prédictions multi-classes par rapport aux autres algorithmes.	<ul style="list-style-type: none">-Naïve Bayes suppose que toutes les caractéristiques sont indépendantes ou non liées, il ne peut donc pas apprendre la relation entre les caractéristiques.

Table 1 : Comparaison des algorithmes d'apprentissage supervisé

Chapitre II Les approches d'analyse des sentiments

3.2.2 L'apprentissage non supervisé

L'apprentissage non supervisé, également appelé clustering, consiste à regrouper des données non étiquetées en identifiant des structures et des relations intrinsèques. Il vise à diviser les données en groupes homogènes où les exemples similaires sont regroupés ensemble. Contrairement à l'apprentissage supervisé, il n'y a pas de valeurs de sortie prédéfinies. L'objectif est de découvrir des patterns et des groupes sans se baser sur des étiquettes préexistantes. Cela permet d'explorer et de révéler des informations cachées dans les données [19] [13]. La Figure 14 présente un exemple sur classification non super visé.

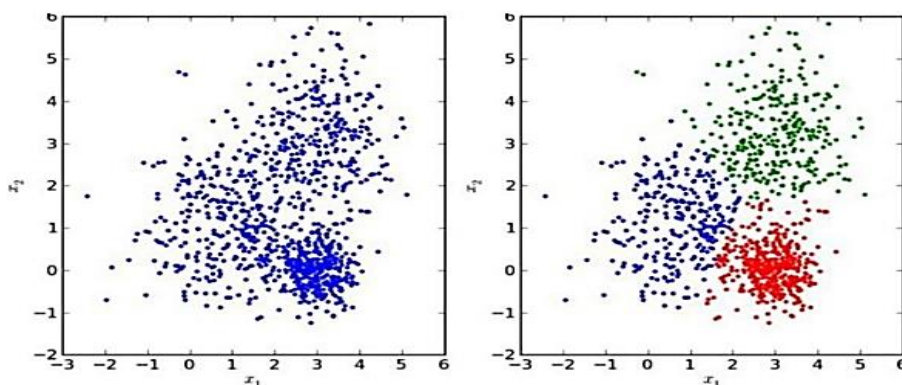


Figure 14 : Exemple de prédiction 3 clusters

4. Approche lexicale

L'utilisation d'un lexique est l'une des principales approches de l'analyse des sentiments. Elle consiste à évaluer le sentiment en se basant sur l'orientation sémantique des mots ou des phrases présents dans un texte. Cette méthode nécessite un dictionnaire de mots positifs et négatifs, avec une valeur de sentiment attribuée à chaque mot. Différentes techniques, manuelles ou automatiques, ont été développées pour créer ces dictionnaires.

Dans les approches basées sur le lexique, un morceau de texte est représenté comme un ensemble de mots. Ensuite, les valeurs de sentiment du dictionnaire sont assignées à chaque mot ou expression positive ou négative présente dans le texte. Une fonction de combinaison, telle que la somme ou la moyenne, est utilisée pour prédire le sentiment global de la phrase.

En plus de la valeur sentimentale, le contexte local d'un mot est souvent pris en compte, notamment la présence de négation ou d'intensification [6].

L'approche basée sur le lexique repose sur l'utilisation d'un lexique contenant un ensemble de mots associés à des valeurs de polarité. Les mots positifs ont des valeurs de polarité supérieures

Chapitre II Les approches d'analyse des sentiments

à zéro, tandis que les mots négatifs ont des valeurs de polarité inférieures à zéro. Les mots qui ne sont pas présents dans le lexique sont considérés comme neutres [3].

La méthode basée sur le lexique utilise un lexique qui contient des termes accompagnés de scores de sentiment correspondants. Chaque terme peut être associé à un mot unique ou à une phrase. Le sentiment d'un texte est déterminé en fonction de la présence ou de l'absence de termes du lexique dans celui-ci. Cette approche englobe deux sous-approches : l'approche basée sur le corpus et l'approche basée sur le dictionnaire [9].

4.1 Les types des approches basé sur lexique

Les approches basées sur le lexique sont caractérisées par deux types différents, sont :

4.1.1 Approche basée sur le dictionnaire

À partir d'un ensemble initial de mots de sentiment connus pour leur orientation positive ou négative, on exploite les ressources telles que les thésaurus et les corpus, comme **Word Net**, pour trouver des synonymes et des antonymes pour chaque mot de la liste. Les mots nouvellement découverts sont ajoutés à la liste initiale, et le processus est répété jusqu'à ce qu'aucun nouveau mot ne puisse être trouvé.

Cependant, cette approche présente un inconvénient majeur elle ne parvient pas à trouver des mots d'opinion qui sont spécifiques à un domaine particulier. Par exemple, si l'on dit "le haut-parleur du téléphone est silencieux", cela exprime une opinion négative, tandis que si l'on dit "la voiture est silencieuse", cela indique une opinion positive [3].

Ressources des dictionnaires open source.

- Word Net.
- Senti Word Net.
- General Inquirer Dataset (General Inquirer Categories).
- MPQA opinion Corpus (MPQA Ressources) [7].

4.1.2 Approche basée sur le corpus

Selon Bing Liu, l'approche basée sur le corpus peut être appliquée dans deux situations. Le premier cas consiste à identifier les mots d'opinion et leurs polarités dans un corpus de domaine en utilisant un ensemble préexistant de mots d'opinion. Le deuxième cas concerne la création

Chapitre II Les approches d'analyse des sentiments

d'un nouveau lexique spécifique à un domaine en se basant sur un autre lexique existant, en utilisant un corpus de domaine [5].

Les résultats suggèrent que, même si les mots d'opinion sont dépendants du domaine, il peut arriver qu'un même mot ait une orientation opposée en fonction du contexte [9].

4.2 Algorithme de l'approche lexicale

1. Initialiser la note totale du sentiment $S=0$.
2. Pour chaque mot de texte, vérifier la présence dans le lexique
 - a. Si le mot est positif, alors $S=S+W$.
 - b. Si le mot est négatif, alors $S=S-W$.
3. Regarder le sentiment total du texte
 - a. Si $S >$ au seuil, alors le sentiment du texte est positif.
 - b. $S <$ au seuil, alors le sentiment du texte est négatif.

4.3 Comparaison entre les deux approches

Approches	Avantages	Inconvénients
Approche basée sur lexicale	-Il ne demande aucune donnée d'entraînement ou des données étiquetées et ceci permet d'introduire moins d'opérations de calcul.	-Moins de capacité de classification en fonction du contexte ou du domaine. -Exige l'existence de ressources linguistiques puissantes qui ne sont pas toujours disponibles
Approche basée sur l'apprentissage automatique	-Il peut être transformé en ce que le domaine détonant pour mieux travailler. -Un dictionnaire n'est pas nécessaire. -Donne de meilleurs résultats en termes de haute précision de classification.	- Peut être affecté par les variations de classes et aussi par l'effet des changements linguistiques. -Les Méthodes qui se sont entraînées sur un domaine spécifique, dans la plupart des cas ne fonctionnent pas avec un autre.

Tableau 1 : Les avantages et les inconvénients d'approches (Lexicale et apprentissage automatique)[5]

Chapitre II Les approches d'analyse des sentiments

5. Conclusion

Les approches d'analyse des sentiments, qu'elles soient basées sur l'apprentissage automatique ou sur des lexiques, sont des méthodes clés pour extraire et comprendre les sentiments exprimés dans les textes. L'apprentissage automatique permet d'exploiter les modèles et les données pour prédire le sentiment, tandis que l'approche basée sur le lexique se concentre sur l'utilisation de dictionnaires de mots pour évaluer le sentiment. Ces approches complémentaires offrent des perspectives différentes et peuvent être utilisées de manière combinée pour améliorer la précision de l'analyse des sentiments et obtenir des informations plus riches sur les opinions des utilisateurs.

Chapitre III

*Implémentation et
expérimentation*

Chapitre III Implémentation et Expérimentation

1. Introduction

Dans ce dernier chapitre de notre mémoire, nous mettons en avant l'importance de l'analyse des sentiments dans différents domaines, en particulier dans le domaine médical. Nous nous concentrons sur l'analyse des sentiments des patients concernant la qualité des soins. Pour ce faire, nous utilisons des algorithmes d'apprentissage automatique et lexical. Nous présentons notre contribution en fournissant les outils, les ensembles de données et les étapes de préparation nécessaires pour appliquer ces algorithmes. Nous comparons également les résultats obtenus et présentons une application pratique pour tester nos modèles sur des données réelles. Ce chapitre offre une vision complète de notre travail et met en évidence l'importance de l'analyse des sentiments médicaux pour améliorer la satisfaction des patients.

2. L'implémentation

L'implémentation est une étape cruciale de notre travail de recherche. Nous décrivons en détail le processus d'implémentation de notre système d'analyse des sentiments médicaux. Nous commençons par présenter les ressources utilisées, notamment l'ensemble de données que nous avons recueilli et annoté, ainsi que le lexique spécifique au domaine médical que nous avons utilisé. Ensuite, nous détaillons l'architecture de notre application.

2.1 Les ressources

Les ressources jouent un rôle crucial dans notre étude d'analyse des sentiments médicaux. Dans cette section dédiée aux ressources, nous décrivons en détail les différentes sources et outils que nous avons utilisés pour collecter, prétraiter et analyser le corpus de données.

2.1.1 L'ensemble de données

La quantité et la qualité des données sont des facteurs cruciaux qui impactent directement la précision et l'efficacité du modèle d'analyse des sentiments. En effet, plus le nombre de données est élevé et plus elles sont fiables, plus le résultat obtenu sera précis et pertinent. Dans cette section consacrée aux ressources, nous mettons en évidence les éléments clés utilisés dans notre étude d'analyse des sentiments médicaux.

2.1.2 Collecte de données

La collecte de données consiste sur des étapes essentielles pour recueillir des corpus précieux permettant de mettre en avant des études éclairées. Dans le cadre de notre mémoire, Nous avons collecté les données manuellement, nous avons mis en place une méthodologie de collecte de données en utilisant les questionnaires. Nous avons eu l'opportunité de faire un stage pratique de 40 jours divisés en deux périodes à l'Etablissement Public de Santé de Proximité (EPSP) de KHENCHELA. Dans ces deux reprises nous sommes arrivés à distribuer les questionnaires aux patients, leur offrant ainsi la possibilité de partager leur opinion sur les services et la qualité des soins qu'ils ont reçus. Cette mise en œuvre concrète nous a

Chapitre III Implémentation et Expérimentation

permettais d'obtenir des données pertinentes, indispensables pour évaluer l'efficacité des soins. En analysant ces données, nous avons pu identifier les domaines spécifiques qui nécessitent des améliorations et proposer des recommandations pour améliorer les services de santé, en mettant l'accent sur les besoins des patients et en garantissant une meilleure qualité de soins pour tous. Des échantillons sur les réponses des patients sur les questionnaires sont présenté dans les figures 15 ,16 et 17.

الرجاء الإجابة على هذه الأسئلة بكل شفافية لتحسين جودة الرعاية

1- كيف وجدت جودة العناية الطبية الذي تلقيته؟
جودة العناية الطبية التي تلقيتها كانت في مستوى
العاملين. فجا من طرف أصحاب الممار البيخاء

2- ما هي الأمور التي أعجبتك في الخدمات الصحية؟
الأجور التي أجبتي في الخدمات الصحية هي المراكز
والخدمات التي أتت بها المستشفى. المسؤوهم

3- ما هي الأمور التي لم تعجبك في الخدمات الصحية؟
الأمور التي لم تعجبني في الخدمات الصحية هي نقص في عملية
استقبال المرضى والمعاملة الجيدة موجودة. الطواق الخاصة
من قبل بعض الناس متوقف. وليس معاملة الجدية
صدق والمعاملة الجيدة هي أهم شيء مفقود في الخدمة.

Figure 15 : Réponse de patient sur questionnaire (Exemple 1)

- الرجاء الإجابة على هذه الأسئلة بكل شفافية لتحسين جودة الرعاية

1- كيف وجدت جودة العناية الطبية الذي تلقينته؟
عناية جيدة (+) متميزة

2- ما هي الأمور التي أعجبتك في الخدمات الصحية؟
كل الأمور على ما يرام على جميع الجند ما شاء

3- ما هي الأمور التي لم تعجبك في الخدمات الصحية؟
/

Figure 16 : Réponse de patient sur questionnaire (Exemple 2)

الرجاء الإجابة على هذه الأسئلة بكل شفافية لتحسين جودة الرعاية

1- كيف وجدت جودة العناية الطبية الذي تلقينته؟
كانت أفضل من جميع المراكز الطبية لعدم الذهاب والنوم
للطبيب الجراح في المستشفى والاطباء

2- ما هي الأمور التي أعجبتك في الخدمات الصحية؟
لا يوجد ما أعجب أم ليس الشاغل

3- ما هي الأمور التي لم تعجبك في الخدمات الصحية؟
لجاء من المستشفى على أسوأ
المعاملة

Figure 17 : Réponse de patient sur questionnaire (Exemple 3)

Chapitre III Implémentation et Expérimentation

2.1.3 Création de l'ensemble de données Data_Quality_Cares

Nous avons suivi plusieurs étapes pour recueillir des informations précieuses. Les principales étapes de création de Data_Quality_Cares sont :

- D'abord, nous avons distribué des questionnaires aux patients et les encourager à partager leurs opinions sur les services et la qualité des soins reçus.
- Ensuite, nous avons soigneusement examiné les réponses des patients, éliminant les commentaires répétitifs et les questionnaires vides (avec des réponses par "/").
- Après, avoir procédé à une sélection attentive, nous avons choisi les commentaires les plus appropriés (élimination des commentaires neutres, néants, etc.).
- Enfin, nous avons créé un ensemble de données, nommé Data_Quality_Cares, dans Excel, où nous avons stocké les informations que nous avons précédemment collectées.

2.1.4 Annotation de données

Dans l'annotation des données, Nous avons adopté une approche impliquant deux annotateurs indépendants pour évaluer et classifier les commentaires recueillis. Chacun des annotateurs a examiné attentivement les commentaires et a exprimé son évaluation de l'opinion, en les catégorisant comme étant positifs ou négatifs en fonction de sa propre interprétation. Cette méthode d'annotation a été utilisée afin de garantir une évaluation plus objective, en minimisant les éventuels biais individuels. Grâce à cette approche d'annotation des données, nous avons pu obtenir une annotation solide des commentaires, fournissant ainsi une base robuste pour l'expérimentation ultérieure et les conclusions de notre étude.

Dans le cas où les deux annotateurs ont des évaluations différentes sur le même commentaire, nous faisons recours à un troisième annotateur pour jouer le rôle d'arbitre. Le rôle de l'arbitre est de résoudre les écarts entre les annotations précédentes et de parvenir à un consensus. Ce processus d'arbitrage est crucial pour assurer la cohérence et la fiabilité des annotations, en particulier lorsque des opinions contradictoires se manifestent, comme illustré dans la troisième ligne du Tableau 2.

Chapitre III Implémentation et Expérimentation

Commentaires	Annotateur(A)	Annotateur(B)	Annotations
نقص كبير في خدمات الاستقبال والنظافة وكذلك عدم توفر مقاعد الانتظار	Nég	Nég	Nég
المتابعة الجيدة للمريض	Pos	Pos	Pos
هناك البعض من فئة الطاقم الطبي تحس نفسك أنك مرتاح اما البعض الآخر فلا	Pos	Nég	Pos

Tableau 2 : Échantillons sur le vote entre les annotateurs

2.1.5 L'agrément entre annotateurs Inter Annotators Agreement (IAA)

Dans le domaine des linguistiques informatiques (computationnel linguistiques), l'accord inter-annotateur, IAA, est une mesure utilisée pour évaluer la concordance entre les annotations effectuées par différents annotateurs sur un même ensemble de données. Il permet de quantifier le degré de concordance entre les annotateurs et de déterminer la fiabilité des annotations. Cette mesure revêt une grande importance en linguistique de corpus, où il est essentiel d'estimer la confiance que l'on peut accorder aux annotations fournies [28].

Dans notre étude, l'annotation des différents annotateurs est présentée dans le tableau au-dessous (Tableau 4).

	Annotateur A			Total
	Positif	Négatif	Total	
Annotateur B	Positif	131	2	133
	Négatif	0	131	131
	Total	131	133	264

Tableau 3 : L'agrément entre annotateurs IAA

Calcul de l'agrément entre annotateurs IAA ;

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \dots\dots\dots (1)$$

Pr(a) : l'accord observé relatif entre les annotations.

Pr(e) : l'accord attendu entre les annotations

Chapitre III Implémentation et Expérimentation

- Nous calculons $Pr(a)$, Sur les 264 documents. A et B ont tous les deux noté "positif" 131 commentaires, et tous deux indiqués "négatif" 131 fois. Par conséquent, le pourcentage d'accord observé est :

$$Pr(a) = \frac{131+131}{264} = 0,99 \text{ (99 \%)} \dots\dots\dots (2)$$

Ensuite, nous calculons $Pr(e)$. Pour ce faire, nous déterminons le pourcentage de fois que chaque annotateur utilise chaque étiquette et multiplions ces pourcentages pour déterminer à quelle fréquence les deux annotateurs utiliseraient la même étiquette sur le même document au même moment. Ensuite, nous ajoutons ces fréquences pour obtenir $Pr(e)$.

$$\text{Positif}(B) = \frac{131+2}{264} = 0,503 \dots\dots\dots (3)$$

$$\text{Positif}(A) = \frac{131+0}{264} = 0,496 \dots\dots\dots (4)$$

$$\text{Positif}(A*B) = (0,503*0,496) = 0,249 \dots\dots\dots (5)$$

$$\text{Négatif}(B) = \frac{131+0}{264} = 0,496 \dots\dots\dots (6)$$

$$\text{Négatif}(A) = \frac{2+131}{264} = 0,503 \dots\dots\dots (7)$$

$$\text{Négatif}(A*B) = (0,503*0,496) = 0,249 \dots\dots\dots (8)$$

$$Pr(e) = 0,249+0,249 = 0,498 \dots\dots\dots (9)$$

Le coefficient de Cohen (κ : Kappa) mesure l'accord entre deux annotateurs, en tenant compte de la possibilité d'un accord par hasard. L'équation est la suivante.

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} = \frac{0,99 - 0,498}{1 - 0,498} = \frac{0,492}{0,502} = 0,98 \dots\dots\dots (10)$$

Coefficient Kappa 0,98 indique un accord presque parfait l'accord au-delà de ce qui est étendu par hasard est étendu [21].

Après l'annotation des données Data_Quality_Cares, nous avons obtenu des statistiques détaillées, qui sont présentées dans le Tableau 5.

Chapitre III Implémentation et Expérimentation

Sentiment	Nombre de commentaires	
Négatif	131	49.62%
Positif	133	50.38%
Total	264	100%

Tableau 4 : Statistiques de Data_Quality_Cares

2.1.6 Description de l'ensemble des données

Notre ensemble de données est composé de deux champs. Le premier champ consiste sur le texte du commentaire, ce texte est extrait des questionnaires que nous avons distribués et collectés auprès des patients. Quant au deuxième champ, il est lié à la classe du commentaire, les classes sont affectées dans l'étape d'annotation. Le Tableau 6 décrit les variables contenues dans l'ensemble de données utilisé et la Figure 18 contient Présentation de Data_Quality_Cares.

Variables	Description	Types
Texte	Ce sont les sentiments et les opinions des patients sur la qualité de soins et service.	Object
Sentiment	Evaluation de l'opinion (Pos, Nég).	Object

Tableau 5 : Description de l'ensemble des données (Data_Quality_Cares)

	Text	Sentiment
0	عناية طبية جيد	Pos
1	...عدم الاستخدام المناسب لمعدات الوقاية شخصية وعد	Neg
2	عناية طبية رائعة	Pos
3	من الجيد ان الخدمة مجانية	Pos
4	...نقص كبير في خدمات الاستقبال والنظافة وكذلك عدم	Neg
...
259	تم تشخيصي بشكل خاطئ في المركز الطبي	Neg
260	...الأطباء لم يقوموا بالتحقق بشكل جيد ولم يهتموا ب	Neg
261	أنا عاضب ومحبط	Neg
262	...النظافة غير مرضية والتعقيم غير جيد. لم يكن هنا	Neg
263	أصبح بالاعتماد على هذا المستشفى	Pos

264 rows × 2 columns

Figure 18 : Présentation de Data_Quality_Cares

2.2 Le lexique utilisé

Le lexique peut représenter les mots d'une langue spécifique, et peut également englober tous les termes utilisés dans les divers domaines de l'expérience humaine, exprimés également dans une langue donnée [27]. Dans notre travail, le lexique utilisé consiste en deux ensembles de mots ; un ensemble des mots de polarité positive et un ensemble de mots de polarité négative.

2.2.1 Obtention du lexique

Le lexique utilisé est téléchargé du site web [Kaggle](https://www.kaggle.com)¹ et les statistiques de ce lexique sont présentées dans le tableau au-dessous (Tableau 6).

¹www.kaggle.com

Chapitre III Implémentation et Expérimentation

Sentiment	Nombre de terme	
Négatif	1652	59.13%
Positif	1142	40.87%
Total	2794	100%

Tableau 6 : Statistiques de lexique télécharger

2.2.2 Prétraitement et enrichissement de lexique

Nous avons remarqué que le lexique téléchargé est d'ordre général. Dans ce sens, les mots de sentiment ne sont pas réellement du domaine de notre étude ; savoir les sentiments médicaux. Pour ces raisons, nous avons procédé un prétraitement on a supprimé les mots qui sont pas important et on a fait aussi un enrichissement du lexique par des mots de polarité de domaine médicale. Les statistiques de lexique après l'enrichissement sont présentées dans le tableau au-dessous (Tableau 7) ; est le (Tableau 8) contient un exemple de contenu de ce lexique.

Sentiment	Nombre de terme	
Négatif	1227	54.53%
Positif	1023	45.47%
Total	2250	100%

Tableau 7 : Statistique du lexique

	Liste des mots négatifs	Liste des mots positifs
Exemple	اهمال مرهق مشوش متالم متوتر تعب شكاو خاطء خايب تاخير غاضب	سعيد متعاطف صبور مهتم متعا انصح يستمع منظم يهتم ستمع راض

Tableau 8 : Exemple des mots de lexique

Chapitre III Implémentation et Expérimentation

3. Architecture de notre application

L'architecture de notre application repose sur plusieurs composants essentiels pour l'analyse des sentiments médicaux. Tout d'abord, nous importons l'ensemble de données pertinentes dans notre système. Ensuite, nous effectuons un prétraitement des données, y compris une étape de prétraitement préliminaire, la tokenisation pour séparer les mots, et l'enracinement pour normaliser les termes. En ce qui concerne les approches d'analyse des sentiments, nous avons mis en place deux méthodes principales : l'approche basée sur l'apprentissage automatique et l'approche lexicale.

Pour l'approche basée sur l'apprentissage automatique, nous extrayons les caractéristiques des données et appliquons une méthode de classification pour prédire les sentiments. Pour l'approche lexicale, nous créons un dictionnaire de termes positifs et négatifs, puis classifions les commentaires en fonction de leur polarité. Enfin, nous évaluons les performances de nos modèles en utilisant des mesures d'évaluation appropriées. L'architecture de notre application est conçue de manière à permettre une analyse précise et efficace des sentiments médicaux, en fournissant des résultats fiables pour améliorer la prise de décision dans le domaine de la santé.

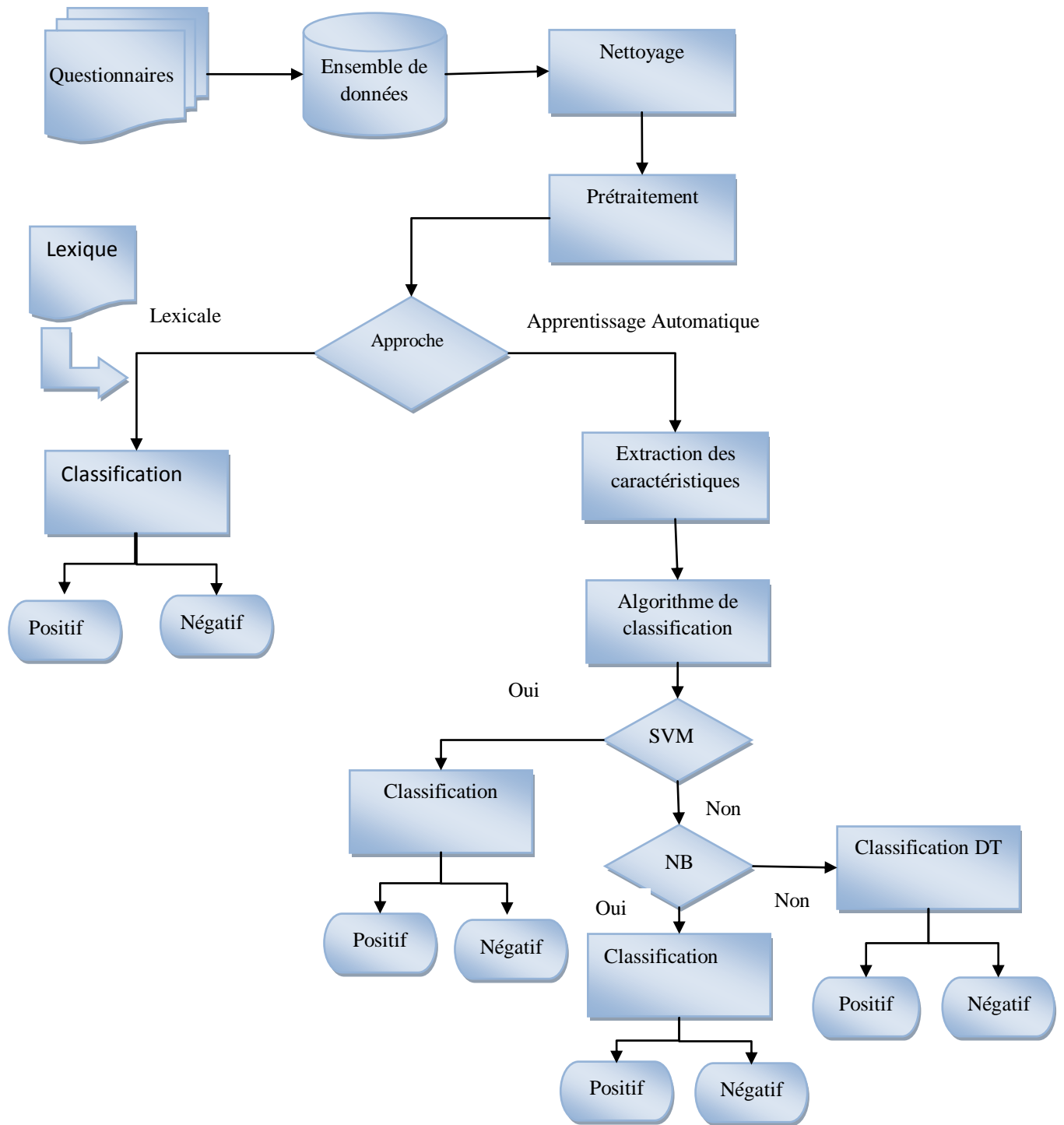


Figure 19 : Les principales étapes de l'analyse du sentiment

Chapitre III Implémentation et Expérimentation

3.1 Importation de l'ensemble de données

Dans cette étape, nous importons l'ensemble de données, également appelé dataset, dans notre système afin de le préparer pour l'analyse. Dans le cadre de ce travail de mémoire, nous avons créé notre propre ensemble de données spécifiquement pour cette tâche `Data_Quality_Cares`.

3.2 Prétraitement des données

Le prétraitement des données est une étape essentielle dans le domaine de l'analyse de données, y compris l'analyse des sentiments. Il s'agit d'un ensemble de techniques et de processus visant à préparer les données brutes pour une analyse ultérieure. L'objectif principal du prétraitement des données est d'améliorer la qualité des données, de réduire le bruit et d'obtenir des données cohérentes et prêtes à être utilisées dans les modèles et les algorithmes d'analyse.

3.2.1 Prétraitement préliminaire

Au cours de cette étape de prétraitement, nous avons utilisé la bibliothèque Python `"re"` (expressions régulières) pour supprimer les caractères spéciaux tels que les symboles, les chiffres et les signes de ponctuation du texte. L'objectif était de conserver uniquement les caractères alphabétiques dans les données pour garantir une cohérence dans l'ensemble du texte. Cette étape de prétraitement contribue à obtenir des données textuelles plus propres, cohérentes et adaptées à l'analyse des sentiments ou à d'autres tâches de traitement automatique du langage naturel.

3.2.2 Séparation de mots (Tokenisation)

La méthode de séparation des mots (tokenisation) est utilisée pour séparer votre texte en mots individuels. La tokenisation est le processus de décomposition d'un texte en unités appelées "tokens", qui correspondent généralement aux mots.

La méthode `word_tokenize ()` que nous avons utilisée est une fonction de la bibliothèque `NLTK` (Natural Language Toolkit) qui permet d'effectuer cette opération. L'objectif de la tokenisation est de segmenter le texte en mots afin de pouvoir les traiter individuellement par la suite. Cela permet de faciliter l'analyse du texte et d'appliquer différentes techniques de traitement du langage naturel, telles que l'analyse de sentiment, la classification de texte, etc.

La tokenisation peut être plus complexe pour certaines langues, comme l'arabe, où les mots peuvent être attachés les uns aux autres sans espaces [23]. La Figure 20 (capture sur la tokens) présente un exemple de séparation de mots d'un commentaire de notre corpus, `Data_Quality_Cares`.



Figure 20 : Exemple de séparation de mots d'un commentaire

3.2.3 Enracinement (Stemming)

La phase d'enracinement est une étape cruciale dans le domaine de l'analyse des sentiments. Elle vise à réduire la diversité du vocabulaire en remplaçant les mots similaires par leur racine (ou "stem"). Cette technique permet de diminuer la redondance et d'améliorer l'efficacité des résultats. Dans notre travail, nous avons utilisé l'algorithme Snow Ball pour effectuer l'enracinement.

L'algorithme Snow Ball permet l'enracinement largement utilisée et offert par la bibliothèque NLTK. Il prend en charge plusieurs langues, dont l'Arabe et plus de 15 autres langues.

En utilisant Snow Ball, nous avons pu réduire le nombre de formes différentes des mots dans notre corpus, ce qui facilite le traitement et l'analyse ultérieure des données textuelles [24]. La Figure 21(capture sur le stem) présente un exemple de racinisation de mots d'un commentaire de notre corpus, Data_Quality_Cares.

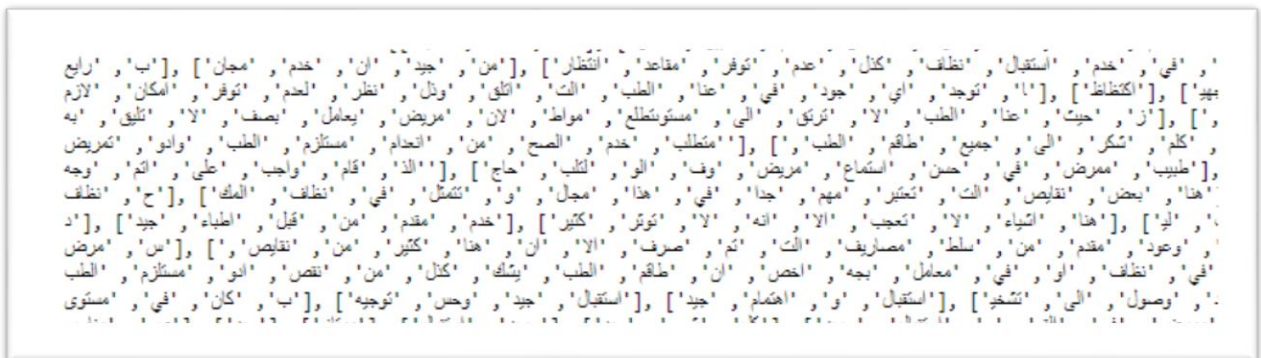


Figure 21 : Exemple de racinisation des mots de commentaires

Chapitre III Implémentation et Expérimentation

3.3.1 Approche basée sur l'apprentissage automatique

L'approche à base d'apprentissage automatique dans l'analyse des sentiments repose sur l'utilisation de divers algorithmes pour classer des données textuelles. Trois des algorithmes couramment utilisés sont appliqués dans cette tâche à savoir ; le Support Vector Machine (SVM), les Arbres de Décision (DT) et le Naïve Bayes (NB).

3.3.1.1 Extraction des caractéristiques

Le processus d'extraction de caractéristiques consiste à convertir des données textuelles en vecteurs numériques afin de les utiliser avec des algorithmes d'apprentissage automatique. Une technique couramment utilisée pour cette conversion est TF-IDF (Term Frequency-Inverse Document Frequency) [1].

TF-IDF attribue un poids à chaque terme en fonction de sa fréquence dans un document spécifique et de son importance dans l'ensemble du corpus. Cette technique permet de représenter chaque document par un vecteur de caractéristiques pondérées, ce qui facilite l'apprentissage et la classification.

Dans notre cas, nous avons utilisé TF-IDF pour convertir notre ensemble de données textuelles en vecteurs de caractéristiques numériques. Ces vecteurs représentent les informations clés des documents et sont utilisés pour l'analyse des sentiments médicaux à l'aide d'algorithmes d'apprentissage automatique [5].

Le TF-idf (Term Frequency-Inverse Document Frequency) permet d'évaluer l'importance d'un terme dans un document par rapport à une collection de documents. Le poids du terme augmente proportionnellement à son nombre d'occurrences dans le document, mais diminue en fonction de sa fréquence dans la collection globale de documents. Ainsi, la fréquence inverse du document (idf) mesure l'importance du terme dans l'ensemble des documents. Dans le cas du TF-idf, on cherche à accorder un poids plus élevé aux termes moins fréquents, car ils sont considérés comme plus discriminants. Pour calculer l'idf, on prend le logarithme de l'inverse de la proportion de documents qui contiennent le terme spécifique [26].

$$TF-IDF=TF*IDF..... (11)$$

$$TF(k) = \frac{\text{le nombre de fois que le terme } k \text{ apparaît dans le document}}{\text{lasommedestermesdansleedocument}} (12)$$

$$IDF(k) = \log \frac{\text{le nombre total de documents}}{\text{lenombrededocumentscontenantletermeK}} (13)$$

Chapitre III Implémentation et Expérimentation

3.3.1.2 La classification

Pour la classification des sentiments des patines nous avons utilisé trois classifieur dans l'approche basée sur l'apprentissage automatique qui sont les Séparateurs à Vaste Marge (SVM), le Naïve Bayes (NB), les arbres de décisions (DT).

3.3.2 Approche lexicale

L'approche lexicale dans l'analyse des sentiments consiste à utiliser un dictionnaire comprenant des mots positifs et des mots négatifs pour évaluer la polarité des textes. L'implémentation de cette approche comprend les étapes suivantes

3.3.2.1 Création de dictionnaire

Nous avons téléchargé un lexique contenant des mots positifs et des mots négatifs pour notre approche lexicale. Pour enrichir le lexique avec des termes spécifiques au domaine médical, nous avons ajouté des mots de polarité liés à la santé, aux soins et aux traitements médicaux. Cette étape permet d'adapter le lexique à notre domaine d'application spécifique, améliorant ainsi la précision de l'analyse des sentiments dans le contexte médical.

3.3.2.2 Règles de classification

Les règles de classification utilisées dans cette approche sont basées sur l'utilisation de mots positifs et des mots négatifs, ainsi que sur la somme des polarités. Dans cette méthode, chaque mot du texte est évalué individuellement en le comparant à une liste de mots positifs et négatifs prédéfinis. Si un mot correspond à un mot positif, il est considéré comme ayant une polarité positive. De même, s'il correspond à un mot négatif, il est considéré comme ayant une polarité négative. Ensuite, les polarités attribuées à chaque mot sont agrégées en effectuant une somme des polarités positives et négatives. Si la somme est strictement positive, le texte est classé comme ayant une polarité positive. En revanche, si la somme est négative, le texte est classé comme ayant une polarité négative. Cette approche permet de quantifier le sentiment global exprimé dans le texte en prenant en compte à la fois les mots positifs et négatifs présents.

4. Mesures d'évaluations

Les mesures d'évaluation permettent de calculer différentes métriques de performance telle que la précision, le rappel et le score F1 et l'accuracy. Ces mesures permettent d'identifier les erreurs de classification spécifiques et d'évaluer la qualité des prédictions du modèle.

➤ La matrice de confusion

La matrice de confusion Tableau 10 est un outil utilisé pour évaluer les performances des algorithmes de classification. Elle est représentée par un tableau à quatre valeurs qui comparent les valeurs attendues avec les valeurs réelles. Chaque ligne du tableau représente une catégorie réelle, tandis que chaque colonne représente la catégorie prédite par le modèle [29].

Chapitre III Implémentation et Expérimentation

	Négative : 0	Positive : 1
Négative : 0	True Négative : TN	False Négative : FN
Positive : 1	False Positive : FP	True Positive : TP

Tableau 9: Matrice de confusion

- La signification de TP, TN, FP et FN est comme suit.
 - TN : True Négative.
 - TP: True Positive.
 - FP : False Positive.
 - FN : False Négative.

TN : Lorsque la prédiction est négative et que cela correspond à la réalité dans l'ensemble de données, cela signifie qu'un commentaire est réellement négatif et qu'il a été correctement prédit comme tel.

TP : Lorsque la prédiction est positive et correspond à la réalité dans l'ensemble de données, cela signifie qu'un commentaire est réellement positif et qu'il a été correctement prédit comme tel.

FP : Lorsque la prédiction est positive mais ne correspond pas à la réalité dans l'ensemble de données, cela signifie que le modèle a incorrectement classifié un commentaire comme positif alors qu'il était en réalité négatif.

FN : Lorsque la prédiction est négative mais ne correspond pas à la réalité dans l'ensemble de données, cela signifie que le modèle a incorrectement classifié un commentaire comme négatif alors qu'il était en réalité positif.

Donc, les métriques nécessaires pour analyser ce tableau (matrice de confusion) sont ; précision, rappel, F1score, accuracy.

➤**Précision** : La capacité du modèle de classification à ne renvoyer que des cas liés peut être mesurée par le taux de précision. Cela correspond au nombre de vrais positifs (TP) divisé par la somme des vrais positifs et des faux positifs (FP).

$$précision = \frac{TP}{TP+FP} \dots\dots\dots (14)$$

➤**Rappel** : (recall en anglais) Le rappel est une mesure de la capacité du modèle de classification à identifier tous les cas pertinents. Il est calculé en divisant le nombre de vrais positifs par la somme des vrais positifs et des faux négatifs.

$$Rappel = \frac{TP}{TP+FN} \dots\dots\dots (15)$$

➤ **Le score F1** : Le score F1 (F1-Score en anglais) est une métrique qui combine le rappel et la précision en utilisant la moyenne harmonique, ce qui permet de prendre en compte les deux mesures dans une seule équation.

$$F1 = 2 * \frac{\text{Précision} * \text{rappel}}{\text{Précision} + \text{rappel}} \dots\dots\dots (16)$$

➤ **Accuracy** : l'accuracy est une mesure de performance utilisée pour évaluer la précision d'un modèle ou d'une méthode de prédiction. Elle représente la proportion de prédictions correctes par rapport à l'ensemble des prédictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (17)$$

5. Expérimentation

L'étude expérimentale repose sur l'utilisation de trois algorithmes d'apprentissage supervisé et une approche lexicale pour l'analyse des sentiments des revues sur la qualité des soins. Nous avons utilisé l'accuracy et le F1_score comme mesure de comparaison entre les différents classifieurs utilisés. Pour évaluer les performances, nous avons utilisé les fonctions de la bibliothèque Sklearn, notamment la matrice de confusion (confusion matrix) et le rapport de classification (classification report), qui nous ont permis d'accéder à un ensemble de métriques de performance détaillées.

5.1 Approche apprentissage automatique

5.1.1 La division (Train/Test Split)

Avant d'utiliser l'algorithme d'apprentissage automatique supervisé, nous avons effectué une division des données du dataset en deux parties : la première partie pour l'entraînement du modèle et la deuxième partie pour le test et l'évaluation des performances. Lors de cette étape, nous avons effectué la division des données de la manière suivante : nous avons utilisé 80% des données pour l'entraînement et 20% pour les tests.

5.1.2 L'évaluation de résultats

Dans cette étape on va évaluer les résultats obtenus.

5.1.2.1 Classifieur Séparateur à Vaste Marge (SVM)

Nous avons utilisé le classifieur (SVM) pour analyser les sentiments des patients à propos des questionnaires et les classer comme positif ou négatif. Les résultats de SVM sur les corpus utilisés sont donnés comme suit le Tableau 10 présente la matrice de confusion du modèle

Chapitre III Implémentation et Expérimentation

généralisé par le classifieur SVM sur l'ensemble de données Data_Quality_Cares le Tableau 11 présente leur performance.

	Négative	Positive
Négative	27	5
Positive	0	21

Tableau 10 : Matrice de confusion de SVM sur Data_Quality_Cares (sans prétraitement)

	Précision	Rappel	F1_measure	Accuracy
Négative	1.00	0.84	0.92	0.90
Positive	0.81	1.00	0.89	
Moyenne	0.90	0.92	0.90	

Tableau 11 : Performances pour SVM sur Data_Quality_Cares (sans prétraitement)

Nous avons ensuite effectué une seconde expérimentation avec prétraitement et nous obtenons les résultats du Tableau 12 et 13.

	Négative	Positive
Négative	24	8
Positive	2	19

Tableau 12 : Matrice de confusion de SVM sur Data_Quality_Cares (avec prétraitement)

	Précision	Rappel	F1_measure	Accuracy
Négative	0.92	0.75	0.83	0.81
Positive	0.70	0.90	0.79	
Moyenne	0.81	0.82	0.81	

Tableau 13 : Performances pour SVM sur Data_Quality_Cares (avec prétraitement)

5.1.2.2 Classifieur Naïve Bayes (NB)

Nous avons utilisé le classifieur naïve bayes (NB) pour analyser les sentiments des patients à propos des questionnaires et classer comme positif ou négatif. Les résultats de NB sur le corpus utilisé sont donnés comme suit ; le Tableau 14 présente la matrice de confusion du modèle généré par le classifieur NB sur l'ensemble de données Data_Quality_Cares et le Tableau 15 présente leur performance.

Chapitre III Implémentation et Expérimentation

	Négative	Positive
Négative	27	5
Positive	1	20

Tableau 14 : Matrice de confusion de NB sur Data_Quality_Cares (sans prétraitement)

	Précision	Rappel	F1_measure	Accuracy
Négative	0.96	0.84	0.90	0.88
Positive	0.80	0.95	0.87	
Moyenne	0.88	0.89	0.88	

Tableau 15 : Performances pour NB sur Data_Quality_Cares (sans prétraitement)

Nous avons ensuite effectué une seconde expérimentation avec prétraitement et nous obtenons les résultats du Tableau 16 et 17.

	Négative	Positive
Négative	25	7
Positive	1	20

Tableau 16 : Matrice de confusion de NB sur Data_Quality_Cares (avec prétraitement)

	Précision	Rappel	F1_measure	Accuracy
Négative	0.96	0.78	0.86	0.84
Positive	0.74	0.95	0.83	
Moyenne	0.85	0.86	0.84	

Tableau 17 : Performances pour NB sur Data_Quality_Cares (avec prétraitement)

5.1.2.3 Classifieur Arbre de Décision DT

Nous avons utilisé le classifieur arbre de décision DT pour analyser les sentiments des patients à propos des questionnaires et les classer comme positif ou négative. Les résultats de DT sur le corpus utilisé sont donnés comme suit le Tableau 18 montre les résultats de la matrice de confusion du modèle généré par le classifieur DT sur l'ensemble de données Data_Quality_Cares et le Tableau 19 présente leur performance.

Chapitre III Implémentation et Expérimentation

	Négative	Positive
Négative	24	8
Positive	0	21

Tableau 18 : Matrice de confusion de DT sur Data_Quality_Cares (sans prétraitement)

	Précision	Rappel	F1_Score	Accuracy
Négative	1.00	0.75	0.86	0.84
Positive	0.72	1.00	0.84	
Moyenne	0.86	0.87	0.85	

Tableau 19 : Performances pour DT sur Data_Quality_Cares (sans prétraitement)

Nous avons ensuite effectué une seconde expérimentation avec prétraitement et nous obtenons les résultats du Tableau 20 et 21.

	Négative	Positive
Négative	24	8
Positive	2	19

Tableau 20 : Matrice de confusion de DT sur Data_Quality_Cares (avec prétraitement)

	Précision	Rappel	F1_Score	Accuracy
Négative	0.92	0.75	0.83	0.81
Positive	0.70	0.90	0.79	
Moyenne	0.81	0.82	0.81	

Tableau 21 : Performances pour DT sur Data_Quality_Cares (avec prétraitement)

Après avoir analysé les résultats obtenus, nous avons constaté que lorsque nous avons appliqué les algorithmes d'apprentissage automatique sur des données prétraitées, l'accuracy obtenue était inférieure à celle obtenue en utilisant des données non prétraitées. Cela indique que le prétraitement des données peut avoir un impact négatif sur les performances du modèle. Par conséquent, dans notre étude, nous avons sélectionné les modèles de classification basés sur des données non prétraitées.

Chapitre III Implémentation et Expérimentation

5.2. Approche à base de lexique

5.2.1 Attribution de polarité

Une fois les données prétraitées, chaque mot est comparé avec les entrées du dictionnaire. Si un mot correspond à une entrée positive, il est considéré comme positif, tandis qu'un mot correspondant à une entrée négative est considéré comme négatif.

5.2.2 Agrégation des polarités

Dans cette étape, les polarités attribuées à chaque mot sont agrégées afin d'obtenir une polarité globale pour le texte. Pour ce faire, différentes méthodes peuvent être utilisées, parmi lesquelles on peut sélectionner la méthode de sommation des polarités. Dans cette approche, les polarités positives et négatives des mots sont additionnées pour obtenir un score global. Si ce score est strictement positif le commentaire est classé positif, dans le cas contraire il est classé négatif. Cette méthode permet de quantifier le sentiment global exprimé dans le texte en prenant en compte les mots positifs et négatifs présents. En additionnant les polarités, on obtient ainsi une mesure de la polarité globale du texte.

5.2.3. L'évaluation de résultats

Nous avons utilisé l'approche basée sur le lexique pour analyser les sentiments des patients à propos des questionnaires et les classer comme positif ou négatif. Les résultats de l'analyse sur le lexique utilisé sont donnés comme suit. Le Tableau 22 montre les résultats de la matrice de confusion sur l'ensemble de données Data_Quality_Cares. Et le Tableau 23 montre les résultats de performances du modèle généré.

	Négative	Positive
Négative	103	28
Positive	7	126

Tableau 22 : Matrice de confusion de Lexique sur Data_Quality_Cares

	Précision	Rappel	F1_measure	Accuracy
Négative	0.94	0.79	0.85	0.86
Positive	0.82	0.95	0.88	
Moyenne	0.885	0.87	0.865	

Tableau 23 : Performances pour Lexique sur Data_Quality_Cares

Chapitre III Implémentation et Expérimentation

6. Comparaison des résultats obtenus

Dans cette étude, nous avons utilisé trois algorithmes d'apprentissage automatique SVM, NB et DT. Dans l'approche lexicale, on a utilisé un algorithme basé sur le lexique pour analyser les sentiments des patients à propos des questionnaires et les classer comme positif ou négatif. Les performances de tous les classifieurs utilisés ont été comparées selon l'échelle d'accuracy et le F1_Score les résultats sont présentés dans le Tableau 24.

	Lexique	SVM	NB	DT
Accuracy	0.86	0.90	0.88	0.84
F1_Score	0.86	0.90	0.88	0.85

Tableau 24 : Performances des quatre classifieurs

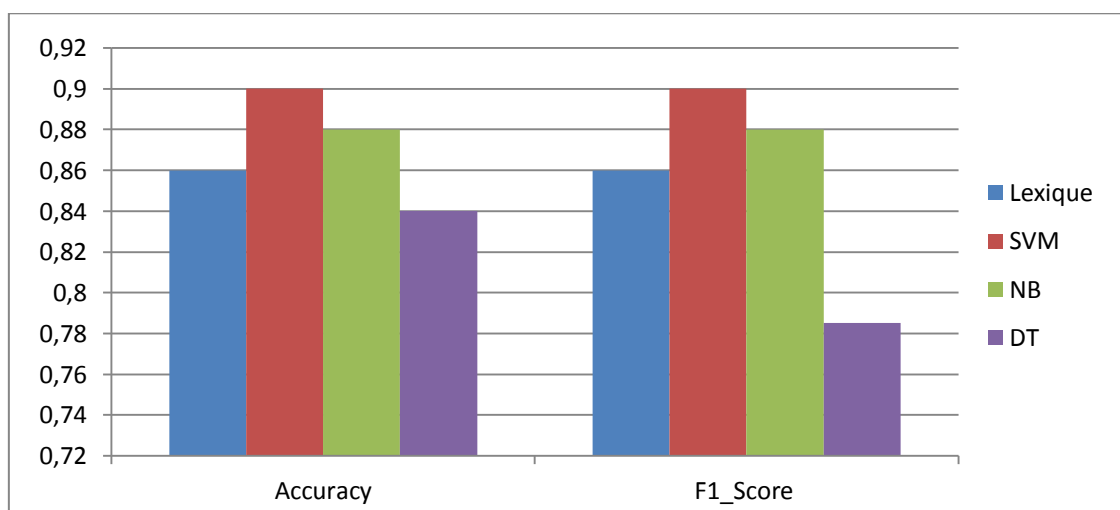


Figure 22 : Comparaison entre les approches utilisées.

Nos résultats démontrent que l'approche basée sur le lexique et l'approche basée sur l'apprentissage automatique (SVM) ont démontré des performances solides dans la classification des commentaires. Le modèle basé sur le lexique et le SVM ont donné des résultats très proches avec la meilleure précision globale, tandis que le Naive Bayes a présenté une précision légèrement inférieure. En revanche, l'arbre de décision a montré des performances inférieures lors de notre étude.

6. L'application développée

Dans le cadre de ce mémoire nous avons développé une application avec interface graphique pour l'analyse de sentiments médicaux.

Chapitre III Implémentation et Expérimentation

Pour organiser l'interface de notre application, nous avons utilisée différents éléments graphiques tels que des fenêtres, des étiquettes, des champs de texte, des boutons, etc. Voici l'organisation de l'interface :

6.1 Fenêtre d'analyse

- **Titre de la fenêtre** : N&N sentiment analysis.
- **Eléments** : Un champ de texte vous permettant de saisir votre commentaire en arabe. Des options pour choisir la méthode d'analyse qui vous convient le mieux parmi celles proposées. Un bouton "Analyser" qui, une fois cliqué, lance l'analyse du sentiment. Comme illustré dans la Figure 23.

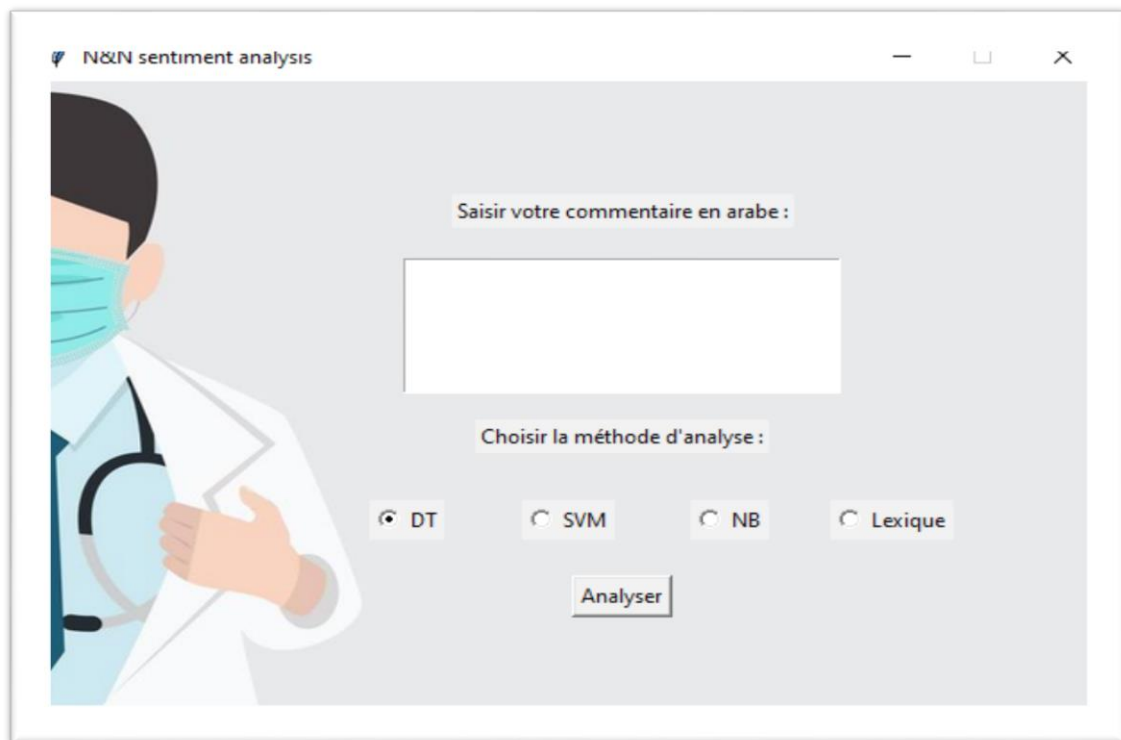


Figure 23 : Fenêtre principale

6.2 Fenêtre de résultats

- **Titre de la fenêtre** : "Résultats"
- **Eléments** : Affiche un message indiquant le sentiment de votre commentaire, qui peut être soit "POS" (positif) soit "Neg" (négatif). Comme illustré dans la Figure 24.



Figure 24 : Fenêtre de résultat

6.3 Exemple sur l'utilisation de l'application

Ajoutez votre commentaire en arabe dans le champ de saisie Figure 25.

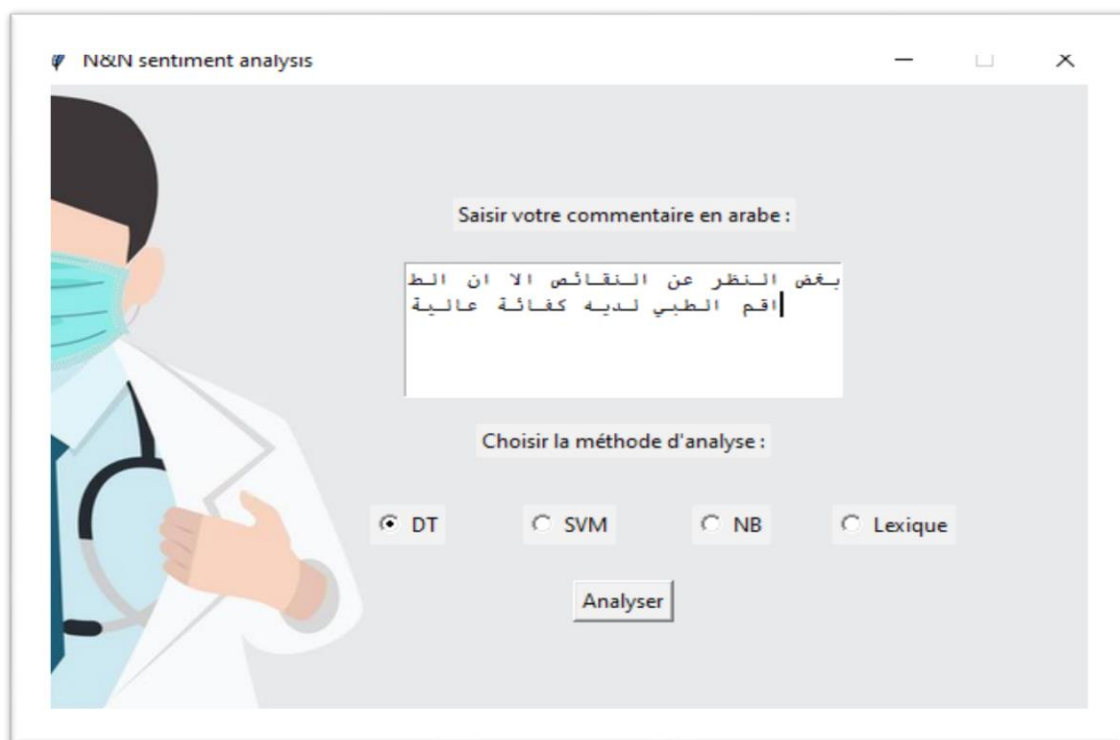


Figure 25 : Saisir de commentaire

Chapitre III Implémentation et Expérimentation

- Choisissez une méthode d'analyse parmi les quatre disponibles. Dans notre exemple, nous avons choisi la méthode lexicale Figure 26.

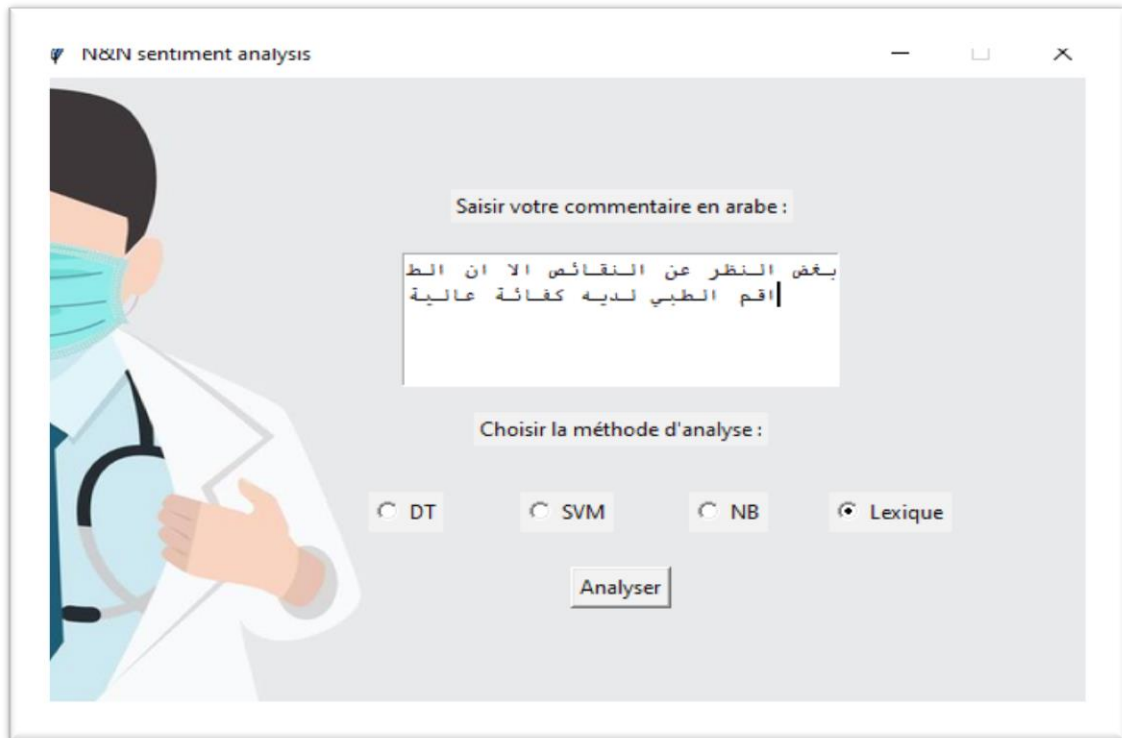


Figure 26 : Choix de méthode

- Afficher les résultats de l'analyse dans la fenêtre d'analyse, comme illustré dans la Figure 27.



Figure 27 : Résultat d'analyse

7. Conclusion

Dans ce chapitre, nous avons développé une application avec une interface graphique conviviale dans le but de permettre aux utilisateurs de classifier leurs commentaires. L'objectif principal de cette application est de fournir une plateforme où les utilisateurs peuvent saisir leurs commentaires et les faire classifier en utilisant différents algorithmes de classification et de transmettre ces commentaires aux société médicales afin d'améliorer leurs services.

Dans cette application, il y a quatre algorithmes de classification SVM, NB et DT, et un algorithme basé sur l'analyse lexicale. Les utilisateurs ont la liberté de choisir l'algorithme de leur choix pour classer leurs commentaires et obtenir les résultats de classification correspondants. Après avoir évalué les performances des différents algorithmes, nous avons constaté que le modèle SVM a donné les meilleurs résultats en termes de précision pour la classification des commentaires. Cela démontre l'efficacité de l'algorithme SVM dans cette tâche.

Conclusion Générale

Conclusion générale

Conclusion générale

Ce travail a traité l'analyse des sentiments médicaux sur la base de deux approches ; à savoir par lexicque et par apprentissage automatique. Nous fournissons ainsi un aperçu approfondi de ce domaine en plein essor. Nous avons met en évidence l'importance cruciale de l'analyse des sentiments dans le contexte médical dans le premier chapitre, "Fouille de sentiments médicaux". Il souligne comment cette analyse permet de comprendre les opinions et les émotions des patients, fournissant ainsi des informations précieuses pour la prise de décisions médicales et l'amélioration des soins de santé. Nous avons présenté dans le deuxième chapitre les approches d'analyse des sentiments, en détaillant les deux approches utilisées dans cette étude. L'approche lexicale, qui repose sur des dictionnaires et des ressources lexicales, permet d'associer des mots à des polarités sentimentales. Par ailleurs, l'apprentissage automatique, basé sur des algorithmes d'apprentissage, permet de construire des modèles capables de prédire les sentiments à partir de données d'entraînement. Le troisième chapitre, "Implémentation et réalisation", s'était concentré sur la mise en œuvre concrète des deux approches étudiées.

Dans le cadre de notre projet, nous avons collecté des données à partir de questionnaires pour obtenir les mots et expressions utilisés par les participants afin de décrire leurs expériences médicales. Ces données ont été utilisées pour effectuer une analyse des sentiments.

L'utilisation du prétraitement des données dans les algorithmes d'apprentissage automatique a donné des résultats moins précis par rapport aux modèles appliqués sur les données brutes, sans prétraitement. Après avoir testé différentes approches, nous avons constaté que le modèle SVM a donné les meilleures performances, suivi de près par l'approche lexicale, tandis que le modèle NB a obtenu des résultats satisfaisants et le modèle DT était moins performant. En ce sens, nous avons sélectionné le modèle SVM comme le plus approprié pour notre application d'analyse des sentiments dans le domaine médical.

Par ailleurs, lors de notre approche lexicale, nous avons exploré l'étape de suppression des mots vides, qui consiste à éliminer les mots courants tels que les pronoms, les prépositions et les conjonctions de notre analyse des sentiments. Cependant, nous avons remarqué que cette étape a eu un impact négatif sur l'exactitude du modèle, obtenant ainsi des résultats inférieurs à ceux obtenus sans la suppression des mots vides. Par conséquent, nous avons pris la décision de ne pas inclure cette étape dans notre application d'analyse des sentiments médicaux, afin d'améliorer les performances globales du modèle.

Ce mémoire fournit un aperçu approfondi de l'analyse des sentiments médicaux en mettant l'accent sur deux approches principales, lexicque et apprentissage automatique. Le travail souligne l'importance de cette analyse dans le domaine médical et présente des méthodologies complémentaires pour l'exploration des sentiments des patients. Les résultats obtenus et les discussions menées ouvrent la voie à de nouvelles recherches et à des applications futures dans ce domaine en constante évolution. Ce mémoire constitue ainsi une contribution précieuse à la compréhension et à l'amélioration de la prise en charge des patients dans le domaine de la santé.

Conclusion générale

Nous proposons comme perspective de ce travail l'enrichissement de notre corpus, Data_Quality_Cares, par d'autres commentaires qu'il soit ainsi plus représentatif des opinions des patients de la santé de Khenchela. La création d'autres corpus sera une autre alternative très bénéfique. Le lexique utilisé et enrichi donne des résultats promoteurs, l'amélioration de ce lexique dans des futurs travaux peut contribuer à un avancement important dans son efficacité et son utilisation.

Annex

Certificat de stage

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة الصحة

المؤسسة العمومية للصحة الجوارية خنشلة
المديرية الفرعية للموارد البشرية
رقم3.02.0...../م.ع.ص.ج.خ/2022

مقرر

إجراء تربص

إن مدير المؤسسة العمومية للصحة الجوارية خنشلة :

- بمقتضى الأمر رقم : 03-06 المؤرخ في : 15 يوليو 2006 والمتضمن القانون الأساسي العام للوظيفة العمومية
- بمقتضى المرسوم الرئاسي رقم: 54-23 المؤرخ في: 23 جمادى الثانية عام 1444 الموافق 16 جانفي 2023 ، يعدل المرسوم الرئاسي رقم: 304-07 المؤرخ في 17 رمضان عام 1428 الموافق 29 سبتمبر سنة 2007 الذي يحدد الشبكة الاستدلالية لمرتبات الموظفين ونظام دفع رواتبهم
- بمقتضى المرسوم التنفيذي رقم : 99-90 المؤرخ في 17 مارس 1990 المتعلق بسلطة التعيين والتسيير الإداري للموظفين ، أعوان الإدارات المركزية، الولايات، البلديات والمؤسسات ذات الطابع الإداري .
- بمقتضى المرسوم التنفيذي رقم 2007/140 المؤرخ في: 19 ماي 2007 المتضمن إنشاء المؤسسات العمومية الاستشفائية والمؤسسات العمومية للصحة الجوارية تنظيمها وتسييرها
- بناء على الطلب المقدم من طرف جامعة عباس لغرور-خنشلة- والمتمثل في إجراء تربص تطبيقي بالمؤسسة العمومية للصحة الجوارية - خنشلة -

باقترح من المدير الفرعي للموارد البشرية

يقرر

- المادة الأولى : (ت) يحول السيد(ة):قاسمي نسرين بصفتها : طالبة لإجراء تربص تطبيقي بالعيادة متعددة الخدمات حمو بوشوارب لمدة 20 يوم ابتداء من : 2022/12/24
- المادة الثانية : يكلف السيد المدير الفرعي للموارد البشرية والمدير الفرعي للمصالح الصحية كل في حدود اختصاصه بتنفيذ ما جاء في هذا المقرر

24 ديسمبر 2022

خنشلة في :

المدير
مدير فرعي للموارد البشرية
حقااص فريد

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة الصحة

المؤسسة العمومية للصحة الجوارية خنشلة
المديرية الفرعية للموارد البشرية
رقم301...../م.ع.ص.ج.خ/2022

مقرر
إجراء تربص

إن مدير المؤسسة العمومية للصحة الجوارية خنشلة :

- بمقتضى الأمر رقم : 03-06 المؤرخ في : 15 يوليو 2006 والمتضمن القانون الأساسي العام للوظيفة العمومية
- بمقتضى المرسوم الرئاسي رقم: 54-23 المؤرخ في: 23 جمادى الثانية عام 1444 الموافق 16 جانفي 2023 ، يعدل المرسوم الرئاسي رقم: 304-07 المؤرخ في 17 رمضان عام 1428 الموافق 29 سبتمبر سنة 2007 الذي يحدد الشبكة الاستدلالية لمرتبات الموظفين ونظام دفع رواتبهم
- بمقتضى المرسوم التنفيذي رقم : 99-90 المؤرخ في 17 مارس 1990 المتعلق بسلطة التعيين والتسيير الإداري للموظفين ، أعوان الإدارات المركزية، الولايات، البلديات والمؤسسات ذات الطابع الإداري .
- بمقتضى المرسوم التنفيذي رقم 2007/140 المؤرخ في: 19 ماي 2007 المتضمن إنشاء المؤسسات العمومية الاستشفائية والمؤسسات العمومية للصحة الجوارية تنظيمها وتسييرها
- بناء على الطلب المقدم من طرف جامعة عباس لغرور-خنشلة- والمتمثل في إجراء تربص تطبيقي بالمؤسسة العمومية للصحة الجوارية - خنشلة -

باقتراح من المدير الفرعي للموارد البشرية

يقرر

- المادة الأولى : (ت) يحول السيد(ة): سنوسي نريمان بصفقتها : طالبة لإجراء تربص تطبيقي بالعيادة متعددة الخدمات حمو بوشوارب لمدة 20 يوم ابتداء من: 2022/12/24
- المادة الثانية : يكلف السيد المدير الفرعي للموارد البشرية والمدير الفرعي للمصالح الصحية كل في حدود اختصاصه بتنفيذ ما جاء في هذا المقرر

خنشلة في : 24 ديسمبر 2022



مدير فرعي للموارد البشرية
حفاص فريد

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة الصحة

المؤسسة العمومية للصحة الجوارية خنشلة
المديرية الفرعية للموارد البشرية
رقم/م.ع.ص.ج.خ/2022

مقرر

إجراء تربص

إن مدير المؤسسة العمومية للصحة الجوارية خنشلة :

- بمقتضى الأمر رقم : 03-06 المؤرخ في : 15 يوليو 2006 والمتضمن القانون الأساسي العام للوظيفة العمومية
- بمقتضى المرسوم الرئاسي رقم : 304/07 المؤرخ في : 29 سبتمبر 2007 الذي يحدد الشبكة الاستدلالية لمرتبات الموظفين و نظام دفع رواتبهم
- بمقتضى المرسوم التنفيذي رقم : 99-90 المؤرخ في 17 مارس 1990 المتعلق بسلطة التعيين والتسيير الإداري للموظفين ، أعوان الإدارات المركزية، الولايات، البلديات والمؤسسات ذات الطابع الإداري .
- بمقتضى المرسوم التنفيذي رقم 2007/140 المؤرخ في: 19 ماي 2007 المتضمن إنشاء المؤسسات العمومية الاستشفائية والمؤسسات العمومية للصحة الجوارية تنظيمها وتسييرها
- بناء على الطلب المقدم من طرف جامعة عباس لغرور خنشلة والمتمثل في إجراء تربص تطبيقي بالمؤسسة العمومية للصحة الجوارية - خنشلة -

باقتراح من المدير الفرعي للموارد البشرية

يقرر

- المادة الأولى : (ت) يحول السيد(ة) : سنوسي نريمان بصفتها : طالبة لإجراء تربص تطبيقي بالعيادة متعددة الخدمات حمو بوشوارب ابتداء من : 2022/11/08 إلى غاية : 2022/11/28
- المادة الثانية : يكلف السيد المدير الفرعي للموارد البشرية ومدير المصالح الصحية كل في حدود اختصاصه بتنفيذ ما جاء في هذا المقرر

خنشلة في : 08 نونبر 2022.....

المدير
المديرية الفرعية للموارد البشرية
للموارد البشرية
خنشلة
فريد

Références Bibliographique

Références bibliographiques

Bibliographies

- [1] HADJI Nour ELhouda, LAICHI Nour ELhouda. L'Analyse Automatique des sentiments dans les textes Arabes. Département d'informatique. Université MOHAMED BOUDIAF – M'SILA. Juin 2022. Mémoire de Master. Champs de recherche et applications de TALN.
- [2] Cheni Rania Wissam. Analyse des sentiments arabes en utilisant L'apprentissage en profondeur. Département d'informatique. Université Mohamed Khider–BISKRA. 2020. Mémoire de Master.
- [3] Habes Yasmine. Application des méthodes d'Apprentissage Automatique dans l'Analyse des Sentiments des Tweets Arabes. Département d'informatique. Université Saad Dahlab– Blida 1. 21 janvier 2021. Mémoire de Master.
- [4] ZIANI Amel. La recommandation via l'analyse d'opinions. Département d'Informatique. Université de Badji Mokhtar – Annaba.2017/2018. Thèse de Doctorat LMD-3ème cycle.
- [5] Ahlam Bouterai, Randa Benamara. Analyse de Sentiments des malades, vers un système d'assistance médicale. Département d'Informatique. Université ABBES LAGHROUR– KHENCHELA. 2021/2022.Mémoire de Master.
- [6] Nedioui Med Abdelhamid. Techniques d'apprentissage automatique pour l'analyse et la fouille des sentiments dans les réseaux sociaux. Département Informatique. Université Mohamed Khi der – BISKRA.2020/2021.THESE.
- [7] <https://medium.com/@mehdihadji/analyse-des-sentiments-g%C3%A9n%C3%A9ralit%C3%A9s-99ab87503a5e>. Visiter 21/05/2023
- [8] Motaz Saad, Wesam Ashour. Arabic Morphological Tools for Text Mining. Conference Paper. Novembre2010.
- [9] MEGHNI Cherif, TAFERGHOUST Ahmed. UN MODELE DE DEEP LEARNING POUR L'ANALYSE DES SENTIMENTS. Département Informatique. Université Abbes Laghrou – Khenchela. 2020/2021. Mémoire de Master.
- [10] <https://www.cigref.fr/wp/wp-content/uploads/2016/09/Gouvernance-IA-CIGREF-LEXING-2016.pdf>.
- [11] <https://www.oracle.com/ca-fr/artificial-intelligence/machine-learning/what-is-machine-learning/>. Visiter le 21/05/2023.
- [12] BEZIEZ Ouail. Etude et comparaison de modèles de prédiction basés sur l'apprentissage automatique. Département Informatique. Université L'arbi Ben M'hidi – Oum El Bouagh. 2020 – 2021. Mémoire de Master.

Références bibliographiques

- [13] Mlle GUENNINECHE Amel. Prédiction des propriétés des matériaux par apprentissage automatique. Département de physique. UNIVERSITE ABOU-BEKR BELKAID – TLEMCEN. 29/06/2019. Mémoire de Master.
- [14] <https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>. Visiter 21/05/2023.
- [15] MELLAH Fouad. Techniques d'apprentissage automatique pour la prédiction de la maladie de l'artère coronaire. Université Abbés Laghrour – Khenchela. Département Informatique. 2020-2021. Mémoire de Master.
- [16] B. M. Randles, I. V. Pasquetto, M. S. Golshan and C. L. Borgman, "Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study," 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Toronto, ON, Canada, 2017, pp. 1-2, doi: 10.1109/JCDL.2017.7991618.
- [17] ATMANIOU Siham, MILI Feriel. Analyse des sentiments sur les avis des clients dans le E-Commerce. Université A. Mira – Béjaïa. Département d'Informatique. 06 Octobre 2021. Mémoire de Master.
- [18] Himeur Younes, Hadjadj DhiaaElhak. Mesure de la satisfaction de clients basée sur les Commentaires en ligne. UNIVERSITE ABBES LAGHROUR–KHENCHELA. Département d'Informatique. 2021/2020. Mémoire de Master.
- [19] Belhabib abdelkader, Lagha Omar. Développement d'une application à base de l'algorithme de classification k-means. Université Abou Bakr Belkaid – Tlemcen. Département d'Informatique. 28 Juin 2012. Mémoire de Licence.
- [20] Ulrick Serge MbaEngonga. Approches d'analyse des données issues des médias sociaux appliquées au domaine de la santé. UNIVERSITÉ DE SHERBROOKE. FACULTÉ DES SCIENCES. Essai présenté au CeFTI. en vue de l'obtention du grade de maître en technologies de l'information.
- [21] James Pustejovsky, Amber Stubbs. Natural Language Annotation for Machine Learning.
- [22] Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. In Proceedings of the 2018 International Conference on Digital Health (DH '18). New York : ACM, 2018.
- [23] Tokenising into Words and Sentences | What is Tokenization and it's Definition? *greatlearning*. En ligne. 31/05/2023. <https://www.mygreatlearning.com/blog/tokenization/>.
- [24] Divya Khyani, Siddhartha B S, Niveditha N M, Divya B M, An Interpretation of Lemmatization and Stemming in Natural Language Processing. 10, 2020, Journal of University of Shanghai for Science and Technology, Vol. 22.

Références bibliographiques

- [25] <https://datascientest.com/tf-idf-intelligence-artificielle>. Visiter 31/05/2023
- [26] FACTOUR Abdelhak. LEBOUAHLA Mounir, Développement d'un outil de recherche d'information sur web, Département Informatique, Université d'Ibn Khaldoun – Tiaret 2018/2019, Mémoire de Master
- [27] Bounab Sarra. L'enseignement du vocabulaire aux enfants Cas des élèves de 3ème année primaire, Université d'Oum El Bouaghi. 2018/2019. Mémoire de Master
- [28] <https://corpuslinguisticmethods.wordpress.com/2014/01/15/what-is-inter-annotator-agreement/>. Visiter 05/06/2023
- [29] Fatoumata Yalcoué, Amor Adnane .Machine Learning pour la maintenance prédictive Université d'Oum El Bouagh. 2020/2022. Mémoire de Master.