

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE DE TECHNOLOGIE

DEPARTEMENT Informatique

N° :.....



FILIERE : INFORMATIQUE

OPTION : IDO

**Mémoire présenté pour l'obtention
Du diplôme de Master Académique**

Par :

Chadi Manel

Bouzidi Fatima Nour El Houda

Intitulé

Classification des Sentiments en langue Arabe

Soutenu devant le jury composé de:

DR.Geurna A

Université de M'sila

Président

DR. BRAHIMI Belkacem

Université de M'sila

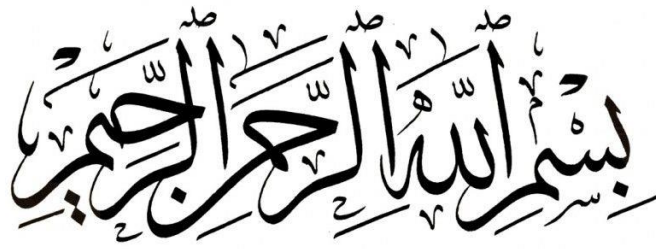
Rapporteur

DR.Yakoubi

Université de M'sila

Examineur

Année universitaire : 2020 /2021



Dédicaces

*Merci au bon Dieu de nous avoir accordé toute la détermination, la volonté et la force
pour qu'on puisse réaliser ce modeste travail*

*A mes chers parents pour tous leurs sacrifices, leur amour, leur soutien et leurs prières
tout au long de mes études*

A ma moitié Houda pour son encouragement et soutien

A mes âmes sœurs Salsabil et Nouha pour leurs appui et soutien moral

*A la personne avec qui j'ai grandi et qui était toujours à mes côtés lorsque personne
n'était SARRA*

A toutes mes sœurs et copines du cœur Lynda, Amina, Houda et Ferial

A mon frère midou pour être toujours présent à mes côtés

A toute ma famille, mes amis pour leurs soutiens durant le parcours universitaire

Merci d'être toujours là pour moi

Chadi Manel,



Dédicaces

Je dédie ce travail :

A mes très chers parents

Aucune dédicace, nul mot ne pourrait exprimer à leur juste valeur la gratitude et l'amour que je vous porte. Merci pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études, leurs encouragements m'ont toujours donné de la force pour persévérer et pour prospérer dans la vie.

Chaque ligne de cette thèse chaque mot et chaque lettre vous exprime la reconnaissance, le respect, l'estime et le merci pour mes parents.

A tous mes proches et les membres de ma famille plus particulièrement mes oncles BILEL, SADEK, OMAR et ma tante HADJIRA qui n'ont jamais douté de mes capacités à réussir et m'ont toujours apporté soutien et réconfort.

A mes chères sœurs SAFA, CHAHINEZ, ACHWAK pour leurs encouragements permanents, et leur soutien moral.

À mes amis avec qui j'ai partagé mes doutes, mes échecs, mais aussi mes réussites.

Également à tous ceux que je n'ai pas cités, et qui m'ont encouragé pendant ces années.

Bouzidi Fatima Nor El Houda,

Remerciements

En préambule nous tenons à remercier ALLAH qui nous a aidées et nous à données la patience et le courage durant cette année.

Nous tenons à remercier très chaleureusement Monsieur Brahimi Belkacem qui nous a permis de bénéficier de son encadrement. Les conseils qu'il nous a prodigué, la patience, la confiance qu'il nous a témoignés ont été déterminants dans la réalisation de notre travail de recherche.

Nous remercions aussi, Les membres du jury de nous avoir fait l'honneur de juger cette thèse. Veuillez accepter l'expression de notre profonde gratitude.

De plus, nous tenons également à remercier toutes les personnes qui ont contribué de près ou de loin, d'une manière directe ou indirecte à l'élaboration de ce travail de fin d'études

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragés au cours de la réalisation de ce mémoire.

Merci à tous et à toutes.

Table des matières

Table de matières	1
Liste des figures	3
Liste des tableaux	4
Introduction Générale	5
Chapitre1 : Concepts de Base	6
1.1 Introduction	7
1.2 Définitions	7
1.2.1 Sentiment.....	7
1.2.2 Opinion	7
1.2.2.1 Types d'opinion.....	8
1.2.3 Émotion	8
1.2.4 Les émotions de base : Paul Ekman.....	9
1.2.5 Réseaux sociaux	10
1.3. Les besoins de connaître des sentiments des autres	11
1.4. Conclusion	11
Chapitre2 : Fouille de Texte	12
2.1 Introduction	13
2.2 Text Mining	13
2.2.1 Définition.....	13
2.2.2 Le processus du text mining	14
2.2.2.1 La sélection des ressources.....	14
2.2.2.2 Les prétraitements	15
2.2.2.3 L'extraction d'information	16
2.2.2.4 La fouille de données	16
2.2.2.5 L'interprétation/validation	16
2.2.2.6 La représentation formelle	16
2.2.2.7 L'annotation et le raisonnement	17
2.2.3. La catégorisation de texte	17
2.2.4 Calcule Tf / Idf.....	18
2.2.5 Les méthodes de la classification automatique	19
2.2.5.1 La classification naïve bayésienne	19
2.2.5.2 Les k plus proches voisins	20
2.2.5.3 Les machines à vecteurs de support (svm)	20

2.3 Conclusion	22
Chapitre3 : Classification des Sentiments	23
3.1 Introduction	24
3.2 Classification des Sentiments	24
3.2.1 Définition.....	24
3.2.2 Les approches de la classification	25
3.2.2.1 Approches basée sur lexique	25
3.2.2.2 Approches basée sur corpus	25
3.2.2.3 Approches hybrides	26
3.3 Classification des sentiments en arabe	26
3.3.1 La langue arabe.....	26
3.3.2 La richesse de la langue arabe	27
3.3.3 Les difficultés de la classification en arabe	27
3.4 Etat de l'art (Travaux Connexes)	28
3.5 Conclusion	29
Chapitre4 : Conception et Réalisation	30
4.1 Introduction	31
Partie I: Construction du Corpus	31
4. I.1 Collection des données.....	31
4. I.2 Le dictionnaire d'émotions	33
4. I.3 Difficultés retrouvées lors de la collection	34
4. I.4 Prétraitements.....	35
4. I.5 Méthodologie suivi	35
4. I.6 Algorithme lexical.....	37
Partie II : Réalisation	40
4. II.1 Outils de développement	40
4. II.1.1 Langage JAVA	40
4. II.2 Environnement de développement	40
4. II.2.1 IntelliJ IDEA 2021.1	40
4. II.2.1 4.II.2.2Fonctionnalités de IntelliJ.....	41
4. II.2. 3 Rapidminer9.9	41
4. II.3 Exemple de la méthode réalisé	43
4. II.3 Etude de résultat	46
4.2 Conclusion.....	50
Conclusion générale	51
REFERENCES BIBLIOGRAPHIQUES	52

Liste des figures

Chapitre I

Figure I.1 La roue des émotions	10
Figure I.2 Enchaînement des réseaux sociaux 1978-2015.....	11

Chapitre 2

Figure 2.1 Processus du texte mining	14
Figure 2.2 Tâche général de fouille de texte.....	17
Figure 2.3 Le fonctionnement de l’algorithme KNN.....	20
Figure 2.4 La séparation du l’hyper plan par les SVM.....	21
Figure 2.5 Les vecteurs de support.	21

Chapitre 3

Figure 3.1 Le processus général de classification d’opinions.....	25
Figure 3.2 Les approches de classification des sentiments	26

Chapitre 4

Figure 4.I.1 Methodologie suivie.....	38
Figure 4.I.2 Schéma de la règle générale de getWords()	38
Figure 4.I.3 Exemple pour extraire un sentiment d’un commentaire	39
Figure 4.II.1 Interface de IntelliJ IDEA Community 2021.1	41
Figure 4.II.2 L’interface de RAPIDMINER9.9	42
Figure 4.II.3 Écran principal design RAPIDMINER.....	42
Figure 4.II.4 Utilisation de la fonction getWords()	43
Figure 4.II.5 Comparaison du fichier de commentaire et le dictionnaire	44
Figure 4.II.6 Mettre le contenu du fichier de commentaires dans le nouveau fichier	44
Figure 4.II.7 Concaténation du nouveau fichier avec les résultats	45
Figure 4.II. 8 Le fichier RESULT.....	45
Figure 4.II.4 Processus de l’analyse de sentiment dans RAPIDMINER.....	46
Figure 4.II.5 Pré-traitement du corpus dans RAPIDMINER.....	47
Figure 4.II.6 Résultats de prédiction	47
Figure 4.II.12 Performance de l’approche KNN	49

Liste des Tableaux

Chapitre 4

Tableau4.I.1 Analyse des travaux antérieurs sur l'analyse des sentiments en arabe	31
Tableau4.I.2 Chaque émotion avec le nombre de commentaires filtrés	33
Tableau4.I.3 Quelques commentaires et leur émotions	33
Tableau4.I.4 Lexique arabe et le nombre de mot dans chaque émotion.....	34
Tableau4.I.5 Lexique d'émotions	34
Tableau4.II.1 Les changements apportés dans chaque modèle	48
Tableau4.II.2 Résultats de performance après l'application de chaque modèle.....	48
Tableau 4.II.2: Résultats de performance de nouveau corpus.....	49

Introduction générale :

Les médias sociaux sont en train de devenir un moyen de communication où les gens peuvent exprimer en ligne leurs sentiments et leurs opinions sur divers sujets d'une manière qu'ils font rarement en personne. La détection des sentiments et des émotions dans les textes a fait l'objet d'une attention considérable ces dernières années. Ainsi, les termes "analyse des sentiments" et "détection des émotions" ont pris leur propre chemin pour devenir des éléments essentiels de la linguistique computationnelle et de l'analyse de textes. Ces termes sont conçus pour détecter les opinions et les émotions des personnes et les émotions des gens qui consistent en des expressions subjectives sur une variété de produits ou de décisions politiques.

Problématique :

Avec l'augmentation de la quantité d'informations disponible sur les réseaux sociaux, la classification des sentiments est l'une des tâches les plus difficiles car les réseaux sociaux deviennent de jour en jour multilingues.

Objectif :

Le but de ce travail est de fournir une méthode de classification de la fouille de texte pour classer chaque commentaire écrit en arabe selon le sentiment exprimé.

Structure de ce mémoire :

Ce mémoire est organisé en 4 chapitres qui sont résumés comme suit :

Dans le premier chapitre nous avons définis quelques notions qui entourent sur la classification des sentiments.

Dans le deuxième chapitre nous avons parlé sur le text mining.

Dans le troisième chapitre nous avons fait une représentation générale sur la classification des sentiments, la langue arabe et ses difficultés ainsi que quelques travaux connexes.

Dans le quatrième chapitre nous avons présenté la réalisation de notre travail commençant par la collecte du corpus, passant par les environnements du développement.

Chapitre 1 :

Concepts de Base

1.1 Introduction :

L'analyse des sentiments est le domaine d'étude qui analyse les opinions et les sentiments, les attitudes et les émotions des humains. C'est l'un des domaines de recherche les plus actifs dans le traitement d'une langue naturelle et est également largement étudié dans l'exploitation minière des données. En fait, cette recherche a été étendue de la science à l'administration des sciences sociales et de la science. L'importance croissante de ressentir que les sentiments coïncident avec la croissance des réseaux sociaux, tels que des forums de discussion, des blogs, des micro-blogues, Twitter et des réseaux sociaux.

Les systèmes d'analyse des sentiments sont appliqués dans presque tous les domaines d'affaires et sociaux parce que les opinions sont au centre de presque toutes les activités humaines et sont des influenceurs clés de nos comportements [1].

Dans ce chapitre, nous allons présenter quelques notions concernant les émotions.

1.2 Définitions :

1.2.1 Sentiment :

Le sentiment est la composante de l'émotion qui implique les fonctions cognitives de l'organisme, la manière d'apprécier. Le sentiment est à l'origine d'une connaissance immédiate ou d'une simple impression. Il renvoie à la perception de l'état physiologique du moment. Le sens psychologique de sentiment qui comprend un état affectif est à distinguer du sens propre de la sensibilité.

(Le sentiment fait partie intégrante de l'émotion, impliquant la fonction cognitive et la méthode d'appréciation d'un organisme. Ce sentiment découle d'une connaissance directe ou d'impressions simples. Il se réfère à la perception de l'état physiologique actuel. Les sensations psychologiques, y compris les états émotionnels, doivent être distinguées des sensibilités appropriées.)[2].

1.2.2 Opinion :

L'opinion est un jugement que l'on porte sur un individu, un être vivant, un phénomène, un fait, un objet ou une chose. Elle peut être considérée comme bonne ou mauvaise. L'opinion peut influencer et peut donner de bonnes ou mauvaises informations sur un sujet étudié au sein d'un groupe, d'une personne, d'un objet. Une opinion (terme issu du verbe latin opinari) est un ensemble de jugements que l'on se fait à propos d'un objet. Selon les Définitions du pseudo-Platon [2].

1.2.2.1 Types d'opinion :

L'opinion en tant qu'information de base est retrouvée en deux catégories principales :

a) Opinion usuelle ou comparative

- **Opinion usuelle** : L'opinion usuelle est une simple expression d'opinion, qui peut être abordée directement ou indirectement le thème principal. Actuellement, les opinions directes sont la méthode la plus couramment utilisée dans la plupart des études d'opinion, car elles sont simples et faciles à utiliser et peuvent facilement identifier différentes parties des opinions elles-mêmes.
- **Opinion comparative** : Une opinion comparative représente une opinion exprimée en comparant un sujet avec un autre. Cette comparaison permet de déterminer la valeur des opinions exprimées sur le premier sujet.

b) Opinion explicite ou implicite:

- **Opinions claires** : les opinions claires sont généralement des opinions subjectives, qui peuvent être exprimées de manière simple ou par comparaison.
- **Opinion implicite** : Une opinion implicite est une opinion objective générale, c'est-à-dire l'expression d'une opinion commune ou comparative. Ce type de point de vue est plus difficile à déterminer, et le nombre d'études sur le premier sujet est évidemment moins discuté.

1.2.3 Émotion :

Dans le dictionnaire *Le Robert* (1993), « émotion » est un mot issu de « motion » qui concerne le mouvement, terme apparaissant au XIII^e siècle en français comme en langues saxonnes et portant l'idée d'un mouvement qui s'accomplit ; la racine latine *emovere* signifiant « mettre en mouvement ». L'émotion va exister au début du XVI^e siècle par le mot « esmotion » qui induira la signification utilisée actuellement : l'émotion est un état de conscience complexe, généralement brusque et momentané, accompagné de signes physiologiques (par exemple : rougissement, sudation). Les **émotions** sont des réactions immédiates, directes et non pathologiques qui entraînent des modifications physiologiques intenses. C'est un état d'excitation qui perturbe la réflexion [3].

1.2.4 Les émotions de base : Paul Ekman :

Une émotion est une réaction affective passagère d'intensité plus ou moins forte, qui survient en réaction à un événement déclencheur. Le psychologue américain Paul Ekman, pionnier dans l'étude des émotions, a observé les expressions faciales dans diverses cultures et a dénombré 6 émotions fondamentales : la joie, la colère, la peur, la tristesse, la surprise, le dégoût. Elles servent de matériau de base à l'élaboration d'autres émotions dites secondaires. Le vocabulaire émotionnel décrit la palette et l'intensité des émotions, il est utile pour mieux les identifier et mieux se connaître.¹

1.2.4.1 La joie :

Est liée à la satisfaction d'un désir, la réussite d'un projet important à nos yeux. C'est un état de satisfaction et de bien-être qui se manifeste par de la gaieté et de la bonne humeur. Elle accroît notre énergie, la motivation et la confiance en soi.

1.2.4.2 La colère :

Est une réaction de protection. Elle résulte d'une frustration, d'un sentiment d'injustice, de la rencontre d'un obstacle, voire de l'atteinte à son intégrité physique ou psychologique.

1.2.4.3 La peur :

Est une émotion d'anticipation. Elle est utile lorsqu'elle nous informe d'un danger, d'une menace potentielle ou réelle car elle nous prépare à fuir, ou à agir. Elle peut être également liée à une appréhension, elle peut alors s'avérer stimulante ou bloquante.

1.2.4.4 La tristesse :

Est liée à une perte, une déception, un sentiment d'impuissance, un souhait insatisfait. Elle se caractérise par une baisse d'énergie, de la motivation.

1.2.4.5 La surprise :

Est provoquée par un événement inattendu, soudain, en lien avec un changement imminent ou par une révélation allant à l'encontre de notre perception, de nos représentations. Elle est généralement brève, puis s'estompe ou laisse place à une autre émotion.

1.2.4.6 Le dégoût :

Correspond à un rejet, une aversion physique ou psychologique envers un objet (nourriture...) ou une personne, perçus comme nuisibles.

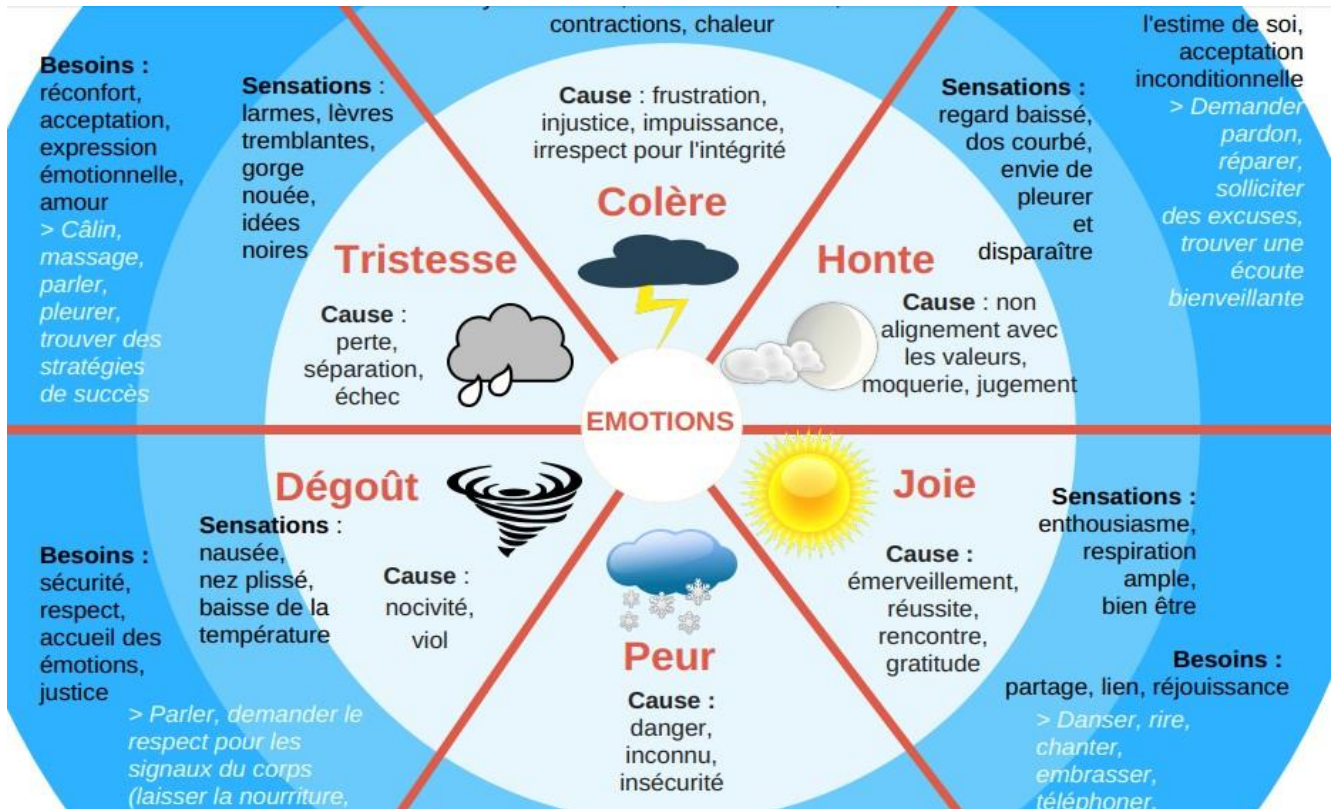


Figure 01 : La roue d'émotions [4].

1.2.5 Réseaux sociaux :

Dans le cadre d'une approche marketing, le terme de réseaux sociaux désigne généralement l'ensemble des sites internet permettant de se constituer un réseau d'amis ou de connaissances professionnelles et fournissant à leurs membres des outils et interfaces d'interactions, de présentation et de communication [5].

L'ère numérique et l'apparition des réseaux sociaux n'apportent rien de nouveau aux besoins initiaux, ils viennent plutôt accélérer la transmission de l'information et réduire le temps de réponse [6].

Les réseaux sociaux les plus connus sont Facebook, Twitter, LinkedIn, Viadeo, Pinterest, etc. Youtube peut également être considéré partiellement comme un réseau social dans la mesure où le service a développé des outils d'interactions entre ses membres [2].

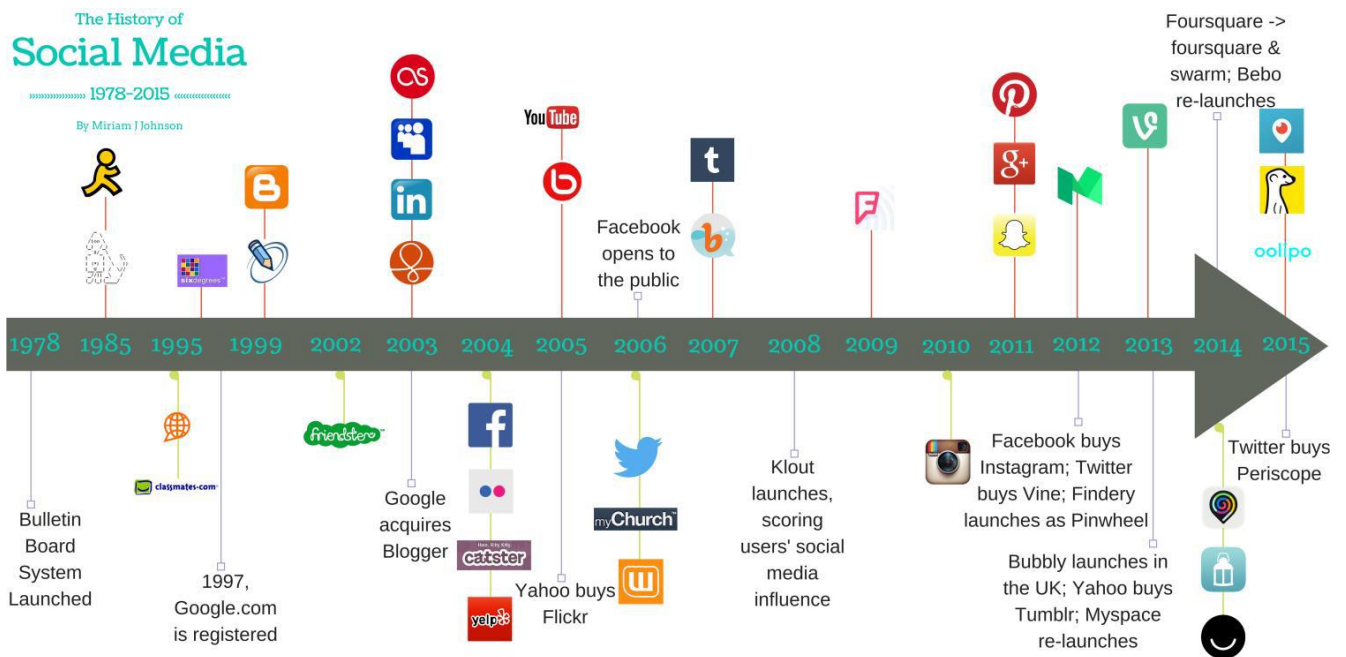


Figure 1-2:Enchainement des réseaux sociaux 1978-2015[7].

1.3 Les besoins de connaître des sentiments des autres :

Connaître l'opinion des autres personnes a toujours été un élément d'information important durant le processus de décision. Les gens demandent très souvent à d'autres de leur recommander un mécanicien automobile ou d'expliquer leur choix de votes aux élections par exemple. Avant de prendre des décisions, les gens s'intéressent énormément aux avis des autres personnes dans différents domaines. Ils consultent les avis des autres consommateurs avant d'effectuer un achat, ou regardent les avis des autres personnes avant de voir un film au cinéma ou avant d'acheter un disque. Grâce à l'Internet nous pouvons découvrir les opinions et les expériences de très grand nombre de personnes qui ne sont ni nos amis, ni les experts de domaines, mais des gens qui peuvent avoir les mêmes goûts que nous, et donc leurs opinions peuvent être très utiles pour nous avant de faire votre choix et d'avoir notre propre idée sur un sujet donné. Aujourd'hui, de plus en plus de personnes donnent leur avis sur différents sujets, ces avis sont à la disposition de tout le monde sur internet [8].

1.4 Conclusion :

Nous avons étudié dans ce chapitre quelques définitions de sentiments et tout ce qui concerne notre sujet, commençant par le sentiment, l'émotion et l'opinion. Passant par définir les émotions de base et les réseaux sociaux .en dernier nous avons mentionné l'importance de connaître les sentiments des autres dans nos jours.

La fouille de texte est le sujet le plus largement étudié nos jours dont nous allons présenter dans le prochain chapitre.

Chapitre 2 :

Fouille de texte

2.1 Introduction :

La recherche accorde ces dernières années, beaucoup d'importance au traitement des données textuelles. Ceci pour plusieurs raisons : un nombre croissant de collections mises en réseau et distribuées au plan international, le développement de l'infrastructure de communication et de l'Internet. Les traitements manuels de ces données s'avèrent très coûteux en temps et en personnel, ils sont peu flexibles et leur généralisation à d'autres domaines est presque impossible. En effet, les faibles coûts des machines en termes de stockage et de puissance ont incité les entreprises à accumuler toujours plus d'informations. Cependant, bien que la quantité de données à traiter ne cesse de croître, les spécialistes du domaine estiment que la quantité de données collectées dans le monde double tous les 20 mois. Jusqu'alors, les entreprises n'étaient pas en mesure de transformer leurs données en connaissances directement exploitables donc Une grande importance a été mise sur la fouille de texte (text mining). Ces documents comprennent souvent des données non structurées ou par fois semi structurées.

Dans ce chapitre nous allons exposer la définition et le processus général de la fouille de texte avec toutes ces étapes, ensuite nous présentons quelques algorithmes de la classification.

2.2 Text mining :

2.2.1 Définition :

Se distingue du data mining également par les moyens techniques spécifiques qu'il le faut employer pour traiter les données textuelles et non structurées.

Une définition générale du text mining est la suivante: l'extraction d'information à partir des formes ou patrons non manifestes (au sens de hidden patterns) dans des grands corpus de textes.

Autrement dit, l'objectif est le traitement de grandes quantités d'information qui sont disponibles sous une forme textuelle et non structurée. (Feldman et al., 1998a; Landau et al., 1998).

L'intelligence économique est sensée assurer aux acteurs économiques une information exploitable et utile, dans la mesure où cette information est textuelle (notes, lettres, rapports techniques, articles scientifiques, brevets, etc.), l'intérêt que le text mining peut représenter pour l'intelligence économique vis-à-vis du simple data mining est alors évident [9].

Dans un langage formel de représentation de connaissances et exploitées pour raisonner sur le contenu des textes [10].

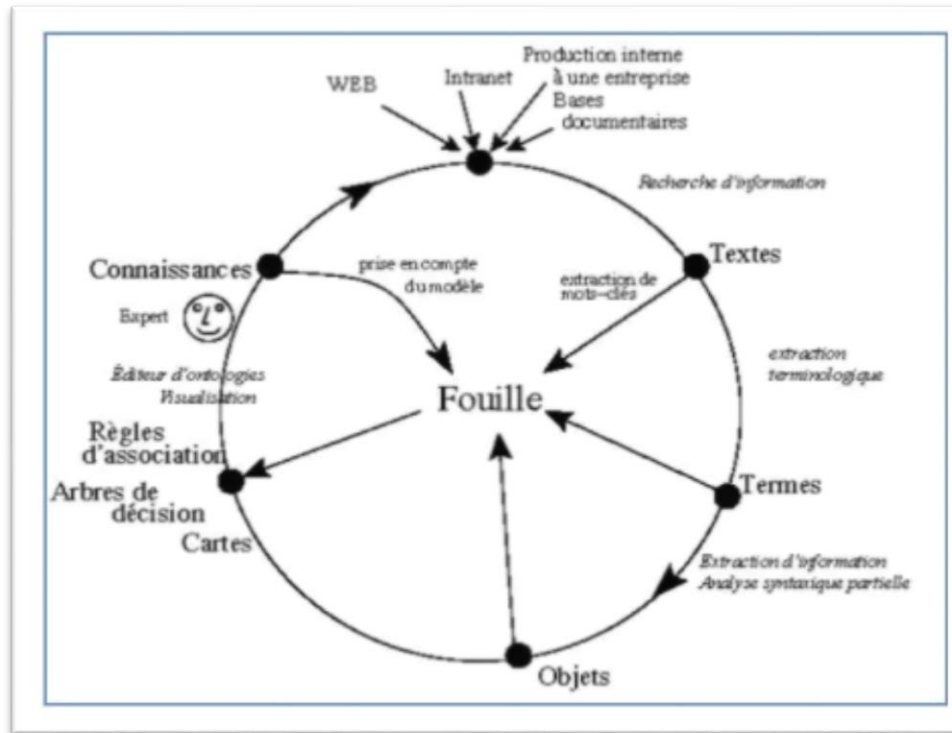


Figure2.1 : Processus du texte mining [11].

2.2.2 Le processus du text mining :

Je propose donc de décomposer la fouille de textes en 7 étapes qui pourraient, bien évidemment, être décomposées ou affinées [12] :

1. Sélection des ressources
2. Prétraitements
3. Extraction d'informations
4. Fouille de données
5. Interprétation/validation
6. Représentation formelle
7. Annotation, raisonnement

2.2.2.1 La sélection des ressources :

Sur un domaine, bien que souvent assez opportuniste, joue un rôle déterminant dans l'adéquation de l'ontologie construite à la tâche pour laquelle elle est construite. Certains textes apportent une connaissance de nature plutôt ontologique (connaissance consensuelle), alors que d'autres relatent une expérimentation dont les résultats pourraient être remis en cause par d'autres expérimentations (conditions différentes, expérience non reproductible. . .). Certaines ressources sont centrales par rapport à la tâche et peuvent être constituées, par exemple, par un ensemble de résumés d'articles scientifiques

sur les phénomènes de mutation génétique des bactéries. D'autres ressources sont plutôt de nature complémentaire et introduisent des connaissances déjà acquises dans un autre contexte. Il s'agit par exemple de prendre en compte les familles des bactéries. Ces connaissances ou informations peuvent se trouver dans des thésaurus, des ontologies, des textes didactiques. Ce type de ressources n'en est pas moins essentiel la qualité de l'ontologie produite.

Enfin, les ressources doivent pouvoir être complétées au cours des différentes itérations du processus de fouille, pour pallier les manques identifiés par les experts lors de la validation des modèles extraits des données.

Les techniques impliquées dans cette étape sont issues de la recherche d'information : requête sur des bases documentaires, classification ou catégorisation des réponses. La difficulté consiste à réduire au mieux la dispersion (éviter des textes sans intérêt pour l'ontologie à construire) tout en préservant une bonne couverture des phénomènes.

2.2.2.2 Les prétraitements :

Sont de natures très différentes. Cela peut concerner des questions d'encodage de caractère, de structure de documents, d'étiquetage morpho-syntaxique. . . Les outils généralement utilisés dans la suite du processus de fouille de textes travaillent sur du texte seul, sans mise en forme (ou une mise en forme très réduite telle que l'italique, la grasse. . .). Il est difficile de délimiter ce qui concerne les prétraitements ou l'extraction d'information dans l'absolu. Ainsi, la segmentation d'un texte peut extraire la partie pertinente d'un texte et constituer un prétraitement ou bien permettre de découper un texte en plusieurs sous-parties et pourrait alors être assimilée aux premières étapes d'extraction d'information.

Dans notre cas La plupart des outils d'analyse des sentiments effectuent trois étapes principales de prétraitement des données avant d'appliquer les techniques de classification, afin de préparer les textes arabes, qui sont :

- **la segmentation d'un texte** peut extraire la partie pertinente d'un texte et constituer un prétraitement ou bien permettre de découper un texte en plusieurs sous-parties
- **Normalisation** : le processus de transformation du texte afin d'être cohérent en convertissant toutes les formes variées d'un mot/lettre en une forme commune. En convertissant toutes les différentes formes d'un mot/lettre en une forme commune. La normalisation pour le texte arabe a consisté en les étapes suivantes [13] :
 - Suppression des diacritiques : par exemple, " العَرَبِيَّةُ " en العربية.
 - Suppression de " ال " au début des mots : par exemple, العربية sera " عربية " .
 - Remplacement de la lettre "ة" par "ه".
 - Remplacer la lettre "ى" par "ي".

2.2.2.7 L'Annotation et le raisonnement :

Dans mon projet de recherche, l'annotation sémantique fait partie du raisonnement sur les textes dans le sens où l'annotation est construite de façon progressive, avec des connaissances de plus en plus riches. Maintenir la cohérence entre les connaissances et les textes annotés se heurte au fait que la trace linguistique d'un concept peut être complexe et qu'un certain nombre d'outils d'annotation qui se rapprochent d'une indexation terminologique (la recherche du terme associé à un concept), ne fonctionnent plus. D'autre part, la trace dans les textes n'est souvent que partielle, des connaissances implicites (et donc externes aux textes) devant être introduites pour la construction de la base de connaissances et la description des concepts. Quant au raisonnement, il est probablement la raison de vivre des bases de connaissances. Un raisonnement erroné dans un système à base de connaissances doit impliquer des modifications ou des enrichissements de la base de connaissances, avec, si possible, un contrôle de la non-régression du système (des réponses du système, précédemment correctes qui ne le seraient plus après modification). Cette dernière étape est donc bien indissociable du processus d'extraction de connaissances.

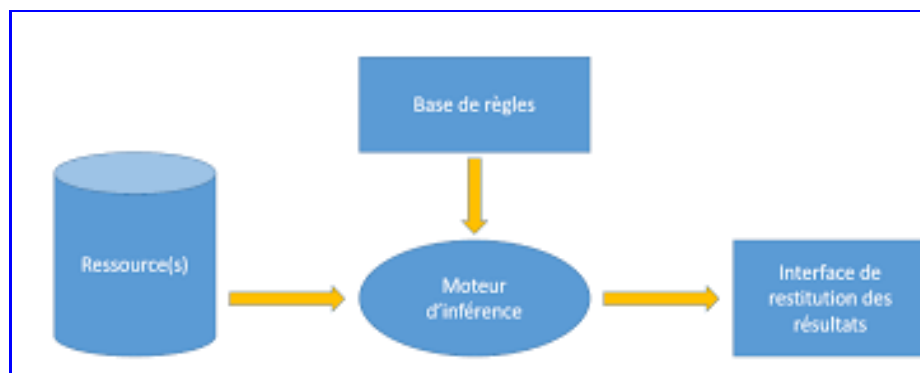


Figure 2.2 : Tâche général de fouille de texte) [31].

2.3 La catégorisation de texte :

La classification du texte est une partie importante du text Mining, la catégorisation de texte (Également connu sous le nom de classification de texte) est la tâche de trier automatiquement un ensemble de documents dans un ensemble prédéfini de catégories (ou classes) Sebastiani, (2005). Ces

documents peuvent être classés en utilisant les méthodes d'apprentissage automatique tel que (naïve bayes, decision tree, K-nearest Neighbors (KNN), Support Vector Machines (SVM), et les réseaux de neuronaux) [14].

Définition du problème Sebastiani, (2002) : La tâche générale de catégorisation de texte peut être formellement définie comme la tâche d'approximation d'une catégorie inconnue par une fonction d'affectation

$$F : D \times C \rightarrow \{0, 1\},$$

Où D est l'ensemble de tous les documents possibles $\{d_1 \dots d_n\}$ et C'est l'ensemble des catégories prédéfinies $\{c_1 \dots c_m\}$.

La valeur de F (d, c) est 1 si le document d appartient à la catégorie c et 0 sinon.

La fonction d'approximation

$$M : D \times C \rightarrow \{0, 1\}$$

Est appelé un classifieur, et la tâche est de construire un classifieur en utilisant des algorithmes d'apprentissage automatique pour faire des prédictions précises fondées sur des observations passées avec la production de résultats comme « proche » que possible de la véritable fonction catégorie d'affectation F.

La classification du texte a été appliquée à une grande variété d'applications pratiques :

- Comprendre et analyser les sentiments de l'audience à partir des médias sociaux.
- Filtrage des emails spams et non-spam.
- Marquage automatique des requêtes des clients.
- Catégorisation des articles d'actualité topics en sujets prédéfinis catalogage des articles de Presse.

2.4 Calcule Tf / Idf :

Le TF-IDF (de l'anglais *termfrequency-inverse document frequency*) est une méthode de pondération souvent utilisée en recherche d'information et en particulier dans la fouille de textes. Cette mesure statistique permet d'évaluer l'importance d'un terme contenu dans un document, relativement à une collection ou un corpus. Le poids augmente proportionnellement au nombre d'occurrences du mot dans le document. Il varie également en fonction de la fréquence du mot dans le corpus. Des variantes de la formule originale sont souvent utilisées dans des moteurs de recherche pour apprécier la pertinence d'un document en fonction des critères de recherche de l'utilisateur [2].

2.4.1 TF :

TF est l'abréviation de l'anglais terme frequency (fréquence du terme). Il détermine la fréquence relative d'un mot ou d'une combinaison de mots dans un document. Cette fréquence du terme sera comparée à la survenance de tous les autres mots restants du texte, du document ou du site web analysé. Cette formule utilise un logarithme qui se lit comme suit [15] :

TF(t) = (Nombre de fois que le terme t apparaît dans un document) / (Nombre total de termes dans le document). (1)

Le logarithme atteste qu'une augmentation visible du mot-clé dans le texte ne mène pas à une amélioration de sa valeur dans le calcul. Alors que la densité du mot-clé calcule principalement la distribution en pourcentage d'un seul mot dans le texte (en relation avec le nombre total de mots restant), le terme frequency factorise également en proportion de tous les mots utilisés dans le texte.

2.4.2 IDF :

L'IDF calcule l'Inverse Document Frequency (la fréquence inverse du document) et complète l'analyse de l'évaluation du mot. Il agit en tant que correct ifdu TF. L'IDF inclut dans le calcul la fréquence des documents pour un mot précis, autrement dit l'IDF compare le chiffre correspondant à tous les documents connus avec le nombre de textes contenant le mot en question. Le logarithme suivant condense les résultats :

IDF (t) = log_e (Nombre total de documents / Nombre de Documents contenant le terme t). (2)

2.5 Les méthodes de la classification automatique :

2.5.1 La classification naïve bayésienne :

Cette méthode se base sur le théorème de Bayes permettant de calculer les probabilités conditionnelles. Un classifieur bayésien naïf suppose que l'existence d'une caractéristique pour une classe, est indépendante de l'existence d'autres caractéristiques. Les classifieurs bayésiens naïfs peuvent être entraînés efficacement dans un contexte d'apprentissage supervisé.

L'avantage du classifieur bayésienne naïf est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables. En effet, l'hypothèse d'indépendance des variables permet de se contenter de la variance de chacune d'entre elle pour chaque classe, sans avoir à calculer de matrice de covariance [16].

2.5.2 Les k plus proches voisins :

C'est une approche très simple et directe. Elle ne nécessite pas d'apprentissage mais simplement le stockage des données d'apprentissage. Son principe est le suivant. Une donnée de classe inconnue est comparée à toutes les données stockées. On choisit pour la nouvelle donnée la classe majoritaire parmi ses K plus proches voisins (Elle peut donc être lourde pour des grandes bases de données) au sens d'une distance choisie.

Afin de trouver les K plus proches d'une donnée à classer, on peut choisir la distance euclidienne [17]. Soient deux données représentées par deux vecteurs x_i et x_j , la distance entre ces deux, données est donnée par :

$$d(x, y) = \sqrt{\sum^m (x - y)^2}$$

Paramètre : le nombre K de voisins
Contexte : un échantillon de L textes classés en $C = c_1, c_2, \dots, c_n$ classes
Début
 Pour chaque texte T faire
 Transformer le texte T en vecteur $T = (x_1, x_2, \dots, x_m)$.
 Déterminer les K plus proches textes du texte T selon une métrique de distance.
 Combiner les classes de ces K exemples en une classe C.
 Fin pour
Fin
Sortie : le texte T associé à la classe C.

Figure2.3 : Le fonctionnement de l'algorithme KNN [17].

Tapez une équation ici.

2.5.3 Les machines à vecteurs de support (svm) :

Les SVM sont un ensemble de techniques d'apprentissage supervisé de nouvelle génération basés sur les progrès récents de la théorie de l'apprentissage statistique [18], ils sont basés sur le principe de la maximisation de la marge (séparation des classes). Principalement conçues pour résoudre ces problèmes de discrimination, permettant de décider à quelle classe appartient un échantillon. Cependant, ils peuvent aussi résoudre des problèmes de régression visant à prédire la valeur numérique d'une variable (Scholkopf et Smola, 2001).

Le but de SVM est de trouver un classificateur qui sépare au mieux les données et maximise la distance entre ces deux classes. Ce dernier est un classificateur linéaire appelé hyperplan. Comme montré dans la Figure2.4, cet hyperplan sépare les deux ensembles de points.

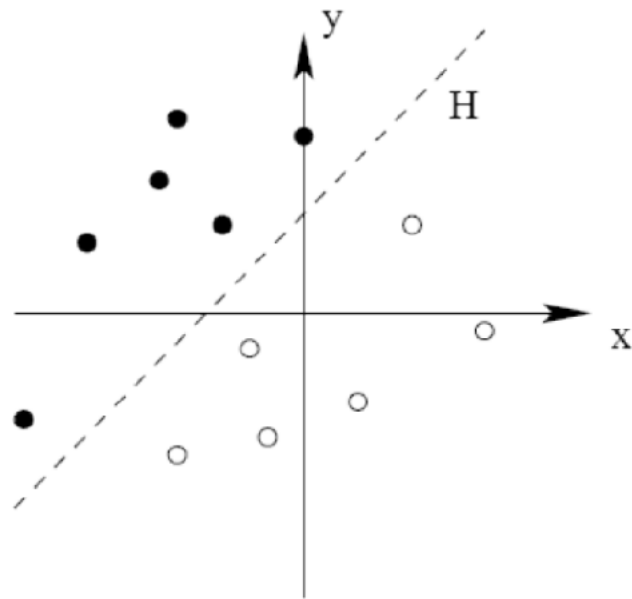


Figure2.4 : La séparation du l'hyper plan par les SVM.

Les points les plus proches, qui seuls sont utilisés pour la détermination de hyperplan, sont appelés **vecteurs de support**.

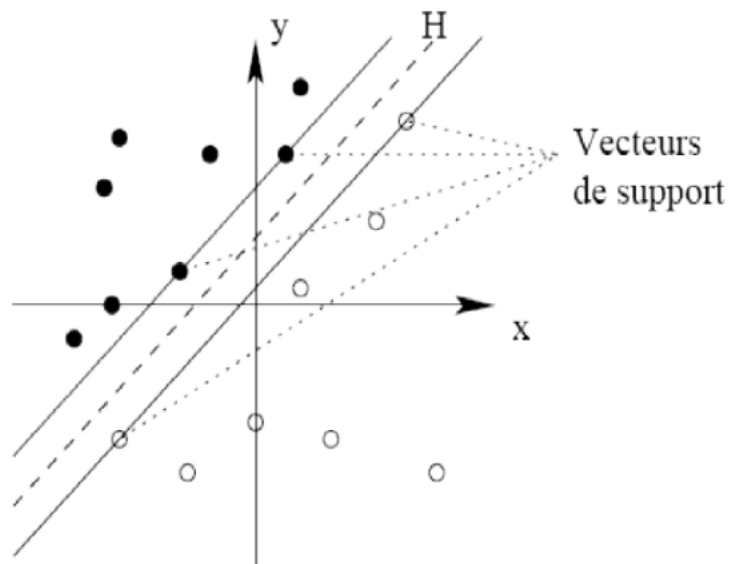


Figure2.5:Les vecteurs de support.

2.6 Conclusion :

Le text Mining est une méthodologie qui automatise la synthèse de connaissances à partir de gros volumes de données. L'essor de cette technologie est le résultat d'un accroissement dramatique de l'information numérique .La préparation des données est une étape importante, si ce n'est primordial, du processus d'extraction de connaissances à partir de données. En schématisant, il s'agit de définir au mieux les individus et la représentation utilisée pour l'apprentissage.

Dans ce chapitre, nous avons présenté le text mining et quelques algorithmes pour la classification. Dans le chapitre suivant, nous décrivons de manière plus détaillée la définition de la classification des sentiments et ces approches.

Chapitre 3 : Classification des Sentiments

3.1 Introduction :

L'intérêt porté à l'arabe et ses dialectes a largement augmenté au cours de ces dernières années. Un intérêt principalement dû à la proportion des locuteurs arabes au sein des médias sociaux. Actuellement, les sites Web de commerce électronique permettent à leurs utilisateurs d'écrire des critiques ou des commentaires sur les produits qu'ils ont achetés. L'information provenant des commentaires des utilisateurs est cruciale pour aider d'autres utilisateurs potentiels à décider d'acheter ou non un article en fonction de l'expérience et des opinions des autres utilisateurs au sujet d'un produit particulier. En outre, les fabricants peuvent également recueillir les commentaires des utilisateurs à partir des évaluations en ligne afin d'améliorer leurs produits. Cependant, avec l'augmentation du nombre d'utilisateurs achetant des produits, le nombre de commentaires augmente également au fil du temps et il n'est pas possible pour les utilisateurs ou les fabricants de lire tous les commentaires pour connaître les opinions des utilisateurs précédents sur un certain produit. En outre, certains des commentaires sont longs, ce qui fait difficile pour les utilisateurs de reconnaître les bonnes et les mauvaises caractéristiques du produit lorsqu'il s'agit de décider si le produit vaut la peine d'être acheté ; ou pour les fabricants de décider si le produit doit être amélioré. Un processus d'analyse d'opinions, qui peut analyser si un utilisateur fournit un bon ou une mauvaise opinion sur un certain produit, est important et hautement souhaitable pour les utilisateurs potentiels et les fabricants car il permet de recueillir facilement des informations utiles sur les produits à partir d'un grand nombre de commentaires, ce qui les aide à prendre des décisions basées sur les opinions des autres.

La richesse des médias sociaux en termes d'opinions, d'émotions et de sentiments a suscité l'intérêt de la communauté de recherche à se pencher beaucoup plus sur les problématiques liées à l'analyse sémantique et plus particulièrement sur l'analyse de sentiments de l'arabe et ses dialectes. L'analyse de sentiments consiste à déterminer la valence (positive, négative ou neutre) d'un texte donné.

Ce chapitre présente une brève revue de littérature sur la classification des sentiments , en mettant l'accent sur ses concepts de base et ses approches, et en se concentrant sur la fouille d'opinions en langue arabe.

3.2 Classification des Sentiments

3.2.1 Définition :

La classification de l'émotion varie selon les chercheurs, l'émotion de base générale trouvée dans la plupart des études de recherche comprend le bonheur, la tristesse, la colère, la peur, le dégoût, la surprise, où ces émotions ont été basées sur un plan bidimensionnel communément appelé le plan de valence-arousal. Les émotions positives peuvent être classées comme le bonheur ou la surprise, tandis

que les émotions négatives peuvent être associées à la tristesse, la colère, la peur et le dégoût. Le stress émotionnel peut être influencé par des expériences négatives sur une période longue et continue [19].

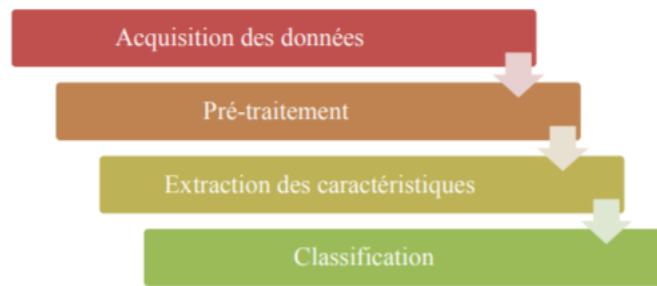


Figure3.1 : Le processus général de classification d'opinions [10].

3.2.2 Les approches de la classification :

Les techniques de classification des sentiments sont habituellement divisées en trois approches :

3.2.2.1 Approches basée sur lexicque :

Appelé aussi symbolique ou linguistique, jusqu'à maintenant, la plupart des études de l'analyse des sentiments se sont basées sur cette méthode. Elle permet d'identifier la polarité d'un texte à l'utilisation de deux ensembles de mots, ceux qui expriment un sentiment positif et ceux qui expriment un sentiment négatif.

Le modèle compte dans le texte le nombre de mots positifs et le nombre de mots négatifs, la somme donne une évaluation globale du sentiment de texte, si le nombre de mots positifs l'emporte sur celle de mots négatifs, le texte considéré comme positif, inversement, le texte est considéré comme négatif, éventuellement neutre si les nombres sont égaux [20]. Exemple d'algorithme : Support Vector Machine (SVM)...

3.2.2.2 Approches basée sur corpus :

Appelé aussi statistique, inversement à la méthode précédente, l'analyse automatique des sentiments basé sur corpus n'a pas besoin d'un lexique de mots positifs et négatifs, elle a besoin de deux corpus annotés (éventuellement un seul, si on utilise l'apprentissage non supervisé), le premier corpus destiné à l'apprentissage (l'entraînement), à partir les annotations, le modèle automatiquement sera capable de faire une analyse équivalente et d'une manière autonome.

Le deuxième corpus destiné au test, son rôle est de vérifier la performance du modèle, si le cas est idéal, le résultat d'analyse devrait être cent pour cent avec ceux du premier corpus. Pour améliorer la performance du modèle, il est important que le corpus d'entraînement doive être représentatif pour le corpus d'évaluation [20]. Exemple d'algorithme : réseau de neurone...

3.2.2.3 Approches hybrides:

Cette approche est appelée aussi classification semi-supervisée, elle combine les points forts de deux approches précédentes, il y a trois façons de faire. La première est d'exploiter les outils linguistiques pour élaborer le corpus puis classer les textes par un outil d'apprentissage supervisé. La deuxième façon est d'utiliser l'apprentissage automatique pour établir le corpus d'opinion nécessaire à l'approche basée sur lexic. La troisième façon est le conjointement des deux approches précédentes et la combinaison de leurs résultats soit par un système de vote soit par un algorithme d'apprentissage [21]. Exemple d'algorithme : S3VM (Semi-Supervised Support Vector Machine).

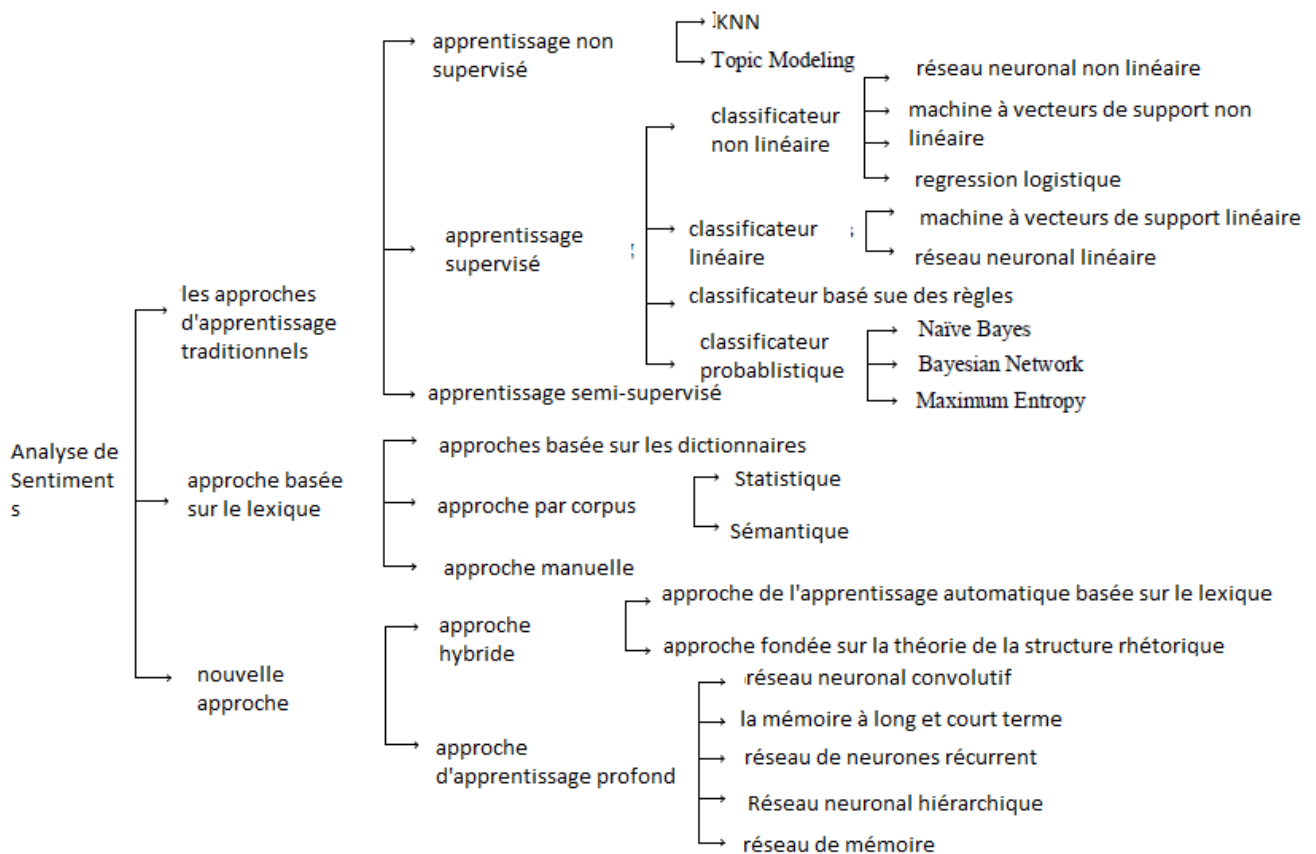


Figure3.2 : Les approches de classification des sentiments [22].

3.3 Classification des sentiments en arabe :

3.3.1 La langue arabe :

L'arabe est une langue qui sert plutôt à la poésie qu'à la science car il est incapable d'accompagner le progrès scientifique, contrairement au français et à l'anglais. Il appartient à la famille des langues sémitiques, tels que l'hébreu, le phénicien, l'araméen, le syriaque, etc. l'histoire des Arabes remonte à environ trois millénaires ; ils ont donné leur nom à la péninsule qu'ils habitaient, l'Arabie.

Leur entrée spectaculaire dans la grande histoire est due à l'avènement de l'Islam au VII^{ème}.s.j. L'arabe fait partie du groupe sud de ces langues et partage avec lui certaines caractéristiques dont :

- Un système phonologique plus proche du sémitique ancien avec une forte proportion de sons gutturaux (q, ʕ, ...).

-Un système morphologique à dérivation comportant principalement des verbes affixés et des pluriels internes (brisés).

Ainsi, l'arabe semble être une langue médiane, de jonction, faisant la synthèse des langues sémitiques. [23] L'origine du mot "Arabe" reste inconnu, malgré il y a beaucoup de recherches, le radical "arab" désigne le désert et c'est un mot araméen "arâbâh", le mot arabe peut dériver aussi de la racine sémitique Abhar "se déplacer", mais l'étymologie arabe considère que le mot "arabe" est dérivé du verbe "exprimer".

3.3.2 La richesse de la langue arabe :

La langue arabe s'écrit au moyen de 28 lettres, le terme "abjad" désigne le système d'écriture, elle dispose le plus grand nombre de mots avec plus de 12 millions de mots où l'Anglais 600 000 mots, le Français 150 000, mots, le Russe 130 000 mots Elle est une langue très riche, Il y aurait 80 termes différents pour identifier le miel, 200 pour le serpent, 500 pour le lion, 1000 pour le chameau et l'épée et jusqu'à 4400 pour définir l'idée de malheur, les grammairiens arabes prétendent que toutes les racines ont été originalement des verbes, où le nombre de ces racines en réalité est de 6000[24].

3.4 Les difficultés de la classification en arabe :

Peu de travaux ont été réalisés dans le domaine de l'analyse des sentiments en arabe. Plusieurs raisons peuvent expliquer le manque d'études dans ce domaine. Assiri dans [25] a mentionné deux raisons principales :

- 1- le financement limité de la recherche dans ce domaine
- 2- l'arabe a une morphologie très complexe par rapport aux autres langues. Par rapport à la morphologie d'autres langues. La complexité et la variété des dialectes arabes exigent des procédures avancées de prétraitement et de construction de [26], [27].
- 3- La grande distance entre MSA et quelques dialectes.
- 4- Une racine peut prendre plusieurs formes en fonction du contexte. [30]
- 5- La répétition d'une lettre plusieurs fois pour intensifier le sens ou le sentiment (بزازاف).
- 6- L'arabe a divers signes diacritiques, la présence ou l'absence de tels signes, peut changer totalement le sens des mots.
- 7- Un seul dialecte (par exemple le dialecte algérien) peut contenir plusieurs sous dialectes [28].

3.5 Etat de l'art (Travaux Connexes) :

Dans cette partie nous allons citer quelques travaux réalisés dans le domaine de la classification de sentiments en arabe. Beaucoup de travaux ont été réalisés dans la classification de sentiments en tant que positif et négatif avec différentes approches.

Ces dernières années, les chercheurs ont relevé le défi d'analyser les sentiments et de détecter les opinions dans le langage dit « d'enrichissement morphologique : MRL ». En fait, le langage MRL est un langage qui exprime des informations importantes sur les unités syntaxiques et les relations au niveau des mots [R. Ayadi, M. Maraoui et M. Zrigui, «Latent Topic Model for Indexing Arabic Documents,» *International Journal of Information Retrieval Research*, vol. 4, pp. 29-45, 2014.].

L'arabe est l'une des langues qui a commencé à susciter un certain intérêt. Parmi les travaux qui s'y intéressent, nous avons cité M. Elhawary et M. Elfeky [M. Elhawary et M. Elfeky, «Mining Arabic Business Reviews,» chez *IEEE International Conference on Data Mining Workshops*, Mountain View, CA, USA, 2010.], qui ont appliqué l'analyse des sentiments aux critiques arabes et utilisé des dictionnaires de mots arabes pour extraire des descripteurs permettant d'identifier la polarité (positive, négative ou neutre) du commentaire.

Mortaz K. Saad et W. Ashour [M. Khaled Saad et W. Ashour, «Arabic Text Classification Using Decision Trees,» pp. 25-26, Novembre 2010.] ont étudié l'influence de l'étape de prétraitement du texte, du radical ou du désuffixe (extraction du radical), de la normalisation du texte et du poids des termes sur la classification des textes arabes.

M. Hijjawi et Z. Bander [M. Hijjawi et Z. Bandar, «An Arabic Stemming Approach using Machine Learning with Arabic Dialogue System,» *ICGST AITML-11 Conference*, 12-14 April 2011.] représentent une méthode d'identification des opinions basée sur l'exploration d'ontologies de texte. Le dictionnaire des sentiments et les techniques de classification supervisée sont utilisés pour extraire les opinions et détecter la polarité des opinions.

A. Ziani, Y. Tlili Guiassa et N. Azizi [A. Ziani, Y. Tlili Guiassa et N. Azizi, «Détection de polarité d'opinion dans les forums en langues arabe par fusion de plusieurs SVMs,» *TALN-RÉCITAL*, 17-21 Juin 2013.] ont proposé un système qui fonctionne en trois étapes : la première étape consiste à construire et prétraiter manuellement un corpus collecté dans les journaux arabes algériens. La deuxième étape consiste à sélectionner les caractéristiques qui caractérisent le commentaire. Enfin, la troisième étape consiste à implémenter le module de classification du classifieur SVM combinant quatre fonctions du noyau différentes.

3.6 Conclusion :

Dans ce chapitre, nous avons introduit un certain nombre de techniques visant à analyser le sentiment public à partir des big social data. La classification des sentiments peut être obtenue par apprentissage automatique ou par des méthodologies lexicales ou une combinaison des deux. Cependant, nous avons vu que la plupart de ces techniques souffrent encore de quelques lacunes. En effet, les contributions précédentes ont utilisé des dictionnaires génériques qui ne tiennent pas compte du contexte étudié, ou ont fait appel aux méthodes d'apprentissage automatique qui utilisent des données d'apprentissage pour un domaine spécifique. Cependant, dans les deux cas, les méthodes énoncées sont peu efficaces.

Nous avons introduit aussi la classification de sentiments en arabe et nous avons cité les approches de classification utilisée pour cette langue.

Dans le chapitre suivant, nous allons voir comment on a fait le processus de la collection du corpus ainsi que les difficultés retrouvées et comment on a réussi à réaliser un système de classification de sentiments en arabe.

Chapitre 4 : Conception et Réalisation

4.1 Introduction

Ce chapitre sera consacré aux étapes fondamentales de réalisation de notre système d'analyse de sentiments en arabe.

Cette partie sera affectée à la collection de corpus qui est l'étape la plus importante dans notre travail, qui sera ensuite un fichier de test. Nous présentons aussi les outils de développement.

L'objectif de ce travail est de faire une étude comparative entre les algorithmes classification (KNN, NB, DecisionTree) selon leurs exactitudes.

Partie I : Construction du corpus

Le développement d'un système d'extraction d'émotions nécessite de rassembler un nombre suffisant de commentaires qui serviront au corpus d'analyse et test.

Pour cela on a réussi à collecter ce corpus dans la manière suivante :

4. I.1 Collection des données

L'extraction d'opinions et l'analyse de sentiments découlent de la nécessité de recueillir l'opinion publique. Ce processus prenait généralement plus de temps et constituait un travail de longue haleine. Le corpus se compose des commentaires en arabe qui traitent des sujets d'actualité (politique, sports, cinéma, émissions de télévision, artistes, etc.) collectées entre 01/03/2021 et le 31/05/2021 pour avoir le maximum de commentaires possibles. La plupart des données ont été collectées manuellement à partir du réseau social twitter ainsi que d'autres ont été téléchargés à partir de la plateforme facepager et le site ExportComments.. Les données ont été sélectionnés selon le présence de mots-clés (حزنفرح,اشمنزاز).

Une autre partie du corpus a été exploitée du corpus de plusieurs travaux communes le tableau 4.I.1 montre les travaux trouvés :

Reference	Nombre de commentaires	MSA ouDialecte	Polarité
SemEval:2017:task4, author = {Sara Rosenthal and NouraFarra and	672 commentaires	Dialecteégyptien	Positive, négative, neutre

Preslav Nakov}, title = {{SemEval}-2017 Task 4: Sentiment Analysis in {T}witter}, booktitle = {Proceedings of the 11th International Workshop on Semantic Evaluation}, \$			
<u>mahmoudnabil/ASTD</u> arabic sentiment tweet dataset	10 000 tweets	Different dialects	Positif, subjectif négatif, subjectif mixte et objectif
Arabic-sentiment-analysis [motazaad April 2019]	47 000 tweets	Different dialects	Positive, negative
A. Altowayan and L. Tao "Word Embeddings for Arabic Sentiment Analysis", IEEE BigData 2016 Workshop	3543 tweets	Different dialects	Positive, negative
DE-CNN Arabic Sentiment Classification Using Convolutional Neural Network and Differential Evolution Algorithm	Près de 8000 commentaries	Différents dialecte Khalji et égyptien	Positive, negative, neutre
Emotion analysis in DSI 2020	Près de 2000 commentaires	Dialect khaliji	Colere, degout, peur, joie

Tableau4.I.1 : Analyse des travaux antérieurs sur l'analyse des sentiments en arabe.

Le langage utilisé était un mélange entre l'arabe standard et quelques dialectes comme le dialectes algérien, égyptien et khaliji.

Le domaine ou le sujet approuvé :

Le corpus s'agit d'un ensemble de commentaires de différents sujets et thèmes. . Les données sont sauvegardées dans un fichier CSV pour les classer et les donner comme entrée à l'algorithme d'entraînement.

Cet ensemble de données compris 1216 commentaires Base de données textuelle .

Nous avons classé toutes les données en 6 classes selon les 6 émotions de base :

Emotion	فرح	خوف	اشمزاز	مفاجأة	حزن	غضب
Nombre de commentaires	209	200	196	189	219	203

Tableau4.I.2 : Chaque émotion avec le nombre de commentaires filtrés.

Et voici quelques exemples de chaque classe d'émotion

Commentaire	Emotion
مساء النور والله صورة الخلفية جد رائعة مع هذا المنظر الجميل والبحر	فرح
ما اخاف البعد انا خايف عليك وين بتلقى لي بهالدنيا مثيل	خوف
يارب عوضني فرحاً عن كل حزن دخل قلبي	حزن
!!أهلها عم يتعاملوا مع اللي صار كأنه وفاة عادية مفاجئة	مفاجأة
بعض الأفعال تدرج تحت تصنيف "مُقرز" من ضمنها إن حد يمشى بكلب في الشارع	اشمزاز
هذا المستخدم غاضب بشدة حتى غسل المواعين اللي كان يمتصه راح تأثيره	غضب

Tableau4.I.3 : Quelques commentaires et leur émotion.

4. I.2 Le dictionnaire d'émotions :

Ce lexique est une extension du lexique des émotions WordNet-Affect (WNA) [Strapparava et Mihalcea, 2007][25], qui est un sous-ensemble du WordNet anglais. Chaque entrée (synset) de ce lexique est annotée avec une des six émotions (colère, dégoût, peur, joie, tristesse et surprise), qui sont considérées comme les émotions humaines de base selon l'étude psychologique menée dans [Ekman, 1992][26].

Le tableau ci-dessous décrit les lexiques arabes, où émotion est le nombre des mots synonymes qui sont regroupés en ensembles et Nombre est le nombre de mots associés à chaque émotion.

Nombre	Émotion
1156	فرح
425	حزن
155	اشمئزاز
748	غضب
201	مفاجأة
3207	Total

Tableau4.I.4 : Lexique arabe et le nombre de mot dans chaque émotion.

Mentionnons aussi quelques exemples de chaque lexique

Lexique d'émotion	Exemple
فرح	إيجابي. ابتهاج، حب. بهجة. سعادة. نشوة
حزن	دمع. محبط. تعاسة. كآبة. حزين. دموع
مفاجأة	تعجب. روعة. دهشة. ذهول. عجيب. رائع
اشمئزاز	إشمئزاز. متقزز. قذر. فساد. ملعون
خوف	هلع. رعب. خوف. ذعر. قلق. تهديد
غضب	غيظ. إزعاج. جريمة. غل. انتقام. كره

Tableau4.I.5 : Lexique d'émotions.

4. I.3 Difficultés retrouvées lors de la collection

- La collection des données était la phase la plus durable dans notre recherche. Toutes les recherches retrouvées s'occupaient de la classification en positif, négatif, ou neutre, ce qui nous a poussés à classer chaque commentaire manuellement
- Les émojis changeaient l'émotion du commentaire par exemple lorsqu'un commentaire contient des mots tristes et un emoji de joie la détection de l'émotion du commentaire et la classification devient moins exacte et plus difficile

- La langue arabe comme nous avons mentionné précédemment est une langue hautement flexionnelle et dérivationnelle avec de nombreuses formes de mots et de diacritiques. Plusieurs suffixes, axes et préfixes dans les mots arabes rendent plus difficile pour les analyseurs lexicaux ou morphologiques d'extraire la racine des mots correctement

4.1.4 Prétraitements :

Dans cette phase nous avons supprimé tout le bruit manuellement. La suppression contenait :

- Les emojis (😊 😞 <3 :(</3)
- La punctuation (/,.?!&’#{[|^@]%*()}).

4. I.5 Méthodologie suivie :

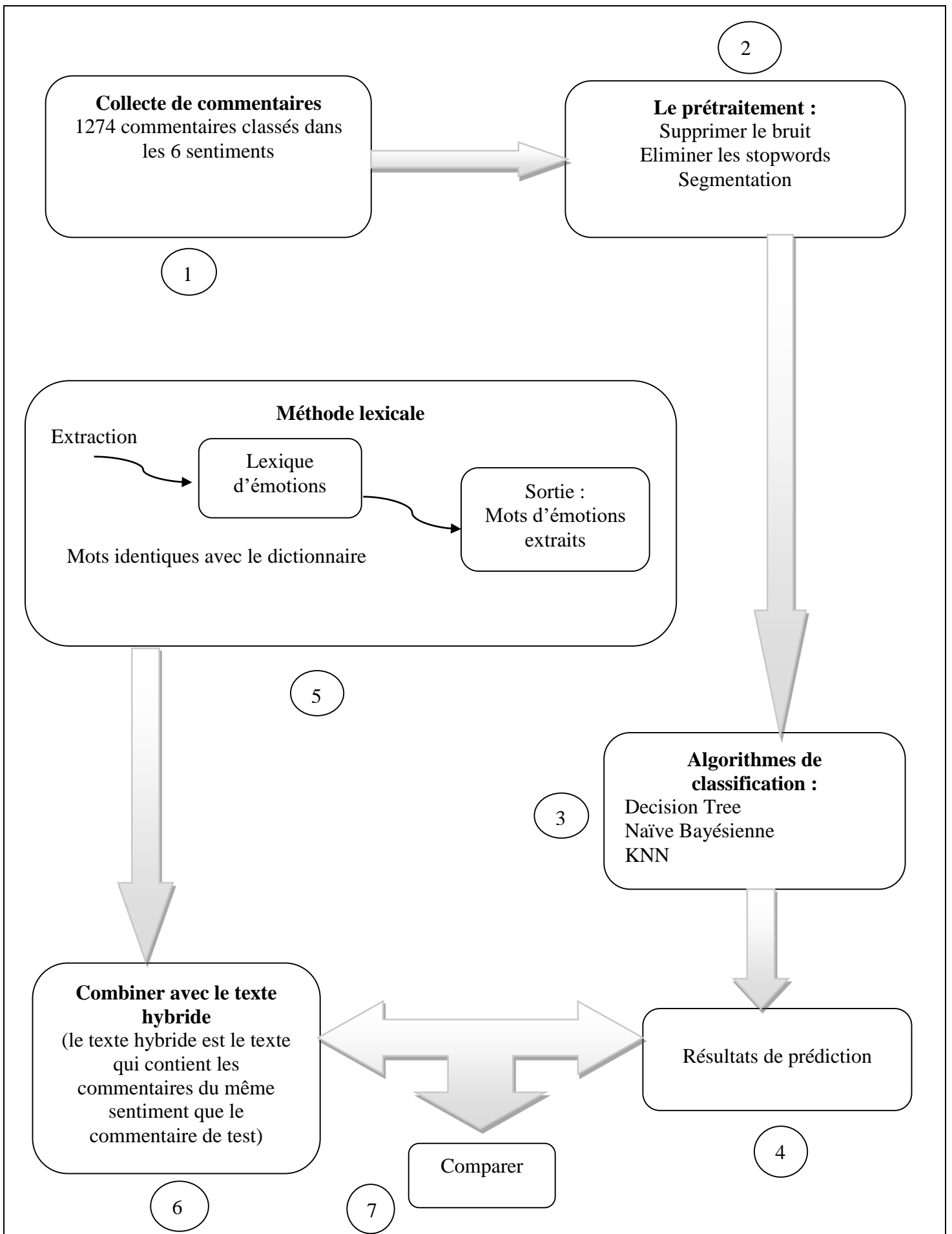


Figure4.I.1 : Méthodologie suivie.

Etape 1 le processus a commencé par collecter les commentaires et les classer dans les 6 sentiments selon le dictionnaire de lexicon.

Etape 2 prétraitement : le prétraitement faisait manuellement en éliminant le bruit et les stopwords.

Etape 3 appliqué le corpus sur des algorithmes de classification.

Etape 4 résultats de prédiction : cette étape affiche les résultats de la classification après l'implémentation de chaque algorithme.

Etape 5 méthode lexicale cette phase doit extraire un sentiment d'après un nouveau commentaire

Etape 6 combiné avec texte hybride dans cette étape le mot du sentiment détecté doit être combiné avec le fichier de test.

Etape 7 comparaison des résultats dans le cas d'utilisation de la méthode lexicale et les algorithmes de classification.

4. I.6 Algorithme lexical

Règle générale :

On a utilisé la fonction `getWords ()` pour sélectionner mot par mot et chercher à quel dictionnaire de sentiments correspond à chaque mot. Après si le mot existe dans notre dictionnaire on utilise une autre fois la fonction `getWords ()` pour garder le mot dans le fichier de test correspond au fichier de sentiment la figure suivante présente brièvement.

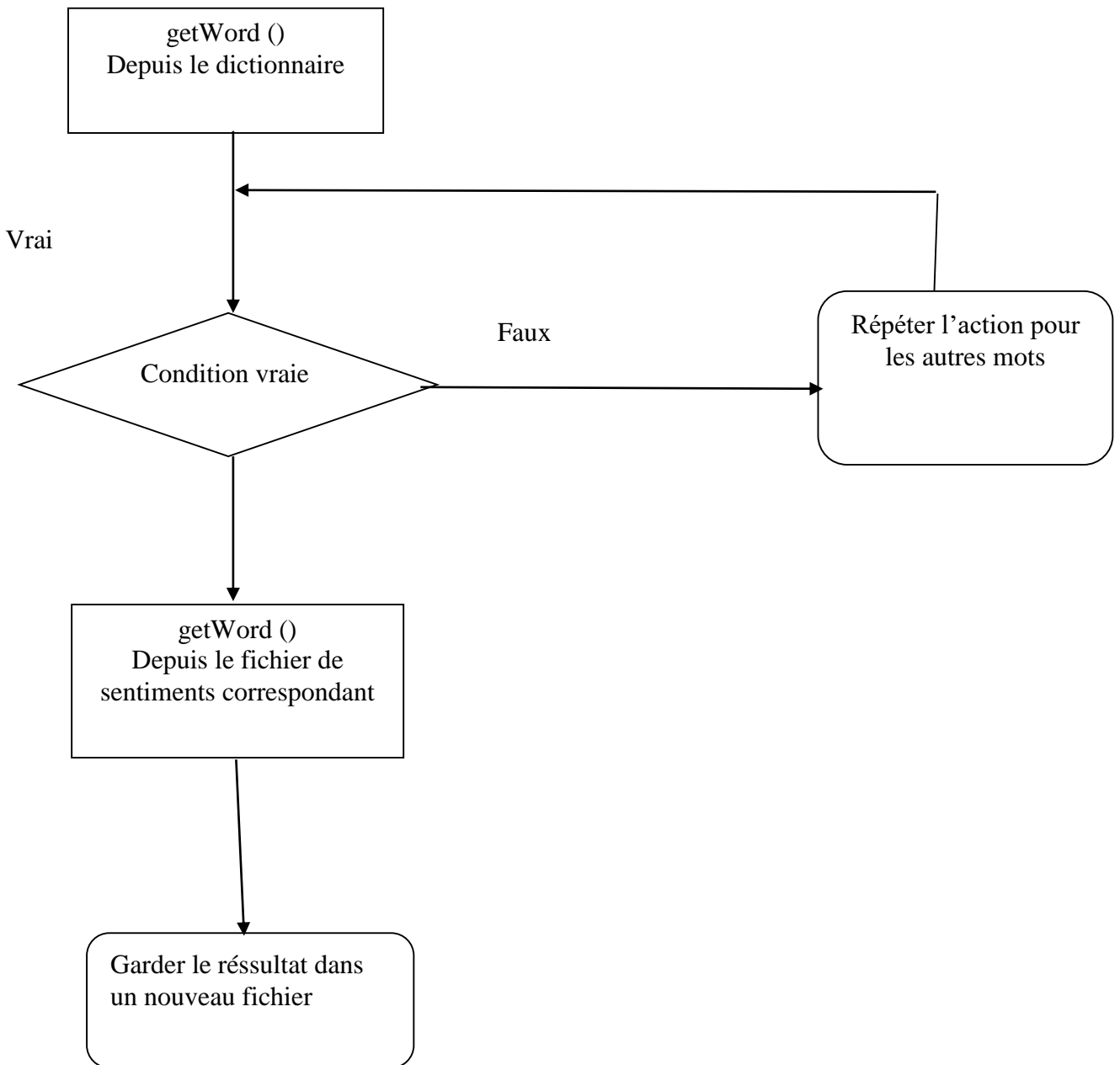


Figure4.I.2 : Schéma de la règle générale de getWords ().

On a appliqué cette fonction dans l'exemple suivant :

أشمنزاز تام من جميع البشر كويس بأني مو منهم

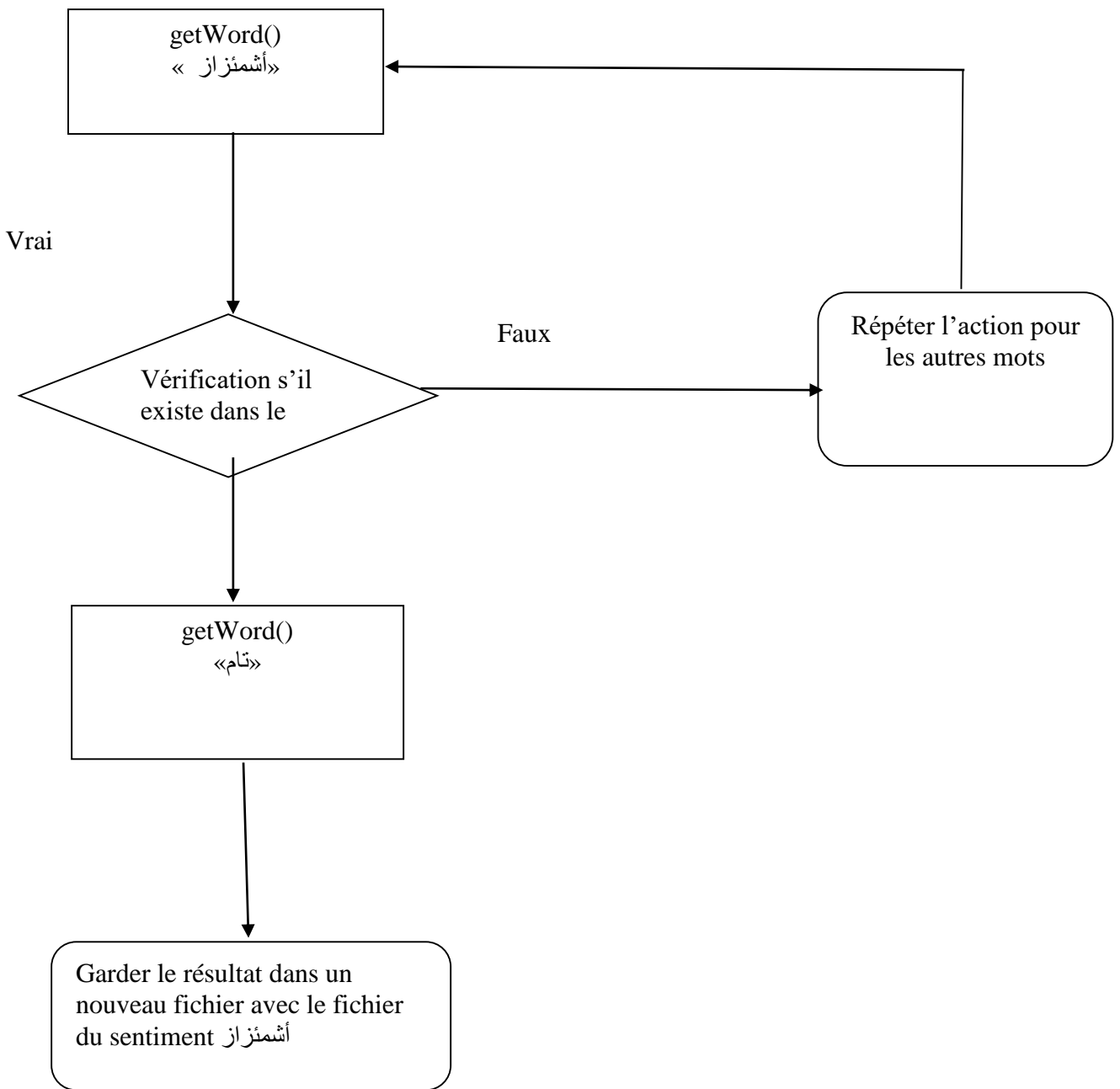


Figure4.I.3 : Exemple pour extraire un sentiment d'un commentaire.

Partie II : Réalisation

4. II.1 Outils de développement

4. II.1.2 Langage JAVA :

Notre choix de langage de programmation se porte sur le langage JAVA, car c'est un langage simple orienté objet qui réduit les risques d'incohérence, et il dispose d'une riche bibliothèque de classes, comprenant diverses fonctions, telles que des fonctions standards, des fichiers Le système de gestion est aussi ainsi que de nombreuses fonctions qui peuvent être utilisées pour développer diverses applications. Il existe un grand nombre de bibliothèques développées et fournies pour JAVA. L'API (Application Programming Interface) des langages autres que JAVA n'est pas encore finalisée et doit encore être mise à jour.

4. II.2 Environnement de développement

L'environnement de développement utilisé ici est IntelliJ IDEA 2021.1

4. II.2.1 IntelliJ IDEA 2021.1:

IntelliJ IDEA est un environnement de développement intégré (IDE) pour les langages JVM conçu pour maximiser la productivité des développeurs. Il effectue les tâches routinières et répétitives à votre place en fournissant une finalisation de code intelligente, une analyse de code statique et des remaniements, et vous permet de vous concentrer sur le bon côté du développement logiciel, le rendant non seulement productif mais aussi une expérience agréable.²

² https://en.wikipedia.org/wiki/IntelliJ_IDEA



Figure4.II.1 : Interface de IntelliJ IDEA Community 2021.1.

4. II.2.2 Fonctionnalités de IntelliJ :

L'outil fournit Plusieurs fonctions et plugins qui permettent l'analyse du code source.

- Il permet Produire des indicateurs de qualité du code source intéressants.
- Il permet notamment d'analyser plusieurs aspects pouvant être classés comme mauvais code Pratique de programmation (c'est-à-dire "odeur de code").
- Il permet également l'analyse Dépendance entre les composants logiciels et analyse des flux de données, y compris L'objectif est de mieux comprendre le fonctionnement des parties de code complexes. [Mesures des sous-caractéristiques de la maintenabilité en génie logiciel Rapport Technique Présenté à l'École de Technologie Supérieure dans le cadre du cours MGL804 - Réalisation et Maintenance de Logiciels Été 2018 Prof. Alain April Par, ELWARTITI, Alkhalil Code permanent: ELWA19129201 Montréal, le 16 juillet 2018].

4. II.2.3 Rapidminer9.9 :

RAPIDMINER est un logiciel libre et open source du projet Yale de l'équipe d'intelligence artificielle de l'Université de Dortmund. Le projet a été repris par Rapid-I, qui maintient deux versions en même temps

- Une version distribuée gratuitement « Community Edition » ;
- Une version commerciale « Enterprise Edition ».

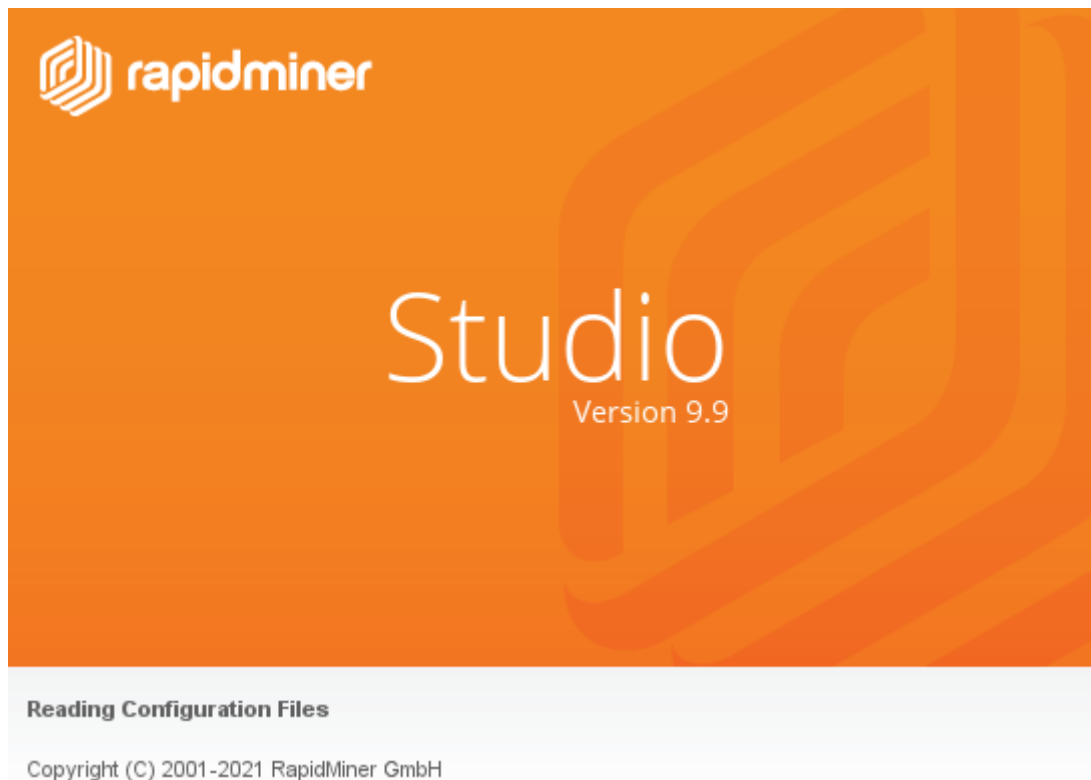


Figure4.II.2 : L'interface de RAPIDMINER9.9.

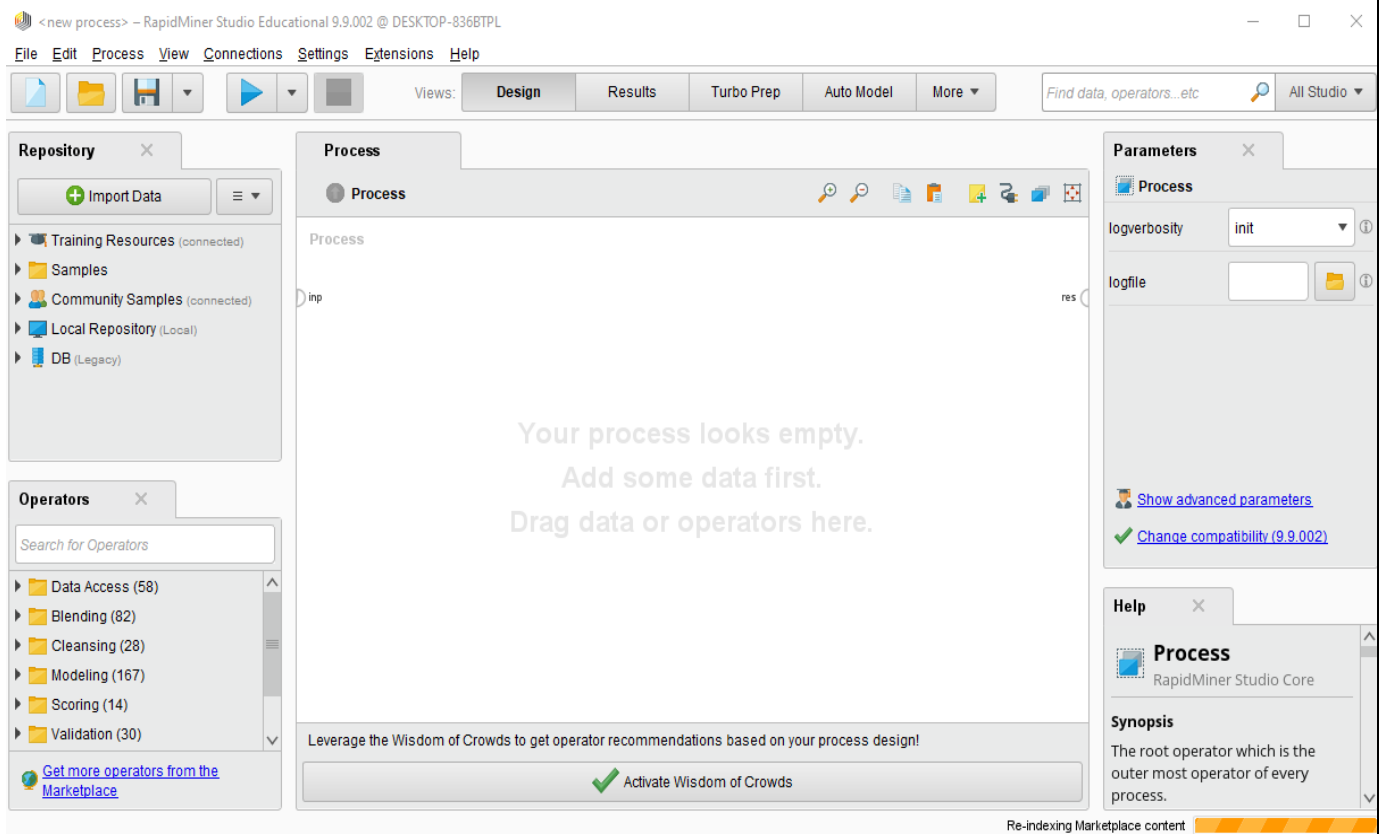


Figure 4.II.3: Ecran principal design RAPIDMINER.

Sur cette interface il existe les blocs suivants : [<https://docs.rapidminer.com/latest/studio/getting-started/ui-overview.html>]

1. Operators : Blocs de construction utilisés pour créer des processus RapidMiner.

2. Repository : Stockage dans RapidMiner Studio pour les données et les processus RapidMiner.
3. Process Panel : Zone de travail pour les processus de construction.
4. Views : Zone de travail permettant d'accéder à une fonctionnalité spécifique.
5. Ports : Mécanismes d'entrée et de sortie pour les opérateurs et les processus. Voir la liste des abréviations et des significations des ports.
6. parameters Paramètres qui modifient le comportement de l'opérateur.
7. Help : Aide contextuelle pour l'opérateur sélectionné.

4. II.3 Exemple de la méthode réalisé :

La figure4.II.4 : présente un morceau de code qui permet d'enregistrer les données des fichiers dans un tableau dynamique en utilisant la fonction `getWords ()`.

```
public class main {
    private static final String WORD_FILE="wordsFile.txt";
    private static final String DICTIONARY="dictionary.txt";
    private static final String RESULT="result.txt";

    public static void main(String[] args) {
        ArrayList<String> wordFile =new ArrayList<>();
        ArrayList<String> wordDictionary =new ArrayList<>();
        ArrayList<String> result =new ArrayList<>();
        wordFile = getWords(WORD_FILE);
        wordDictionary = getWords(DICTIONARY);
    }
}
```

Figure 4.II.4 : Utilisation de la fonction `getWords ()`.

La figure 4.II.5 les instructions nécessaires pour comparer le fichier des commentaires et le dictionnaire et mettre les mots d'emoions identiques dans le fichier de résultats.

```

for (String word : wordFile ){
    for (String d :wordDictionary){
        if (word.equals(d)){
            result.add(word);
            break;
        }
    }
}
System.out.println("length "+ result.size());
for (String r : result){
    System.out.println(r);
}

```

Figure 4.II.5 : Comparaison du fichier de commentaire et le dictionnaire.

La figure 4.II.6 montre les instructions nécessaires pour mettre le contenu du fichier des commentaires dans le nouveau fichier RESULT.

```

try {
    FileWriter myWriter = new FileWriter(RESULT);

    //
    try {
        File myObj = new File(WORD_FILE);
        Scanner myReader = new Scanner(myObj);
        while (myReader.hasNextLine()) {
            String line = myReader.nextLine().trim();
            myWriter.write(line);
            myWriter.write( str: "\n");
        }
        myReader.close();
    } catch (FileNotFoundException e) {
        System.out.println("An error occurred.");
        e.printStackTrace();
    }
    myWriter.write( str: "-----");
}

```

Figure 4.II.6 : Mettre le contenu du fichier de commentaires dans le nouveau fichier.

La figure 4.II.7 montre les instructions de l'ajout des résultats dans le nouveau fichier formant une concaténation entre le fichier de commentaires et les résultats obtenus.

```
for(String str: result ) {  
    myWriter.write( str: str + System.lineSeparator());  
}  
myWriter.close();  
} catch (IOException e) {  
    System.out.println("An error occurred.");  
    e.printStackTrace();  
}  
}
```

Figure 4.II.7 : Concaténation du nouveau fichier avec les résultats.

Les résultats sont dans la figure 4.II.8 suivante :

out of bidirectional text can depend on the base direction (View | Bidi Text Base Directio. Choose direction Hide notification Don't show again

9

و وزيركم للخارجية يخاطب الامم بالفرنسية ، لغة المستعمر غضب
زعيم عربي ، نعم زعيم عربي و نصف لكن في المكر و الخداع و النفاق و المتاجرة في القضايا العربية غضب
حتى تنفرض الوحوش التي تحكمن غضب
الانترنات نتاعكم رأها ضعيفة جدا اعطونا سبب مقنع غضب
برشلونه بدون ميسي ، كالارض بدون مطر غضب
اي سلفية تتكلم عنها و أبرياء يموتون في اليمن ويحك غضب
يلزم نتعلم كل واحد يعطي رايه و الرد يكون باحترام مو بالسبب غضب
الهوى علي النهار ده هو مفيش غيري فل مطبخ اه ده انا حيلي اتهد فل مطبخ الصبح غضب
نت و ايران ملة واحدة متفقون عاي التخريب غضب
يرحم امك تغيير شو مستني مات ماريانو الله لا يوفقه غضب
الله لا يوفقه اذا هذا مدرب من ملا بقولها غضب
كذا يعني غضب شديد و ماسكة نفسي باليالله غضب
كذا يعني غضب شديد و ماسكة نفسي باليالله غضب
" مزاجي عصبي متهور احب المغامرات واحب بنفس الوقت " الوحدة غضب
الأخيره تسبب غضب شديد غضب
تستفزني لا أستطيع تجاهلها أشياء سخيطة صايرة تثير غضبي بسرعة طيب إيش تغير يا ريم أنا جالسة أتعب والله غضب
سيلفي حلقة اليوم منرفزة طريقة كلام الجريم جابت لي القلق غضب

غضب
قلق
في
الخطر
في
في
رهيب

Figure 4.II.8 : Le fichier RESULT.

4. II.4 Etude de résultats :

Après avoir testé la méthode de classification sur notre corpus qui nous a permis d'ajouter le commentaire avec les mots de lexique qui appartient, Nous avons fait le test de l'analyse sur le corpus en implémentant des algorithmes de classification et en ajoutant et supprimant quelques opérateurs pour comparer l'exactitude des résultats à chaque fois.

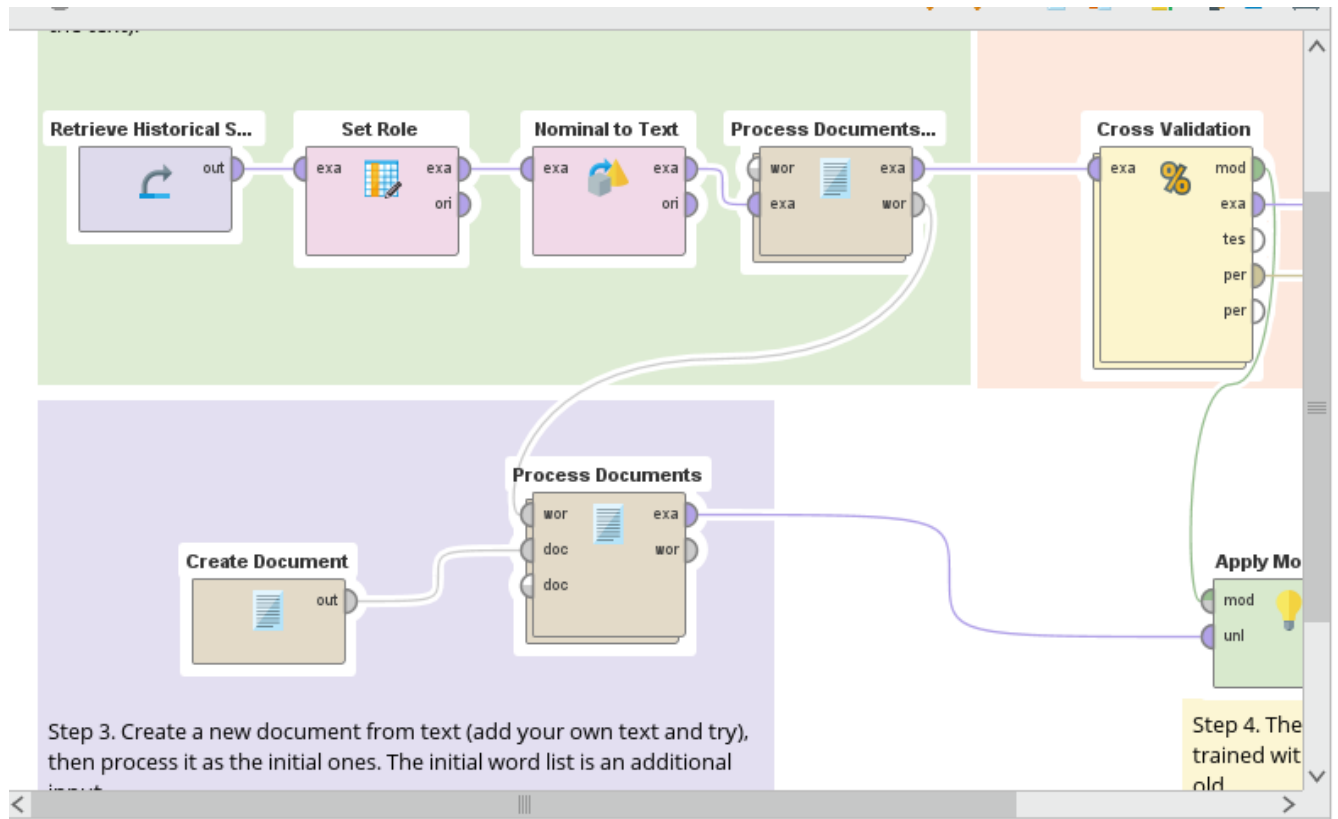


Figure 4.II.9 : Processus de l'analyse de sentiment dans RAPIDMINER.

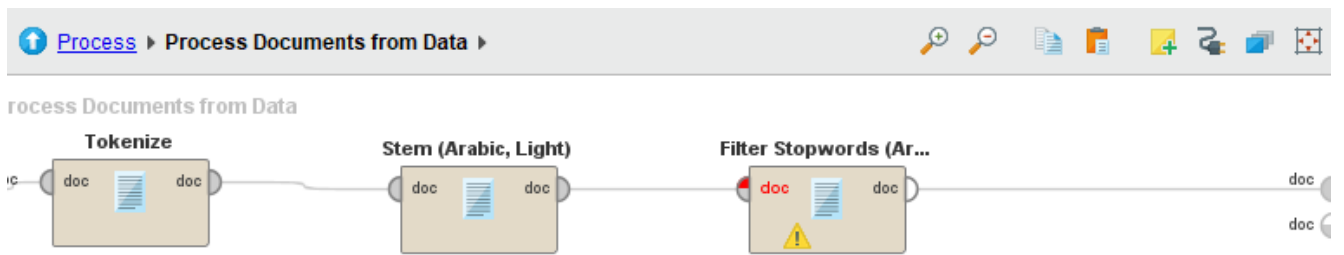


Figure 4.II.10 : Prétraitement du corpus dans RAPIDMINER.

Après avoir donné le commentaire « خايف أقول اللي فقلبي » on obtient les résultats montrés dans la figure 4.II.6

Open in Turbo Prep Auto Model Filter (1 / 1 examples): all

Row No.	prediction(e...)	confidence(ف...	confidence(ع...	confidence(ح...	confidence(م...	confidence('...	confidence(ع...	text	تحيات n
1	خوف	0	1	0	0	0	0	خوف أقول قلب	0

Figure 4.II.11 : Résultats de prédiction.

Le tableau 4.II.1 montre les changements apportés dans chaque modèle après avoir donné le commentaire utilisé précédemment comme un exemple de test et l'appliquer avec les algorithmes de classification et comparer les résultats de la performance de chaque algorithme

Modèle/opérateurs	Tokenize	Stem(arabic,light)	Stem(arabic)	FilterStopWords
Mod1	✓		✓	✓
Mod2	✓	✓		✓
Mod3	✓		✓	

Tableau 4.II.1 : Les changements apportés dans chaque modèle.

En appliquant les algorithmes de classification sur chaque modèle

Algorithme / Modèle	Naïve Bayesian	KNN	DecisionTree
Modèle1	56.88%	59.08%	48.52%
Modèle2	64.74%	63.90%	35.00%
Modèle3	58.15%	57.64%	46.57%

Tableau 4.II.2 : Résultats de performance après l'application de chaque modèle.

La figure 4.II.9 montre un exemple de la performance de l'approche KNN

accuracy: 59.08% +/- 3.60% (micro average: 59.09%)

	true فرح	true خوب	true حزن	true مفاجئة	true اشمزاز	true محصب	class precision
pred. فرح	103	6	25	13	13	11	60.23%
pred. خوب	44	167	29	46	22	37	48.41%
pred. حزن	27	19	128	35	27	40	46.38%
pred. مفاجئة	11	1	4	77	2	8	74.76%
pred. اشمزاز	13	6	7	10	130	11	73.45%
pred. محصب	4	0	5	7	1	94	84.68%
class recall	50.99%	83.92%	64.65%	40.96%	66.67%	46.77%	

Figure 4.II.12 : Performance de l'approche KNN.

Selon les résultats obtenus, nous concluons que la performance de l'algorithme Naïve Bayesian et KNN sont les meilleurs approches que DecisionTree, car le pourcentage de l'exactitude est plus élevé dans ces approches.

Le processus de l'analyse de sentiments utilisé peut être amélioré par ajouter d'autres opérateurs qui aident à arriver au résultat voulu avec une performance plus élevé.

On a finalement fait le test sur le fichier RESULT qui concatène l'ancien corpus avec les mots d'opinion de chaque commentaire dans le système de l'analyse de sentiments de RAPIDMINER, en appliquant les approches de classification sur chaque modèle.

Les résultats de performances sont les suivants:

Algorithmes / Modèle	Naïve Bayesian	KNN	DecisionTree
Modèle1	68.47%	41.12%	42.88%
Modèle2	73.65%	38.47%	42.47%
Modèle3	69.59%	38.82%	42.24%

Tableau 4.II.2: Résultats de performance de nouveau corpus.

D'après les résultats obtenus, nous pouvons conclure que :

- La performance de l'approche Naive Bayesian est meilleure que les autres algorithmes kNN et DecisionTree ;
- De plus, le meilleur modèle de commentaire est le modèle 2 qui consiste à utiliser le stemmer léger, et dont il essaye de garder le sens du mot (sentiment) ;
- La technique hybride qui ajoute les mots de sentiments par la méthode lexicale améliore les résultats dans un seul algorithme ; Naive Bayesian ;

Les résultats peuvent être améliorés en appliquant d'autres opérateurs qui augmentent la performance des autres approches.

4.4 Conclusion

Dans cette dernière partie nous avons présenté la conception et la réalisation de notre travail. Commencant par la collection du corpus jusqu'aux résultats de classification avec une méthode créée et les résultats en utilisant RAPIDMINER .on a aussi exploré les outils utilisés ainsi que les environnements de réalisation.

5. Conclusion générale :

Le domaine de l'analyse de sentiments présente un axe de recherche qui permet de faciliter et améliorer la vie quotidienne en analysant les sentiments et les opinions des internautes. Malgré l'existence de plusieurs travaux dans ce domaine mais ils étaient tous limités et notre réalisation a pu lever :

- Pratiquement tous les travaux calculaient la polarité (positif, négatif, neutre)
- Le langage utilisé était soit l'arabe ou un spécifique dialecte
- Les résultats étaient en pourcentage ou en représentation graphique
- Au terme de ce projet nous pouvons confirmer que nous avons réussi à réaliser notre but fixé et répondre aux objectifs

Ce travail nous a permis d'acquérir nos connaissances dans le langage JAVA, et d'améliorer nos connaissances dans certaines plateformes de DATAMINIG

Pour les perspectives de ce travail nous proposons les points suivants pour l'amélioration de cette étude :

- Un système d'extraction multilingue
- Une interface pour afficher les résultats aux utilisateurs

Références bibliographiques :

- [1] Abdelkrim, B. E. G. H. D. A. D., & Amina, O. U. S. E. R. I. R. (2018). Une approche Deep Learning pour l'analyse des Sentiments Sur Twitter..
- [2] <https://ar.wikipedia.org/> consulté le 5/5/2021.
- [3] Weber, M., & Claudon, P. (2009). L'émotion. Contribution à l'étude psychodynamique du développement de la pensée de l'enfant sans langage en interaction. *Devenir*, 21(1), 61-100.
- [4] <https://apprendreaeducer.fr/roue-des-emotions-enfants/> consulté le 16/04/2021.
- [5] B.Bathelot. (2016, 01 27). *L'encyclopédie illustrée du marketing*. Consulté le 05 26, 2018, sur définition marketing: <https://www.definitions-marketing.com/definition/reseaux-sociaux/>].
- [6] *Historique des réseaux sociaux*. (s.d.). Consulté le 05 26, 2018, sur cadre21: <https://www.cadre21.org/ressources/historique-des-reseaux-sociaux/>.
- [7] Fraisier, O., Cabanac, G., Pitarch, Y., Besancon, R., & Boughanem, M. (2017). Expression politique sur Twitter: les communautés comme indicateurs de point de vue.
- [8] Girardot, M. J. J. (2008). *Grzegorz DZICZKOWSKI* (Doctoral dissertation, École Nationale Supérieure des Mines de Paris).
- [9] Polanco, X. (2001, January). Textmining et intelligence économique: aujourd'hui et demain. In *Colloque Veille technologique, Intelligence économique et Bibliométrie*.
- [10] Toussaint, Y. (2011). *Fouille de textes: des méthodes symboliques pour la construction d'ontologies et l'annotation sémantique guidée par les connaissances* (Doctoral dissertation, Université Henri Poincaré-Nancy I).
- [11] <https://touriaelouahabi.wordpress.com/text-mining/definition-du-text-mining/> consulté le 12/05/2021
- [12] Grivel, L. (2007). Maîtriser le processus de textmining dans le cadre d'applications d'intelligence économique, de gestion de la relation client ou de gestion de connaissances. *Présenté à Veille Stratégique Scientifique et Technologique (VSST), Marrakech (Maroc)*.
- [13] Alowaidi, S., Saleh, M., & Abulnaja, O. (2017). Semantic sentiment analysis of Arabic texts. *International Journal of Advanced Computer Science and Applications*, 8(2), 256-262.
- [14] BENKHELIFA, R. *Modélisation sémantique des intérêts des internautes dans les réseaux sociaux* (Doctoral dissertation, Université de Ouargla-KasdiMerbah).
- [15] Mekki, A. B. I. R. (2018). *Détection automatique des sentiments Dans les réseaux sociaux* (Doctoral dissertation, UNIVERSITE MOHAMED BOUDIAF-M'SILA FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE. Filière: Informatique).

- [16] OUALI, C. (2014). *Classification automatique de textes* (Doctoral dissertation, UNIVERSITE MOHAMED BOUDIAF M'SILA: FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE: Département d'Informatique).
- [17] Chamroukhi, F. (2013). Classification supervisée: Les K-plus proches voisins. *mémoire de fin d'étude, Université du Sud Toulon-Var*.
- [18] Ouakkaf, H., & Berkane, M. (2017). Reconnaissance automatique des expressions faciales par support vector machine.
- [19] Bulagang, A. F., Weng, N. G., Mountstephens, J., & Teo, J. (2020). A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. *Informatics in MedicineUnlocked*, 20, 100363.
- [20] KIHHEL, C. Analyse des sentiments en utilisant l'apprentissage Profond: Cas de la langue Arabe.
- [21] Poirier, D., Fessant, F., Bothorel, C., De Neef, E. G., & Boullé, M. (2009). Approches statistique et linguistique pour la classification de textes d'opinion portant sur les films. *Revue des Nouvelles Technologies de l'Information*, Pages-147.
- [22] <https://www.sciencedirect.com/science/article/pii/S2090447914000550> consulté le 12/06/2021.
- [23] Baccouche, T. (2009). Dynamique de la langue arabe. *Synergies Tunisie1*, 17-24.
- [24] Duwairi, R. M., Marji, R., Sha'ban, N., & Rushaidat, S. (2014, April). Sentiment analysis in arabic tweets. In *2014 5th international conference on information and communication systems (ICICS)* (pp. 1-6). IEEE.
- [25] Assiri, A., Emam, A., & Aldossari, H. (2015). Arabic sentiment analysis: a survey. *International Journal of Advanced Computer Science and Applications*, 6(12), 75-85.
- [26] Alhumoud, S. O., Altuwaijri, M. I., Albuhaire, T. M., & Alohaideb, W. M. (2015). Survey on arabic sentiment analysis in twitter. *International Science Index*, 9(1), 364-368.
- [27] Shoukry, A., & Rafea, A. (2012, May). Sentence-level Arabic sentiment analysis. In *2012 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 546-550). IEEE.
- [28] Alhumoud, S. O., Altuwaijri, M. I., Albuhaire, T. M., & Alohaideb, W. M. (2015). Survey on arabic sentiment analysis in twitter. *International Science Index*, 9(1), 364-368.
- [29] Strapparava and Mihalcea, 2007] Strapparava, C. and Mihalcea, R. (2007). Semeval-2007 task 14: affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07*, pages 70–74, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [30] Ekman, 1992] Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200
- [31] https://www.lattice.cnrs.fr/sites/itellier/poly_info_ling/linguistique009.html consulte le 03/03/2021.

Résumé

Dans ce travail, nous abordons le problème de la classification d'émotions (bonheur, tristesse, colère...) en langue arabe. Le but est d'étudier le rôle de l'étape de prétraitement des commentaires dans les résultats de classification. Les algorithmes d'apprentissage automatique Naive bayes, le plus proche voisin et arbre de décision sont utilisés pour la prédiction d'opinion. En plus, une méthode basée sur le lexique est proposé.

Enfin, les techniques sont combinées pour améliorer les résultats de classification. Les résultats ont montré que Naive bayes est le meilleur algorithme.

Mots clés : classification d'émotions, analyse de sentiments, langue arabe.

Abstract

In this work, we address the problem of classification of emotions (happiness, sadness, anger...) in Arabic. The goal is to study the role of the comment preprocessing step in the classification results. Naive Bayes, nearest neighbor and decision tree machine learning algorithms are used for opinion prediction. In addition, a lexicon-based method is proposed.

Finally, the techniques are combined to improve the classification results. The results showed that Naive Bayes is the best algorithm.

Keywords: emotion classification, sentiment analysis, Arabic language.

ملخص

في هذا العمل نتناول مشكلة تصنيف العواطف (السعادة ، الحزن ، الغضب ...) باللغة العربية. الهدف هو دراسة دور عملية المعالجة المسبقة للتعليق في نتائج التصنيف. يتم استخدام خوارزميات التعلم الآلي Naive bayes, nearest neighbour و decision tree للتنبؤ بالرأي.

بالإضافة إلى ذلك ، تم اقتراح طريقة قائمة على المعجم.

أخيرًا ، يتم الدمج بين التقنيات لتحسين نتائج التصنيف. أظهرت النتائج أن Naive bayes هي أفضل خوارزمية. **كلمات مفتاحية :** تصنيف العواطف، تحليل المشاعر، اللغة العربية.