

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE Des mathématiques et de l'informatique
DEPARTEMENT D'INFORMATIQUE
N° :.....



DOMAINE : Mathématique et informatique
FILIERE : Informatique
OPTION : SIGL

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par : Mekki ABIR

Intitulé

**Détection automatique des sentiments
Dans les réseaux sociaux**

Soutenu devant le jury composé de :

<i>Dr</i> Kadri Said	Université de M'sila	Président
<i>M^{me}</i> Ahlem Bouzaroura	Université de M'sila	Rapporteur
<i>M</i> Ariouat Youcef	Université de M'sila	Examineur

Année universitaire : 2017 /2018

Dédicace

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

A ma chère sœur Amal pour leurs encouragements permanents, et leur soutien moral,

A mes chers frères, Ala Eddine, Djamel Eddine, El-moàtassim Billah, pour leur appui et leur encouragement,

A toute ma famille pour leur soutien tout au long de mon parcours universitaire,

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible,

Merci d'être toujours là pour moi.

Remerciement

Nous tenons tout d'abord à remercier Allah, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

En second lieu, nous tenons à remercier notre encadreur Mme: (**Ahlem Bouzaroura**), son précieux conseil et son aide durant toute la période du travail.

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

De plus, nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

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 des matières.....	i
liste des figures	iv

Introduction générale

Description de cette étude	- 1 -
Objectif de cette étude	- 1 -
Organisation du manuscrit.....	- 2 -

CHAPITRE 1

Les réseaux sociaux

1.1. Introduction.....	- 3 -
1.2. Caractéristiques principales des réseaux sociaux.....	- 3 -
2.1 Le modèle participatif.....	- 3 -
2.2 Réseau social en pratique	- 4 -
2.3 Le modèle économique des réseaux sociaux.....	- 4 -
1.3. Gains à rejoindre un réseau social	- 4 -
1.4. Les dangers liés à l'utilisation des sites de réseaux sociaux.....	- 4 -
1.4.1 Des données personnelles utilisées à des fins commerciales.....	- 5 -
1.4.2 De l'importance de bien gérer son identité.....	- 5 -
1.5 . Réseaux sociaux les plus connus	- 6 -
1.5.1 Facebook	- 6 -
1.5.2 LinkedIn	- 7 -
1.5.3 Google+	- 7 -
1.5.6 Instagram	- 8 -
1.5.7 Twitter	- 8 -
1.6. Avantages et les inconvénients Des Réseaux Sociaux.....	- 9 -
1.7. Conclusion	- 10 -

CHAPITRE 2

Analyse des sentiment et opinion mining

2.1. Introduction.....	- 12 -
2.2. L'analyse des sentiments et opinion mining.....	- 12 -
2.3 Niveaux d'analyse des sentiments.....	- 12 -

2.4 Disciplines en relation avec l'analyse de sentiments.....	- 14 -
2.4.1 fouille de texte.....	- 14 -
2.4.2 Traitement automatique du langage naturel (TALN).....	- 14 -
2.4.3 Apprentissage automatique	- 15 -
2.5 Méthodes de classification des sentiments	- 16 -
2.5.1 Approche basée sur le TALN.....	- 17 -
2.5.2 Approche basée sur un dictionnaire.....	- 18 -
2.5.3 Approche basée sur le corpus.....	- 18 -
2.5.4 l'approche basée sur apprentissage automatique.....	- 19 -
2.6 Outils d'analyse des sentiments	- 19 -
2.6.1 Werfamous	- 19 -
2.6.2 AFINN	- 20 -
2.6.3 General Inquirer	- 20 -
2.6.4. SenticNet	- 20 -
2.6.5 WordNet.....	- 21 -
2.6.6 SentiWordNet.....	- 21 -
2.6.7. Sentiment140	- 22 -
2.6.8. Tweetfeel	- 22 -
2.6.9 Twitrratr	- 22 -
2.6.10. Tweet Sentiments Analyses	- 23 -
2.7 Conclusion.....	- 23 -

CHAPITRE 3

Conception et réalisation

Partie 1

3.1. Introduction	- 24 -
3.2. Travaux réalisés dans le domaine de l'Opinion Mining.....	- 24 -
3.2.1 analyse des sentiments basée sur le lexique (2010)	- 24 -
3.2.2 Analyse des sentiments basés sur LSA (2010)	- 24 -
3.2.3 SVM multi-classes (2011).....	- 25 -
3.2.4 Analyse des sentiments basés sur le lexique, NLP (2012).....	- 25 -
3.2.5. Analyse de sentiment basée sur l'analyse Sémantique (2013)	- 26 -
3.2.6 Analyse de sentiment basé sur apprentissage automatique (2014).....	- 27 -
3.3. Notre système d'analyse du sentiment.....	- 27 -
3.4. Les étapes nécessaires pour notre système à l'analyse du sentiment.....	- 28 -

3.5.Prétraitement des tweets.....	;- 29 -
3.6. Calcul TF- IDF	;;- 29 -
3.6.1 TF.....	- 30 -
3.6.2 IDF	- 30 -

Partie 2

3.1 Présentation de Java.....	- 33 -
3.1.1 Les caractéristiques java	- 33 -
3.2 Environnement d'exécution.....	- 33 -
3.3 Présentation d'application Analysis_Sentement.....	- 34 -
3.3.1 Fonctionnement de l'application.....	- 34 -
3.3.1.1 Classe LatentSemanticAnalysis :.....	- 34 -
3.3.1.2 Classe TF-Idf	- 34 -
3.3.1.3 Classe LatentSemanticAnalysisTest	- 35 -
3.4 Les étapes de l'exécution.....	- 36 -
3.5 Résultat de l'exécution	- 38 -
3.6 Conclusion.....	- 40 -
Conclusion	
générale.....	- 41 -

liste des figures

figure 2.1 processus d'analyse des sentiments sur les avis sur les produits.....	- 13 -
Figure 2.2 méthodes de classification du sentiment.....	- 17 -
Figure 3.1 Architecture Analyses Sentiment.....	- 28 -
Figure 3.2 : les fonctions de la class TF-IDF.....	- 35 -
Figure 3.3: L'interface de l'application.....	- 36 -
Figure 3.4: Training LSA.....	- 30 -
Figure 3.5: Résultat de test 1.....	- 38 -
Figure 3.6: Résultat de test 2.....	- 39-

INTRODUCTION GENERALE

L'Internet est devenu un outil incontournable d'échange d'information tant au niveau personnel que professionnel. Le Web nous offre un monde de l'information prodigieux, il a évolué du simple ensemble de pages statiques vers des services et des applications web de plus en plus complexes. Actuellement, le web nous offre l'achat de tous les produits, la lecture de son journal préféré en ligne, la discussion sur de multiples forums ou la possibilité de s'exprimer sur les blogs.

Par conséquent, le web est pour la plupart d'entre nous c'est grand portail pour trouver les informations, réserver l'avion ou l'hôtel, acheter des produits, consulter les avis d'autres utilisateurs sur les produits qui nous intéressent, lire les commentaires avant de choisir le film à voir au cinéma, voir des propositions d'autres personnes avant de choisir les cadeaux etc. le problème maintenant n'ai pas lié à l'Internet mais plutôt à des considérations sociologiques, est que la globalisation nous envahit, spécifiquement avec l'apparition des réseaux sociaux qui sont devenu un espace où les gens expriment leurs sentiments en toute liberté. Les sociétés commerciales ont profité de ça pour exposer leur produit aux critiques pour avoir un feedback permanent.

Un nouveau domaine est né, l'analyse des sentiments et l'opinion Mining. Ce domaine est le descendant de son fameux ancêtre : fouille de données, plus connu sous le nom du Data Mining. Comme ce dernier, il vise à rechercher des informations pertinentes et des patterns dans les avis, les points de vue et les opinions des internautes.

L'analyse des sentiments peut être appliquée dans des nombreux domaines, spécifiquement le domaine du e-commerce. Mais, elle peut être difficile à implémenter , comme le langage humain est complexe à interpréter pour les systèmes d'apprentissage basés sur la machine, le traitement des sentiments s'avère plus complexe, ce qui entraîne la nécessité d'emploi d'autres domaines tels que : le traitement automatique de langage naturel, l'apprentissage automatique.

Objectif :

L'objectif de ce mémoire sera de détecter automatiquement les sentiments des internautes et leurs avis pour ou contre un produit ou un phénomène social. En développant un système d'analyse des sentiments pour classer les opinions en trois catégories : positif, négatif et neutre.

Organisation du manuscrit :

Après cette introduction générale, le reste de notre travail est structuré comme suit :

Le premier chapitre introduit les caractéristiques principales des réseaux sociaux, les dangers liés à leur utilisation , leurs avantages et inconvénients, ainsi leur influence sur le plan personnel est social.

Le deuxième chapitre est consacré pour le domaines d'analyse des sentiments et opinion Mining. On présentera les différents niveaux d'analyse des sentiments, les disciplines en relation avec l'analyse de sentiments , leurs méthodes de classification et les outils d'analyse des sentiments.

Le troisième chapitre est composé de deux parties : la première partie présente les travaux réalisés dans le domaine de l'analyse des sentiments et l'opinion Mining , Les étapes nécessaires et la démarche détaillée de notre système. La deuxième partie présente les outils de programmation et l'implémentation de notre application, présentation des interfaces et les résultats d'exécution .

Finalement, nous clôturons ce mémoire par une conclusion générale.

CHAPITRE 1

LES RÉSEAUX SOCIAUX

1.1. Introduction

Les réseaux sociaux sont des applications web ayant comme objectif de relier des amis, des connaissances ou des associés.

Les réseaux sociaux présentent des orientations plus ou moins personnelles ou professionnelles, c'est-à-dire que l'objectif des utilisateurs peut être de retrouver des amis et de partager des outils avec eux (photos, messages, commentaires, applications ludiques...) ou de tisser un réseau professionnel (rencontrer des partenaires potentiels, trouver un nouvel emploi, trouver des collaborateurs, annoncer des événements ou des activités professionnelles...). Le réseau le plus connu est évidemment Facebook.

La principale idée d'un réseau social est de retrouver des personnes que vous connaissez, qui à leur tour, vous permettront de rentrer en contact avec d'autres personnes. La communication est évidemment un élément central des réseaux sociaux qui proposent tous des outils de communication synchrones (chat ou vidéoconférence) et asynchrones (commentaires, forum). Il existe aussi des systèmes de communication indépendants tels que Skype ou Yahoo Messenger.

Ce chapitre est consacré pour présenter les principaux réseaux sociaux connues leurs caractéristiques, et leur impactes sur le plan personnel ou social.

1.2. Caractéristiques principales des réseaux sociaux

Trois éléments mêlés à des degrés divers selon les réseaux sociaux :

- le caractère communautaire
- le modèle participatif
- la personnalisation de l'information

2.1 Le modèle participatif

Les utilisateurs d'un réseau social en fournissent le contenu

- existe depuis le début d'Internet, principe même du Web (faire un site consultable depuis le monde entier)
- **”Web 2.0”** : facilite donnée aux utilisateurs de mettre en ligne (partager) du contenu : textes, images, vidéos, liens...

2.2 réseau social en pratique :

-L'utilisateur en s'inscrivant sur un réseau social (le plus souvent gratuitement) il peut:

- créer un profil qu'il va remplir en fournissant des informations personnelles (ASV, profession, centres d'intérêt, religion...)
- sélectionner les "amis" ou "contacts" à intégrer dans son réseau personnel
- L'utilisateur peut ensuite mettre en ligne du contenu (de manière privée ou publique) et accéder à celui partagé par ses "amis"

2.3 Le modèle économique des réseaux sociaux :

Quasiment tous étant gratuits pour les utilisateurs, quelles sont leurs sources de revenus ?

- l'affichage de publicité ciblée, selon vos goûts, activités, géolocalisation ou amis tels que connus du réseau social
- un pourcentage sur les ventes effectuées depuis le réseau social (jeux, musiques, films, articles...)
- l'appropriation des contenus mis en ligne par les utilisateurs
- la constitution et la revente de fichiers de données sur les utilisateurs. [1]

« *Si vous ne payez pas un service, c'est que vous n'êtes pas le consommateur, vous êtes le produit vendu* ». [2]

1.3. Gains à rejoindre un réseau social :

- Garder / (re)prendre contact avec des connaissances (amis, familles) ou des tiers (Institutions, entreprises, experts, artistes, associations, employeurs, personnes partageant mêmes centres d'intérêts...)
- Interagir et partager facilement du contenu avec eux
- S'informer de ses centres d'intérêts (personnels ou professionnels) de manière personnalisée
- Bénéficier de suggestions (achats, activités) adaptées. [1]

1.4. Les dangers liés à l'utilisation des sites de réseaux sociaux :

Les réseaux sociaux dans ce temps est devenu indispensable car il nous facilite beaucoup la communication et l'accès à l'information mais d'une autre coté ces sites là il nous pose dans des différents risques.

1.4.1 Des données personnelles utilisées à des fins commerciales :

Tout d'abord, il faut rappeler que ces informations représentent une véritable manne pour les entreprises de marketing et les publicitaires. En effet, les professionnels de marketing peuvent recueillir ainsi de précieuses informations sur les habitudes de vie, les goûts ou les préférences de millions d'internautes et ainsi mieux cibler leurs opérations de marketing en fonction du profil des internautes. Il s'agit donc de proposer des publicités personnalisées.

Et la popularité des sites communautaires ne cesse d'attirer les annonceurs car elle représente une formidable occasion de profiter de cette forte fréquentation et de tester l'impact sur les consommateurs.

1.4.2 De l'importance de bien gérer son identité :

A travers ces espaces personnels, les internautes souvent ne mesurent pas les risques encourus en éparpillant leurs informations personnelles sur ces sites.

Ils n'ont pas conscience qu'il s'agit d'informations très personnelles voire sensibles (opinions politiques ou religieuses...). Et que toutes ces informations qui font partie de la sphère privée de chaque individu vont être propulsées sur un espace public, visibles de tous (ou presque).

De plus, les nombreuses applications offertes sur les sites de socialisation permettent aux internautes de diffuser, d'échanger ou de copier toutes sortes de contenus multimédias (musique, films, textes, etc.), ce qui peut constituer un acte de contrefaçon lorsque les œuvres sont protégées par le droit d'auteur. Et quant à la diffusion de photos sur son profil, les internautes ignorent souvent que la simple diffusion de l'image d'un ami sans son consentement peut porter atteinte à l'image de cette personne.

L'usurpation d'identité. Il est très facile de créer une fausse identité à partir d'un nom, d'un e-mail et d'une photo. Rien ne prouve que la personne qui se cache derrière cette identité virtuelle est bien celle qu'elle prétend être. Le risque résidant dans le fait que l'abuseur puisse gagner la confiance du réseau et lui soutirer des informations. Et ensuite, dans le fait qu'il peut sérieusement nuire à votre réputation en se livrant à toutes sortes de méfaits sous votre identité.

Et surveiller sa réputation sur le réseau internet c'est s'éviter des problèmes dans la vie réelle. A l'heure actuelle, de plus en plus de recruteurs googlent notre nom avant de nous recevoir pour un entretien, s'il tombe sur une photo de nous désavantageuse, il ne

nous recevra probablement pas. D'où l'importance de faire attention aux informations que l'on met en ligne et de s'assurer que personne n'usurpe votre identité.

D'autant que faire marche arrière après la création de son profil n'est pas toujours chose aisée. [3] .

1.5 . Réseaux sociaux les plus connus :

On pose toujours la question : quel est le réseau social le plus utilisé à l'échelle mondiale ? pour répondre à cette question on doit disposer un outil informatique qui classe ces réseaux selon des critères prédéfinies, Alexa à l'un de ces outils.

Alexa est une société Internet qui analyse le trafic des autres sites avec des outils et des algorithmes qui lui sont propres. Une fois ces analyses faites, elle établit un classement des sites les uns par rapport aux autres. Ce classement est plus ou moins contesté, cependant il est un indicateur utile. Il est possible d'améliorer le classement Alexa de son site, moyennant quelques transformations. Il faut se polariser sur trois aspects-clés : le trafic, l'optimisation et la visibilité de votre site.

Selon les réseaux sociaux classifiés selon Heure quotidienne sur le site et les Pages vues quotidiennes par visiteur et d'après % du trafic de la recherche et Nombre total de sites liés . [4]

1.5 .1 Facebook :

Facebook est un réseau social en ligne qui permet à ses utilisateurs de publier des images, des photos, des vidéos, des fichiers et documents, d'échanger des messages, joindre et créer des groupes et d'utiliser une variété d'applications.

Troisième site web le plus visité au monde après Google et YouTube selon Alexa, il compte, en décembre 2015, 1,04 milliard d'utilisateurs actifs quotidiens sur un total de 1,59 milliard d'utilisateurs actifs mensuels. Le 24 août 2015, pour la première fois, un milliard de personnes ont utilisé Facebook dans la même journée.

En juin 2017, Facebook déclare avoir franchi le nombre de 2 milliards d'utilisateurs actifs.

Facebook est né en 2004 à l'université Harvard ; d'abord réservé aux étudiants de cette université, il s'est ensuite ouvert à d'autres universités américaines avant de devenir accessible à tous en septembre 2006.

Le nom du site provient des albums photo (« trombinoscopes » ou « facebook » en anglais) regroupant les photos des visages de tous les élèves prises en début d'année universitaire.

Facebook fait régulièrement l'objet de débats tant sur le plan politique que juridique, économique, culturel et social. Son influence dans la sphère publique et la manière dont il affecte la vie sociale de ses utilisateurs, son usage des données personnelles, son rôle dans la propagation des fake news ou encore sa politique de régulation des contenus sont ainsi souvent discutés dans l'actualité.[5]

1.5.2 LinkedIn :

LinkedIn est un réseau social professionnel en ligne créé en 2003 à Mountain View (Californie). En novembre 2015, le site revendique plus de 400 millions de membres issus de 170 secteurs d'activités dans plus de 200 pays et territoires. Les profils des utilisateurs LinkedIn sont des identités réelles souvent avec un CV pour trouver du travail, des employeurs, des prestataires, développer les affaires, etc.

La société est devenue bénéficiaire avec un modèle économique reposant sur trois sources de revenus :

- La publicité (18 % des ressources) .
- Les abonnements (20 %) .
- Les services de recrutement (62 %) .

LinkedIn peut être utilisé pour tout ce qui concerne la vie professionnelle : il constitue aujourd'hui un moyen efficace pour construire, développer et enrichir son capital social. Il vise à créer une relation de confiance entre des professionnels, des étudiants et des entreprises afin que chacun puisse mobiliser ces ressources en ligne pour acquérir ou développer de nouvelles idées, obtenir des opportunités d'emploi, bénéficier des communautés d'experts qui existent sur le réseau. [6]

1.5.3 Google+ :

Google+ (prononcé et parfois écrit **Google Plus**, souvent abrégé **G+**) est l'application web de réseau social de l'entreprise américaine Google lancée le 28 juin 2011, et accessible pendant près de 90 jours sur invitation, avant d'être rendue accessible au grand public le 20 septembre 2011. Elle est présentée par nombre de médias comme un produit destiné à concurrencer Facebook. En termes d'utilisateurs, Google+ est le deuxième plus grand réseau social au monde, ayant dépassé Twitter en janvier 2013 ; néanmoins cette statistique est à nuancer étant donné que chaque individu

possédant un compte Gmail est automatiquement inscrit à Google+. Très peu des utilisateurs enregistrés à Google+ sont réellement actifs..

Les utilisateurs de Google+ peuvent voir les mises à jour de leurs contacts grâce à des cercles à travers le « Stream », qui est semblable aux « flux de nouvelles » de Facebook. La zone de saisie permet aux utilisateurs de se mettre à niveau sur les états ou l'utilisation des icônes à télécharger et partager des photos et vidéos.

En décembre 2011, il y a plus de 60 millions d'utilisateurs. Fin janvier 2012, Google+ a passé la barre des 90 millions d'utilisateurs. Toutefois en février 2012, deux études montrent que le temps moyen passé par chaque utilisateur de ce réseau social est faible comparé à Facebook (un peu plus de 3 minutes par mois pour Google+ contre plus de 6 h pour Facebook) et même qu'il a diminué jusqu'en décembre 2011 ; les données de ces études ont été dénoncées comme « particulièrement sous-évaluées » par Google . [7]

1.5.6 Instagram :

Instagram est une application, un réseau social et un service de partage de photos et de vidéos disponibles sur plates-formes mobiles de type iOS, Android et Windows Phone.

instagram permet de partager ses photographies et ses vidéos avec son réseau d'amis, de fournir une appréciation positive (fonction « j'aime ») et de laisser des commentaires sur les clichés déposés par les autres utilisateurs. Elle permet aussi de dialoguer avec les membres via l'utilisation de la messagerie interne appelée « Instagram direct ». Les applications telles qu'Instagram contribuent à la pratique de la phonographies, ou photographie avec un téléphone mobile . [8]

1.5.7 Twitter :

Twitter est un réseau social de microblogage géré par l'entreprise Twitter Inc. Il permet à un utilisateur d'envoyer gratuitement de brefs messages, appelés tweets, sur internet, par messagerie instantanée ou par SMS. Ces messages sont limités à 280 caractères.

Twitter a été créé le 21 mars 2006 par Jack Dorsey, Evan Williams, Biz Stone et Noah Glass, et lancé en juillet de la même année. Le service est rapidement devenu populaire, jusqu'à réunir plus de 500 millions d'utilisateurs dans le monde fin février 2012. Au 5 mars 2017, Twitter compte 313 millions d'utilisateurs actifs par mois avec 500 millions de tweets envoyés par jour et est disponible en plus de 40 langues.

Le siège social de Twitter Inc. se situe aux États-Unis à San Francisco. L'entreprise dispose de plus de 35 bureaux supplémentaires à travers le monde et de serveurs informatiques à New York. [9]

1.6. Avantages et les inconvénients Des Réseaux Sociaux:

Il faut dire que les réseaux sociaux jouent un rôle important dans la vie quotidienne. Facebook, Twitter, instagram,, ils sont partout! Presque tous les jeunes utilisent, au moins, un réseau social chaque jour. Ils sont devenus les principaux moyens de communication. Mais, il y a toujours des avantages et des inconvénients .

Un avantage des réseaux sociaux, c'est qu'ils sont un moyen gratuit et facile de communiquer avec les autres. On peut communiquer avec des amis ou des membres de la famille de partout dans le monde presque tout de suite. En outre, les réseaux sociaux, comme Facebook, sont un bon moyen de rester en contact avec vos amis anciens et nouveaux. On peut partager ses pensées, photos, vidéos avec le clic.

Les réseaux sociaux sont un outil utile pour les entreprises ainsi que les jeunes. Ils permettent aux entreprises d'avoir un contact direct avec leurs clients. La plupart des entreprises ont leur propre page Facebook ou compte twitter où ils peuvent informer les clients des nouveaux produits ou des offres spéciales ainsi que de recevoir les commentaires des clients. Les réseaux sociaux offrent aux entreprises une grand espace de publicité.

Les réseaux sociaux peuvent aussi être considérés comme des outils de promotion pour des entreprises commerciales et des fournisseurs de service grâce à des « pages » permettant de faire connaître l'entreprise.

Un individu peut aussi, grâce aux réseaux sociaux, acquérir une notoriété publique. En effet, s'il crée un blog ou une chaine Youtube, et poste des photos ou vidéos faisant du « buzz », avec un certain nombre de «vue» et de «like», il pourra devenir célèbre grâce à ses publications et même avoir la possibilité d'être payé.

Les réseaux sociaux sont aussi un remède à la solitude. En effet ils comblent le vide relationnel et peuvent apporter à des personnes seules une certaine compagnie.

Des outils valorisants, Ces mêmes réseaux sociaux contribuent également, pour plusieurs, à renforcer leur ego. Ils y étalent leurs bons coups, leurs succès, y minimisent leurs échecs. Bref, les adeptes peuvent s'y montrer sous un jour avantageux, voire narcissique, en plus de bonifier le sentiment d'appartenance à un groupe défini. [10]

Mais, aussi, il y a des inconvénients, réseaux sociaux peuvent être très addictif. Il y a des gens qui ont besoin d'être sur ces sites toutes les heures ou toutes les minutes ! Cela peut créer de graves problèmes pour eux. Beaucoup d'étudiants trouvent qu'il est difficile à étudier car ils sont trop distraits par leur Facebook. Souvent, les résultats des examens sont pauvres à cause de cela. D'autres personnes peuvent prendre du retard au travail parce qu'ils sont trop occupés avec les comptes.

Un problème inquiétant avec les réseaux sociaux est la cyber intimidation. C'est à ce moment que les utilisateurs disent des choses méchantes et blessantes sur d'autres utilisateurs. Il peut avoir de graves conséquences.

La vie privée, Informer vos « amis » de certaines facettes de votre vie privée peut entraîner des problèmes. Combien d'adeptes de Facebook ont été victimes de cambriolage après avoir annoncé, en grande pompe, un voyage dans les Caraïbes? Ou cette autre femme qui s'est vue retirer ses prestations d'assurance chômage après avoir publié des photos d'elle croquées lors d'un voyage dans le Sud.

Beaucoup de gens qui utilisent ces sites ne sais pas si elles sont publiques ou privées. Souvent, les gens mettre des photos sur leur page qu'ils ne veulent pas tout le monde de voir. Il est important d'être conscient des policiers confidentialité de ces sites.

Le bilan des impacts des réseaux sociaux sur la société est plutôt partagé. Sauf que la vraie problématique porte sur l'usage que nous faisons des médias sociaux. L'inconscience mène aux dérives, alors qu'à l'inverse un usage réfléchi peu de révéler enrichissant. Il en est des réseaux sociaux comme de la sécurité, l'humain est le maillon faible... [11]

Les réseaux sociaux ont des inconvénients mais s'ils sont utilisés correctement et avec modération, ils sont utiles et amusants. Facebook, Twitter, Instagram, LinkedIn... Tous ses réseaux sociaux permettent de rapprocher les gens au niveau privé ou professionnel.

1.7. Conclusion

La plupart des réseaux sociaux se développent tellement qu'il est difficile d'avoir un point de vue critique. En y réfléchissant, personne n'a forcément besoin d'avoir 500, 800 ou même 5000 amis... L'amitié se construit dans la réalité avec des personnes en qui on a confiance. Le plus important est la qualité du lien social et non pas la quantité. La vie professionnelle, Pour de nombreux employeurs, ces réseaux sont une véritable mine de renseignements pour recruter ou congédier du personnel. Ou pour vendre un produit ou présenter un service. Les avis et les points de vue de utilisateurs est un moyen pour

mesurer l'importance de tel service vis-à-vis un autre. Collecter et analyser ces opinions est devenue une tâche primordiale est c'est l'objectif de l'analyse des sentiment et l'opinion mining qui sera détaillé dans le prochain chapitre.

CHAPITRE 2

ANALYSE DES SENTIMENT ET OPINION MINING

2.1 Introduction :

L'analyse des sentiments est une notion proche de celle d'analyse des émotions, mais elle va généralement moins loin dans l'analyse et se contente souvent de déterminer trois niveaux (favorable, neutre, défavorable) à l'égard de la marque, de l'individu ou de l'événement.

L'analyse de sentiment demande bien plus de compréhension de la langue que l'analyse de texte et la classification par sujet. En effet, si les algorithmes les plus simples considèrent uniquement les statistiques de fréquence d'apparition des mots, cela se révèle en général insuffisant pour définir l'opinion dominante dans un document, surtout lorsque le contenu est court comme des messages dans un forum ou des tweets.

2.2 L'analyse des sentiments et opinion mining :

L'analyse des sentiments est une démarche principalement basée sur le text mining et l'analyse sémantique qui permet de déterminer la « position » des individus étudiés à l'égard d'une marque ou d'un événement. L'analyse des sentiments peut cependant également reposer sur d'autres éléments que les données textuelles. Elle peut par exemple être basée sur l'usage des émoticônes (emojis), sur les « émotions » facebook, sur l'analyse de la voix ou même sur le facial coding / decoding.

Lorsqu'elle est basée sur la fouille textuelle, l'analyse des sentiments peut porter sur des verbatims issus des réseaux sociaux, des avis, forums, de données d'études qualitatives ou quantitatives, etc. Les sentiments sont généralement classés en trois types : négatifs, neutres et positifs. [12]

2.3 Niveaux d'analyse des sentiments :

L'analyse des sentiments est faite selon des niveaux selon l'objectif :

- **Au niveau du document :** détermine l'opinion générale de l'ensemble du document. Cette analyse fonctionne bien pour des documents qui présentent un

point de vue précis, mais moins pour des comparaisons car elle ne fera pas la différence entre les sujets abordés.

- **Au niveau de la phrase** : détermine l'opinion générale d'une phrase (positive, négative ou neutre). Cette analyse peut donner une mesure de la "neutralité" d'un texte par exemple pour analyser des entrées de Wikipédia. Les méthodes utilisées sont celle de l'analyse de subjectivité.

- **Au niveau des aspects** : (appelé en anglais Feature level) : au lieu de déterminer les entités à analyser en fonction de critère structuraux (phrase, paragraphe, document) ces méthodes se basent sur l'analyse de corrélation entre l'opinion émise et la cible de cette opinion. Par exemple, la phrase "Le sujet du cours me passionne mais le professeur est ennuyeux." présente deux sentiments sur l'entité "cours" : le sujet qui est perçu comme positif et le professeur, qui est perçu comme négatif. Ce niveau d'analyse permet de différencier les aspects qui sont aimés ou non par les auteurs des textes et ainsi permet plus facilement de déterminer des remédiations possibles. En revanche il est très difficile à mettre en place car ce type d'analyse est extrêmement complexe.[13]

2.4 processus d'analyse des sentiments sur les avis sur les produits

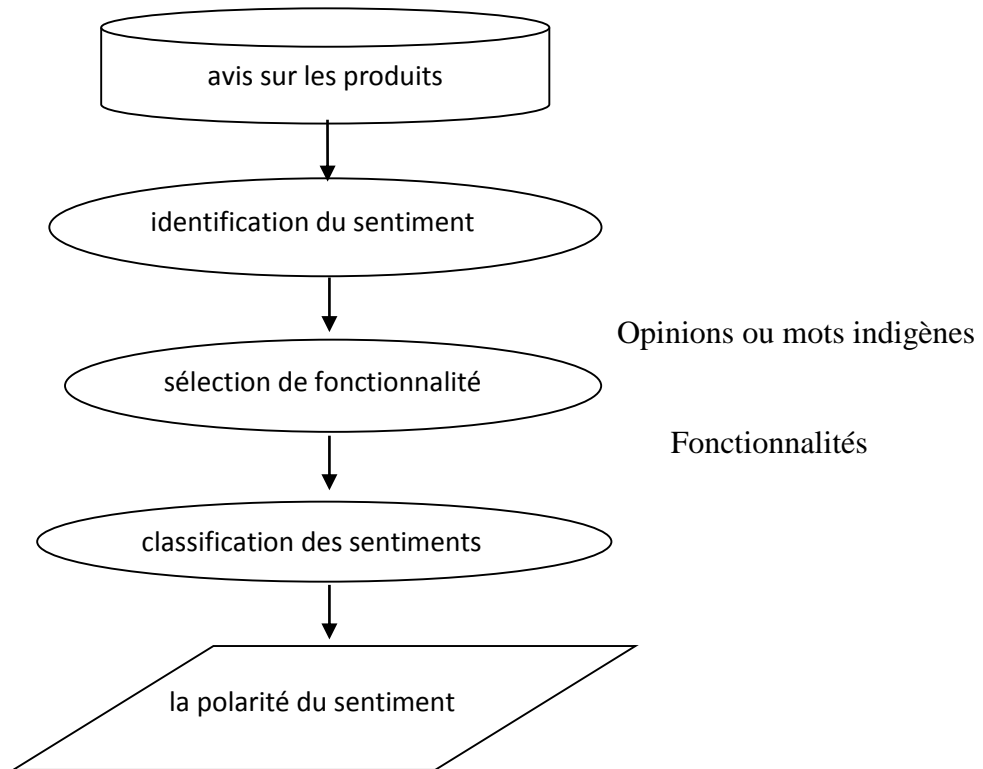


figure 2.1 processus d'analyse des sentiments sur les avis sur les produits

2.5 Disciplines en relation avec l'analyse de sentiments :

Plusieurs disciplines ont une relation directe ou moins directe avec l'analyse de sentiments et l'opinion mining. Intelligence artificiel, le traitement automatique du langage naturel, texte mining et même le data mining offrent des outils et algorithmiques indispensables pour le traitement et la classification des sentiments.

2.5.1 fouille de texte :

La fouille de texte ou Text mining est l'analyse des données contenues dans un texte en langage naturel. L'application de techniques d'exploration de texte pour résoudre des problèmes métier est appelée analyse de texte.

Text mining peut aider une organisation à tirer des informations commerciales potentiellement précieuses à partir de contenus textuels tels que des documents Word, des e-mails et des publications sur des réseaux sociaux tels que Facebook, Twitter et LinkedIn. L'exploration de données non structurées avec le traitement automatique du

langage Naturel (TALN), la modélisation statistique et les techniques d'apprentissage automatique peut toutefois être difficile, car le texte en langage naturel est souvent incohérent. Il contient des ambiguïtés causées par une syntaxe et une sémantique incohérentes, y compris l'argot, le langage spécifique aux industries verticales et aux groupes d'âge, les doubles sens et le sarcasme.

Les logiciels d'analyse de texte peuvent aider en transposant des mots et des phrases dans des données non structurées en valeurs numériques qui peuvent ensuite être liées à des données structurées dans une base de données et analysées avec des techniques d'exploration de données traditionnelles. Avec une approche itérative, une organisation peut utiliser avec succès l'analyse de texte pour mieux comprendre des valeurs spécifiques au contenu telles que le sentiment, l'émotion, l'intensité et la pertinence. Parce que la technologie d'analyse de texte est toujours considérée comme une technologie émergente, les résultats et la profondeur de l'analyse peuvent varier énormément d'un fournisseur à l'autre. [14]

2.5.2 Traitement automatique du langage naturel (TALN) :

La TALN est un moyen pour les ordinateurs d'analyser, de comprendre et de dériver le sens du langage humain d'une manière intelligente et utile. En utilisant le langage TALN, les développeurs peuvent organiser et structurer les connaissances pour exécuter des tâches telles que la synthèse automatique, la reconnaissance d'entités nommées, l'extraction de relations, l'analyse des sentiments, la reconnaissance de la parole et la segmentation des rubriques.

"En dehors des opérations courantes de traitement de texte qui traitent le texte comme une simple séquence de symboles, la TALN considère la structure hiérarchique du langage : plusieurs mots forment une phrase, plusieurs phrases forment une phrase et, finalement, des phrases qui transmettent des idées." expert de Meltwater Group, a déclaré dans Comment le traitement du langage naturel aide à découvrir le sentiment des médias sociaux. "En analysant le langage pour sa signification, les systèmes TALN ont depuis longtemps rempli des rôles utiles, tels que la correction de la grammaire, la conversion de la parole en texte et la traduction automatique entre les langues."

La TALN est utilisée pour analyser le texte, permettant aux machines de comprendre comment l'humain parle. Cette interaction homme-ordinateur permet des applications réelles telles que la synthèse automatique de texte, l'analyse de sentiment, l'extraction de sujet, la reconnaissance d'entité nommée, l'étiquetage de parties de discours, l'extraction

de relations, le stemming, etc. La TALN est couramment utilisée pour l'exploration de texte, la traduction automatique et la réponse automatique aux questions.

Les algorithmes TALN sont généralement basés sur des algorithmes d'apprentissage automatique. Au lieu de coder manuellement de grands ensembles de règles, la TALN peut s'appuyer sur l'apprentissage automatique pour apprendre automatiquement ces règles en analysant un ensemble d'exemples (un grand corpus, comme un livre, jusqu'à une collection de phrases) et en faisant une inférence statique. . En général, plus le nombre de données analysées est élevé, plus le modèle sera précis. [16]

2.5.3 Apprentissage automatique :

L'apprentissage automatique (en anglais machine learning, littéralement « l'apprentissage machine »), champ d'étude de l'intelligence artificielle, concerne la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine (au sens large) d'évoluer par un processus systématique, et ainsi de remplir des tâches difficiles ou problématiques par des moyens algorithmiques plus classiques.

L'analyse peut concerner des graphes, arbres, ou courbes (par exemple, la courbe d'évolution temporelle d'une mesure ; on parle alors de données continues, par opposition aux données discrètes associées à des attributs-valeurs classiques) au même titre que de simples nombres. [17].

Les algorithmes utilisés permettent, dans une certaine mesure, à un système piloté par ordinateur (un robot éventuellement), ou assisté par ordinateur, d'adapter ses analyses et ses comportements en réponse, en se fondant sur l'analyse de données empiriques provenant d'une base de données ou de capteurs.

La difficulté réside dans le fait que l'ensemble de tous les comportements possibles compte tenu de toutes les entrées possibles devient rapidement trop complexe à décrire (on parle d'explosion combinatoire). On confie donc à des programmes le soin d'ajuster un modèle pour simplifier cette complexité et de l'utiliser de manière opérationnelle. Idéalement, l'apprentissage visera à être non supervisé, c'est-à-dire que la nature des données d'entraînement n'est pas connue.

Ces programmes, selon leur degré de perfectionnement, intègrent éventuellement des capacités de traitement probabiliste des données, d'analyse de données issues de capteurs, de reconnaissance (reconnaissance vocale, reconnaissance de forme, d'écriture...), de data-mining, d'informatique théorique...[18]

Lorsque les données ne sont pas étiquetées, l'apprentissage supervisé n'est pas possible, et une approche d'apprentissage non supervisée est nécessaire, qui tente de trouver un

regroupement naturel des données à des groupes, puis mapper de nouvelles données à ces groupes formés. L'algorithme de cluster vectoriel de support créé par Hava Siegelmann et Vladimir Vapnik applique les statistiques des vecteurs de support développés dans l'algorithme des machines vectorielles de support pour catégoriser les données non étiquetées et est l'un des algorithmes de clustering les plus utilisés dans les applications industrielles. [17]

2.6 Méthodes de classification des sentiments

L'analyse des sentiment et l'opinion mining se base essentiellement sous la classification des sentiments en trois catégories (positif, négatif, et neutre) selon [19] il existe deux approches principales :

- Approche basée sur l'apprentissage automatique
- Approche basée sur le TALN

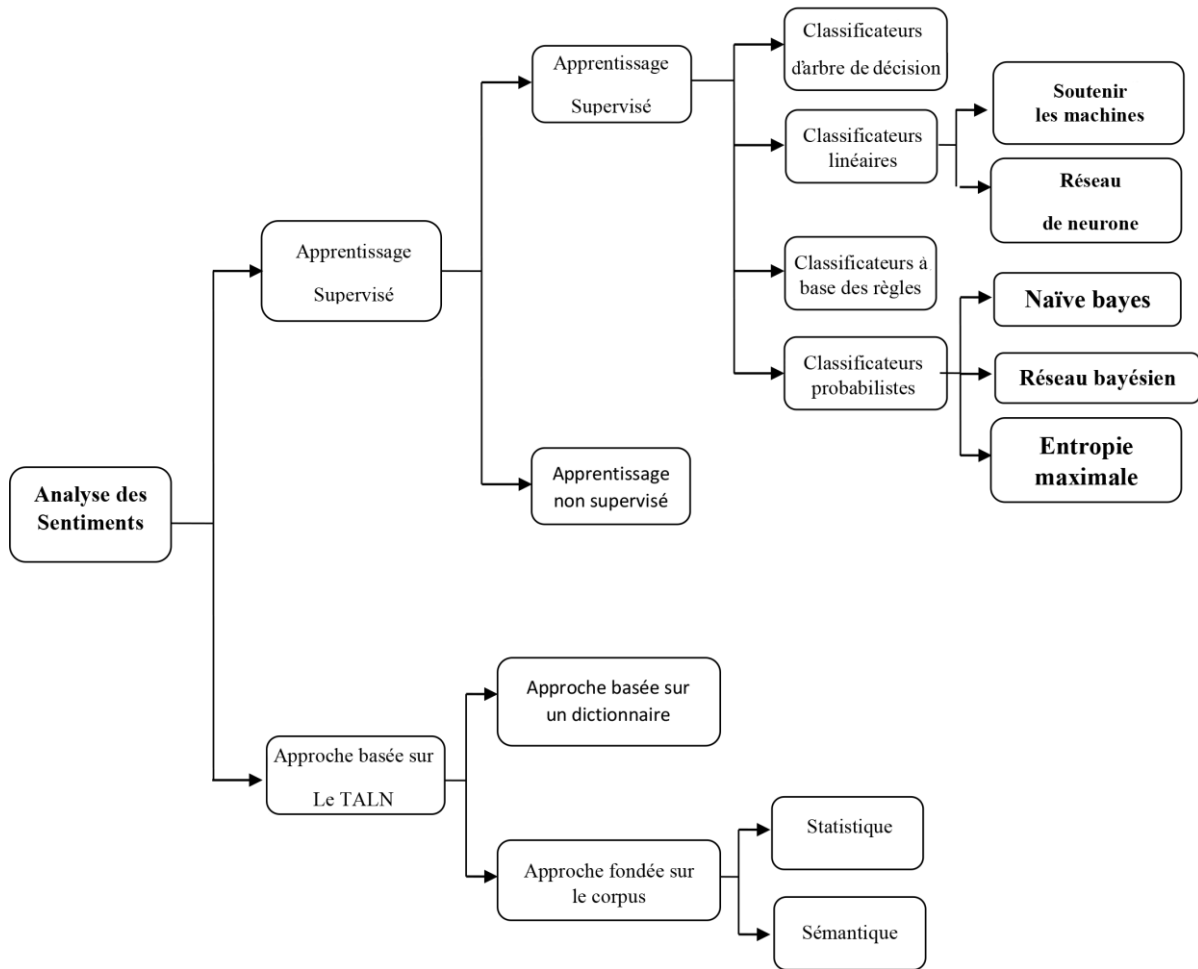


Figure 2.2 méthodes de classification du sentiment

Les techniques de l'apprentissage automatique sont connues, dans ce qui suit on va détailler l'approche basée TALN qui sera le cœur de notre travail :

2.6.1 Approche basée le TALN :

Le traitement automatique du langage naturel est basé sous le traitement des phrases, tandis que les phrases sont constituées des mots. Dans l'analyse de sentiment on trouve des mots qui sont utilisés pour exprimer l'opinion, un lexique ou vocabulaire bien déterminé est employé dans la classification de sentiment. Il y a des mots d'opinion positifs qui sont utilisés pour exprimer des sentiments positifs tel que : il me plaît ; beau, joli , intéressant...etc et des mots d'opinion négatifs qui sont utilisés pour exprimer les sentiments négatifs. Il y a aussi des phrases d'opinion et des idiomes qui ensemble sont

appelés lexique d'opinion. Il y a trois principales approches afin de compiler ou de recueillir la liste des mots d'opinion.

L'approche manuelle prend beaucoup de temps et n'est pas utilisée seule. Elle est généralement combinée avec les deux autres automatisées approches comme une vérification finale pour éviter les erreurs qui ont résulté à partir de méthodes automatisées. Les deux approches automatisées sont les suivantes.[19]

2.6.2 Approche basée sur un dictionnaire:

Un petit ensemble de mots d'opinion est collecté manuellement avec des orientations connues. Ensuite, cet ensemble est cultivé par la recherche dans les corpus bien connus WordNet ou thesaurus pour leurs synonymes et antonymes. Les mots nouvellement trouvés sont ajoutés à la liste de départ alors l'itération suivante commence. L'itératif le processus s'arrête quand aucun nouveau mot n'est trouvé. Après le processus est terminée, une inspection manuelle peut être effectuée par un expert pour supprimer ou corriger les erreurs.

2.6.3 Approche basée sur le corpus:

L'approche basée sur le corpus aide à résoudre le problème de trouver des mots d'opinion avec des orientations spécifiques au contexte. Ses méthodes dépendent des modèles syntaxiques ou des modèles qui se produisent ensemble avec un corpus de mots d'opinion pour trouver d'autres mots d'opinion dans le même contexte. Une de ces méthodes était représentée par Hatzivassiloglou et McKeown. Ils commencent avec une liste d'adjectifs d'opinion de semences, et les a utilisés avec un ensemble de contraintes linguistiques pour identifier d'autres mots d'opinion adjectifs et leurs orientations. Les contraintes sont pour des connecteurs comme ET, OU, MAIS, SOIT OU-OU. . . . la conjonction ET par exemple dit que les adjectifs conjoints ont généralement la même orientation. Cette idée est appelée la cohérence du sentiment, qui n'est pas toujours cohérente dans la pratique.

Il y a aussi des expressions adversatives telles que mais, cependant, qui sont indiqués comme des changements d'opinion. Afin de déterminer si deux adjectifs conjoints sont identiques ou différentes orientations, l'apprentissage est appliqué à un grand corpus. Alors, les liens entre les adjectifs forment un graphique et le regroupement est effectué sur le graphique pour produire deux ensembles de mots: positif et négatif. [19]

2.6.4 l'approche basée sous apprentissage automatique :

Dans l'apprentissage automatique, les machines vectorielles de support (SVM, également support des réseaux vectoriels) sont des modèles d'apprentissage supervisé avec des algorithmes d'apprentissage associés qui analysent les données utilisées pour la classification et l'analyse de régression. Étant donné un ensemble d'exemples d'apprentissage, chacun marqué comme appartenant à l'une ou l'autre des deux catégories, un algorithme d'apprentissage SVM construit un modèle qui attribue de nouveaux exemples à une catégorie ou à une autre, ce qui en fait un classificateur binaire linéaire non probabiliste. Un modèle SVM est une représentation des exemples sous la forme de points dans l'espace, mappés afin que les exemples des catégories séparées soient divisés par un espace libre aussi large que possible. De nouveaux exemples sont ensuite cartographiés dans ce même espace et prédits pour appartenir à une catégorie en fonction de quel côté de l'écart ils tombent.

En plus d'effectuer une classification linéaire, les SVM peuvent effectuer efficacement une classification non-linéaire en utilisant ce que l'on appelle l'astuce du noyau, en mettant implicitement en correspondance leurs entrées dans des espaces de caractéristiques de grande dimension.

2.6.5 Analyse sémantique latente :

L'analyse sémantique latente (LSA, de l'anglais : Latent semantic analysis) ou indexation sémantique latente (ou LSI, de l'anglais : Latent semantic indexation) est un procédé de traitement des langues naturelles, dans le cadre de la sémantique vectorielle. La LSA fut brevetée en 1988 et publiée en 1990.

Elle permet d'établir des relations entre un ensemble de documents et les termes qu'ils contiennent, en construisant des « concepts » liés aux documents et aux termes.

2.7 Outils d'analyse des sentiments :

Il existe des outils permettant d'identifier le sentiment dégagé par un texte. Voici une liste non exhaustive des outils les plus connus :

2.7.1 Werfamous :

outil d'analyse en ligne gratuit, donnant un score de sentiment sur une échelle de -100 à 100, ainsi qu'un niveau de confiance lié à ce score.

Le site sauvegarde des traceurs textes (cookies) sur votre appareil afin de vous garantir de meilleurs contenus et à des fins de collectes statistiques. Vous pouvez désactiver l'usage des cookies en changeant les paramètres de votre navigateur. En poursuivant votre

navigation sur notre site sans changer vos paramètres de navigateur vous nous accordez la permission de conserver des informations sur votre appareil.

Vous placez votre requête dans la boîte de dialogue et obtenez une analyse sur le web. Les requêtes sont sauvegardées dans leur Jeu de données et Historique sections.[20]

2.7.2 AFINN :

AFINN est une liste de mots valant pour valence avec un nombre entier entre moins cinq (négatif) et plus cinq (positif) .

évalue la positivité/négativité d'un mot à l'aide d'un dictionnaire contenu dans une archive

Bien que le titre du document associé suggère qu'il est basé sur le corpus étiqueté ANEW, il ne l'est pas. Le titre est simplement un wordpun. Il a été développé indépendamment de la liste de mots, et ce n'est pas une révision de celui-ci. Par rapport à ANEW, la liste de mots AFINN a plus de mots et comprend des mots obscènes. ANEW d'autre part a (outre la valence) l'excitation et la dominance pour chaque mot et chaque mot a été étiqueté par plusieurs personnes et la moyenne et la déviation standard sont données. L'AFINN n'a été étiqueté que par Finn Årup Nielsen. Finn Årup Nielsen n'était aucunement impliqué dans le développement d'ANEW. ANEW a été développé par Margaret M. Bradley et Peter J. Lang.[21]

2.7.3 General Inquirer :

lemmatise les mots, effectue une analyse graphique et statistique et produit un rapport contenant des phrases avec les mots les plus significatifs .

Ce site est divisé en plusieurs sections, donnant à la fois des informations sur l'Inquirer et des points vers d'autres systèmes. Les pages de notre site Web contiennent des liens vers les 100 premiers mots de chaque catégorie dans les dictionnaires Harvard et Lasswell. [22]

2.7.4. SenticNet :

Parler de SenticNet, c'est parler d'analyse des sentiments au niveau conceptuel, c'est-à-dire effectuer des tâches telles que la détection de polarité et la reconnaissance des émotions en s'appuyant uniquement sur la sémantique et la linguistique.

Dans ce contexte, SenticNet peut être l'une des choses suivantes:

- une base de connaissances conceptuelle;
- un cadre multidisciplinaire;
- une entreprise privée.

En tant que base de connaissances, SenticNet fournit un ensemble de sémantiques, de sentiques et de polarité associées à 100 000 concepts de langage naturel. En particulier, la sémantique est la plus sémantiquement liée au concept d'entrée (ie les cinq concepts qui partagent plus de caractéristiques sémantiques avec le concept d'entrée), les sentics sont des valeurs de catégorisation des émotions exprimées en termes de quatre dimensions affectives (Agréable, Attention, Sensibilité, et Aptitude) et la polarité est un nombre flottant entre -1 et +1 (où -1 est la négativité extrême et +1 est la positivité extrême). La base de connaissances est téléchargeable gratuitement en tant que fichier XML autonome et sa dernière version (publiée tous les deux ans) est également accessible en tant qu'API.

En tant que cadre, SenticNet se compose d'un ensemble d'outils et de techniques pour l'analyse des sentiments combinant le raisonnement de bon sens, la psychologie, la linguistique et l'apprentissage automatique. Dans ce contexte, SenticNet est plus communément appelé «informatique sismique», un paradigme multidisciplinaire qui va au-delà des simples approches statistiques pour se concentrer sur une représentation préservant la sémantique des concepts de langage naturel et sur la structure des phrases.[23]

2.7.5 SentiWordNet :

SentiWordNet est une ressource lexicale pour l'extraction d'opinion. SentiWordNet attribue à chaque synset de WordNet trois scores de sentiment: positivité, négativité, objectivité.

Dans le domaine de l'analyse de sentiment, une étude comparative a été effectuée afin de déterminer quels étaient les avantages et inconvénients de chaque source de données. Dans le cadre d'analyse de tweets relatifs à des événements majeurs, l'étude met en avant le fait que plusieurs de ces tweets n'ont pas pu être reconnus par les sources de données. On peut y voir que SentiWordNet, SenticNet et SentiStrength semblent couvrir un plus grand nombre de tweets. Cependant l'article met également en évidence que le taux de couverture n'est pas synonyme de reconnaissance efficace et que la polarité d'un mot donné n'est pas fiable. C'est pourquoi l'article se propose de combiner plusieurs de ces méthodes afin d'exploiter les avantages de chacun et d'obtenir le résultat le plus proche possible de la réalité.[25]

2.7.6. Sentiment140

Sentiment140 (anciennement connu sous le nom "Twitter sentiment") est un outil en ligne gratuit qui a été créé par trois étudiants en computer science de Stanford, donc c'est un projet académique. Cet outil, contrairement à la plupart des autres sites d'analyse

de sentiments, n'utilise pas de listes de mots positifs ou négatifs mais est fondé sur les algorithmes d'apprentissage automatique [26].

Sentiment140 permet de découvrir des sentiments des tweets d'une marque, un produit ou un sujet sur Twitter. Site officiel : <http://www.sentiment140.com> .

2.7.7. Tweetfeel

Tweetfeel est un service qui s'appuie sur les capacités temps réels de **Twitter** pour vous donner le sentiment des utilisateurs de Twitter sur un mot clé, une marque ou encore une star.

L'évaluation de TweetFeel se fait sur la base de présence de mots clés précis dans les tweets tels que Good, Bad, etc... (Uniquement anglais pour l'instant), a quand un service similaire en français.

Ensuite un pourcentage est calculé selon le nombre de tws ou négatifs un sentiment global de Twitter sur la marque. [27]Site officiel : <http://www.tweetfeel.com> .

2.7.8 Twitrratr

Twitrratr est un outil en ligne gratuit, qui a émergé à partir d'un projet Startup Weekend. Twitrratr fonctionne à partir d'une liste de mots positifs et d'une liste de mots négatifs [37].

Cet outil classe une opinion sur le mot clé de la requête s'il est capable de le croiser avec un mot d'une des deux listes. Les mots positifs et négatifs qui servent à classer les tweets sont surlignés dans l'interface [28].Site officiel <http://twitrratr.com> .

2.7.9 Tweet Sentiments Analyses

Tweet Sentiments Analyses est un outil en ligne gratuit et open source d'analyse du sentiment sur Twitter. Il peut donner des sentiments positifs, négatifs et neutres des tweets sur le mot clé lancé dans la requête. Il peut travailler sur 12 langues. Il donne les résultats sous forme graphique. [29] Site officiel : <http://smm.streamcrab.com> .

2.8 Conclusion :

Dans ce chapitre nous avons fait un survole sur les principaux fondements, méthodes et techniques de classification des sentiments et d'opinion mining. dans le chapitre suivant nous allons détailler plus les travaux faites dans ce domaine en relation avec le traitement automatique du langage naturel et présenter notre système de classification de sentiment.

CHAPITRE 3

CONCEPTION ET RÉALISATION

3.1. Introduction

De nos jours, Opinion Mining est devenu un sujet de recherche émergent en raison du grand nombre de données disponibles sur les blogs et les réseaux sociaux. Suivre différents types d'opinions et les résumer peut fournir des informations précieuses aux différents types d'opinions aux utilisateurs qui utilisent les réseaux sociaux pour obtenir des commentaires sur tout produit, service ou tout sujet. L'analyse des opinions et sa classification sur la base de la polarité (positive, négative, neutre) est une tâche difficile. Beaucoup de travail a été fait sur l'analyse des sentiments des données de Twitter et beaucoup de choses doivent être faites.

3.2. Travaux réalisés dans le domaine de l'Opinion Mining

Plusieurs travaux sont réalisés pour classifier les données dans cette section nous allons présenter quelques un suivant les grands axes :

3.2.1 analyse des sentiments basée sur le lexique (2010) :

La tâche principale abordée dans cette recherche est la classification du texte à l'aide d'étiquettes. Le système @AM développé repose sur le principe de compositionnalité et une nouvelle approche basée sur les règles élaborées pour distinguer sémantiquement les classes de verbes. L'évaluation de cette méthode est faite sur 1000 phrases, qui décrivent des expériences personnelles, a montré des résultats prometteurs : précision moyenne sur le grain fin niveau (14 étiquettes) était de 62%, sur niveau moyen (7 étiquettes) - 71%, et sur le niveau supérieur (3 étiquettes) - 88%. [33]

3.2.2 Analyse des sentiments basés sur LSA (2010) :

Les revues d'utilisateurs en ligne aident les consommateurs à faire face aux surcharges d'information et facilite la prise de décision. Cependant, de nombreux avis d'utilisateurs en ligne ne disposent pas de suffisamment de votes d'utilité pour les autres utilisateurs afin d'évaluer leur véritable niveau d'utilité. Cette étude examine empiriquement l'impact des différentes caractéristiques, c'est-à-dire, les caractéristiques de base, stylistiques et sémantiques des revues d'utilisateurs en ligne sur le nombre de votes d'utilité que ces revues reçoivent. Les techniques d'exploration de texte sont utilisées pour extraire les caractéristiques sémantiques des textes de révision. Les résultats montrent que les caractéristiques sémantiques sont plus influentes que d'autres

caractéristiques en ce qui concerne le nombre de votes d'utilité reçus. Les résultats suggèrent également que les critiques avec des opinions extrêmes reçoivent plus de votes d'utilité que ceux avec des opinions mitigées ou neutres. Ce document met en lumière la compréhension du comportement de vote d'utilité des utilisateurs en ligne et la conception d'un mécanisme de vote de meilleure utilité pour les systèmes d'examen en ligne des utilisateurs.[34]

3.2.3 SVM multi-classes (2011) :

L'omniprésence de Web2.0 fait du Web une source inestimable d'informations commerciales. Par exemple, le produit Les critiques rédigées en collaboration par de nombreux critiques Internet indépendants peuvent aider les consommateurs à décisions d'achat et permettre aux entreprises d'améliorer leurs stratégies commerciales. Comme le nombre de commentaires est

Croissance exponentielle, les techniques d'exploration et de recherche d'opinion sont nécessaires pour identifier les examens importants et des opinions pour répondre aux questions des utilisateurs. La plupart des approches d'extraction et d'extraction d'opinion tentent d'extraire expressions sentimentales ou bipolaires d'un grand volume de critiques. Cependant, le processus ignore souvent qualité de chaque révision et peut récupérer des documents inutiles ou même bruyants. Dans cet article, nous proposons une méthode d'évaluation de la qualité de l'information dans les revues de produits. Nous traitons l'évaluation de l'examen la qualité comme un problème de classification et emploient un cadre efficace de qualité de l'information pour extraire caractéristiques d'examen représentatif. Des expériences basées sur un corpus de données composé d'experts démontrent que La méthode proposée surpasse considérablement les approches de pointe.[35]

3.2.4 Analyse des sentiments basés sur le lexique, NLP (2012) :

Faits saillants - Capacité à gérer la tendance de nombreux utilisateurs à exprimer leurs opinions dans un langage non standard. Détecte les cibles des opinions des utilisateurs dans des scénarios multi-domaines. Modèle de connaissance linguistique modulaire avec adaptabilité à faible coût. Lexique hiérarchique spécialement conçu pour analyser les commentaires des nouvelles. Grâce à la révolution technologique qui a accompagné le Web 2.0, les utilisateurs peuvent interagir de manière intensive sur Internet, comme en témoignent les réseaux sociaux, les blogs, les forums, etc. Dans ces scénarios, les utilisateurs peuvent s'exprimer librement sur tout sujet pertinent. Cependant, le volume élevé de contenu généré par les utilisateurs rend une analyse

manuelle de ce discours non viable. Par conséquent, des techniques d'analyse automatique sont nécessaires pour extraire les opinions exprimées dans les commentaires des utilisateurs, étant donné que ces opinions sont un baromètre implicite d'un intérêt indiscutable pour une grande variété d'entreprises, d'agences et d'organismes. Nous proposons donc un News Sentiment Analyzer (LCN-SA) axé sur les commentaires basé sur le lexique, qui est capable de gérer les éléments suivants: (a) la tendance de nombreux utilisateurs à exprimer leurs opinions dans un langage non standard; (b) la détection de la cible des opinions des utilisateurs dans un scénario multi-domaine; (c) la conception d'un modèle de connaissance linguistique modularisé avec une adaptabilité à faible coût. Le système proposé consiste en un module de détection de focalisation automatique et un module d'analyse de sentiment capable d'évaluer les opinions des utilisateurs sur les sujets abordés dans les nouvelles. Ces modules utilisent une taxonomie-lexique spécifiquement conçue pour l'analyse des nouvelles. Les expériences montrent que les résultats obtenus jusqu'ici sont extrêmement prometteurs. [36]

3.2.5. Analyse de sentiment basée sur l'analyse Sémantique (2013) :

L'estimation de la polarité dans les domaines à grande échelle et multi-sujets est une question difficile. La plupart des solutions de pointe reposent essentiellement sur des fréquences de mots porteurs de sentiments (extraits d'un lexique, par exemple) lors de l'analyse du sentiment véhiculé par un texte en langage naturel. Ces approches ignorent les aspects structurels d'un document, qui contiennent des informations précieuses. La théorie de la structure rhétorique (RST) fournit des informations importantes sur l'importance relative des différentes zones de texte dans un document. Cette connaissance pourrait être utile pour l'analyse des sentiments et la classification de la polarité. Cependant, la RST n'a été étudiée que pour des problèmes de classification de polarité dans des scénarios contraints et à petite échelle. L'objectif principal de cet article est d'explorer l'utilité de RST dans le classement de polarité à grande échelle des articles de blog. Nous appliquons des méthodes au niveau de la phrase pour sélectionner les phrases clés qui expriment le sentiment général d'un article sur le sujet. Ensuite, nous appliquons l'analyse RST à ces phrases de base afin de guider la classification de leur polarité et de générer ainsi une estimation globale de la polarité du document par rapport à un sujet spécifique. Nos résultats montrent que RST fournit des informations précieuses sur la structure discursive des textes qui peuvent être utilisés pour faire un

classement plus précis des documents en termes de leur sentiment estimé dans les blogs multi-sujets.[37]

3.2.6 Analyse de sentiment basé sur apprentissage automatique (2014)

A présenté des variations des classificateurs Naive Bayes pour détecter la polarité des tweets anglais. Deux variantes différentes des classificateurs Naive Bayes ont été construites, à savoir Baseline (formé pour classer les tweets comme positif, négatif et neutre) et Binaire (utilise un lexique de polarité et se classe comme positif et négatif. Les tweets neutres sont négligés). Les caractéristiques considérées par les classificateurs étaient les lemmes (noms, verbes, adjectifs et adverbes), les Lexiques de la Polarité et Multiword de différentes sources et Valence Shifters. [38]

3.3. Notre système d'analyse du sentiment

Pour notre système de fouille d'opinion on a choisi le réseau social Twitter. Twitter est actuellement la plate-forme de micro blog la plus populaire. Son premier slogan était Que faites-vous? néanmoins l'utilisation a pris une autre piste où les utilisateurs échangent des avis et des informations, le slogan devient Quoi de neuf ?. Plusieurs célébrités utilisent Twitter, on y trouve même des chefs d'Etat. Twitter limite le nombre de caractères utilisés dans un message, appelé tweet, à 280et qui peut contenir également des liens hypertextes. Les utilisateurs peuvent recevoir et envoyer des messages via le service SMS. [30]

Notre système se base sur l'approche basée corpus est précisément sous L'analyse sémantique latente LSA est une approche statistique qui est utilisé pour analyser les relations entre un ensemble de documents et les termes mentionnés dans ces documents pour produire un ensemble de modèles significatifs liés aux documents et termes. Cao et Duan ont utilisé LSA pour trouver les caractéristiques sémantiques des textes de revue à examiner l'impact des différentes fonctionnalités. L'objectif de leur travail est de comprendre pourquoi certaines critiques reçoivent beaucoup de serviabilité votes, tandis que d'autres reçoivent peu ou pas de voix du tout. Donc, au lieu de prévoir un niveau utile pour les examens qui n'ont pas votes, ils ont enquêté sur les facteurs qui déterminent le nombre des votes d'utilité qu'un examen particulier reçoit (inclure les votes «oui» et «non»). Ils ont travaillé sur des logiciels les commentaires des utilisateurs de CNET Download.com. Il a montré que les caractéristiques sémantiques sont plus influentes que d'autres caractéristiques en affectant combien de serviabilité avis de vote reçoivent. [19]

LSA repose essentiellement sous un corpus pour déterminer la relation sémantique entre des termes et des documents. Dans notre système le corpus est un ensemble de tweets de différents utilisateurs téléchargé depuis Twitter.

Ensuite les tweets sont prétraités et nettoyés pour la phase de calcul TF-IDF. Cette phase est essentielle pour éliminer les mot vide est déterminer les mots les plus fréquents porteurs d'opinion. Ensuite par l'algorithme LSA les opinions sont classé soit positif, négatif, ou neutre.

3.4 Les étapes nécessaires pour notre système à l'analyse du sentiment :

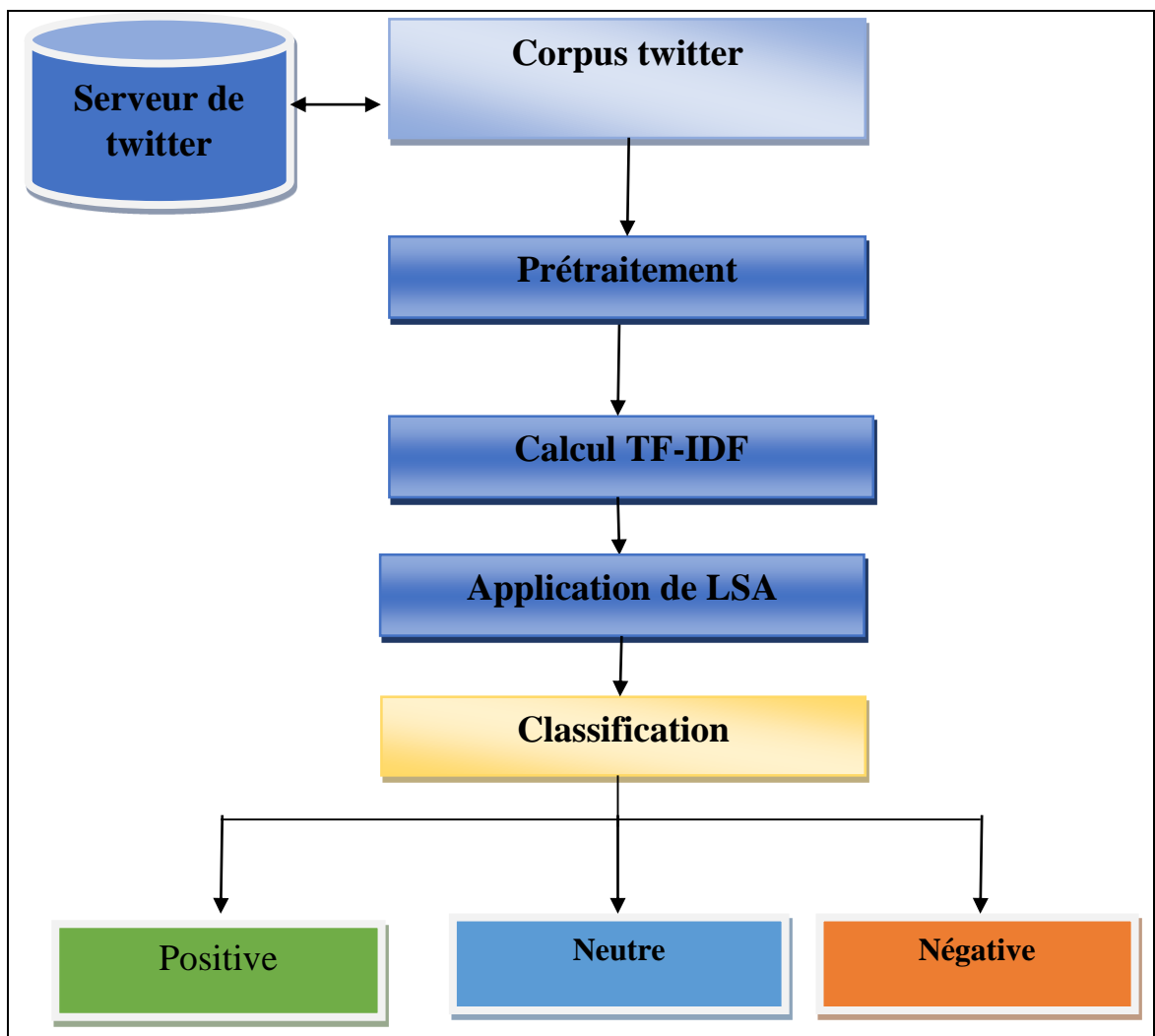


Figure 3.1 Architecture Analyses Sentiment

Dans ce qui suit nous allons présenter notre démarche en détailles :

3.5 Téléchargez les tweets

Connectez-vous à l'API Twitter, téléchargez les tweets d'un utilisateur au format CSV et comptez la fréquence des hashtags et des mots.

En utilisant le programme en ligne de commande `t` pour collecter les données Twitter au format CSV et l'outil `csvfix` pour le traiter, nous pouvons ensuite analyser le texte en utilisant les outils de texte standard Unix pour trouver des tendances intéressantes dans les tweets d'un utilisateur.

Le programme `t` est un wrapper simplifié pour l'API Twitter; Bien qu'il facilite la connexion à Twitter, les données qu'il renvoie ne sont qu'un sous-ensemble de ce que propose l'API de Twitter. Cela dit, suffisamment de données sont révélées dans le texte seul, et donc cet exercice sera principalement une pratique d'expressions `grep` et régulières.

3.6 Prétraitement des tweets

Un tweet contient beaucoup d'opinions sur les données qui sont exprimées de différentes façons par différents utilisateurs. L'ensemble de données de Twitter utilisé dans ce travail d'étude est déjà étiqueté en deux classes, à savoir. Polarité négative et positive et donc l'analyse du sentiment des données deviennent facile à observer l'effet de diverses fonctionnalités. Les données brutes ayant une polarité sont fortement sensibles à l'incohérence et à la redondance. Le prétraitement du tweet comprend les points suivants,

- ✓ Supprimez toutes les URL (par exemple `www.xyz.com`), les balises hash (par exemple, `#topic`), les cibles (`@username`).
- ✓ Corrigez les orthographes ; La séquence de caractères répétés doit être traitée.
- ✓ Remplacez toutes les émoticônes par leur sentiment.
- ✓ Supprimez toutes les ponctuations, symboles, numéros.
- ✓ Supprimer les mots d'arrêt.
- ✓ Expand Acronymes (on peut utiliser un dictionnaire d'acronymes).
- ✓ Supprimer les Tweets non anglais. [31]

3.7 Calcul TF- IDF :

Dans la recherche d'information, Tf-idf ou TF-IDF (de l'anglais : term frequency — inverse document frequency), abréviation de fréquence de document à fréquence

inverse, est une statistique numérique destinée à refléter l'importance d'un mot pour un document dans une collection ou un corpus. Il est souvent utilisé comme facteur de pondération dans les recherches d'extraction d'information, l'exploration de texte et la modélisation d'utilisateur. La valeur de Tf-idf augmente proportionnellement au nombre de fois qu'un mot apparaît dans le document et est compensée par la fréquence du mot dans le corpus, ce qui aide à ajuster le fait que certains mots apparaissent plus fréquemment en général. Tf-idf est l'un des systèmes de pondération de terme les plus populaires aujourd'hui; 83% des systèmes de recommandation textuels dans les bibliothèques numériques utilisent tf-idf.

Les variations du schéma de pondération de Tf-idf sont souvent utilisées par les moteurs de recherche comme un outil central de notation et de classement de la pertinence d'un document en fonction d'une requête de l'utilisateur. Tf-idf peut être utilisé avec succès pour le filtrage des mots d'arrêt dans divers domaines, y compris la synthèse de texte et la classification.

3.7.1 TF :

TF est l'abréviation de l'anglais term 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 :

$$\mathbf{TF(t) = (Nombre\ de\ fois\ que\ le\ terme\ t\ apparaît\ dans\ un\ document) / (Nombre\ total\ de\ termes\ dans\ le\ document).}$$

(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 term frequency factorise également en proportion de tous les mots utilisés dans le texte.

3.7.2 IDF :

L'IDF calcule le Inverse Document Frequency (la fréquence inverse du document) et complète l'analyse de l'évaluation du mot. Il agit en tant que correctif du 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 :

$$\mathbf{IDF (t) = \log_e (\text{Nombre total de documents} / \text{Nombre de documents contenant le terme t}).} \quad (2)$$

En conséquence, l'IDF détermine la pertinence d'un texte en considérant un mot-clé précis.

Par exemple, lorsqu'un document de **100** mots contient le terme «**chat**» **12** fois, le **TF** pour le mot «**chat**» est

$$\mathbf{TF \text{ chat} = 12/100} \quad \text{c'est-à-dire} \quad \mathbf{0,12}$$

L'IDF (fréquence de document inverse) d'un mot est la mesure de la signification de ce terme dans l'ensemble du corpus.

Par exemple, disons que le terme «**chat**» apparaît **x** fois dans un corpus de **10 000 000** millions de documents (c'est-à-dire web). Supposons qu'il y ait 0,3 million de documents qui contiennent le terme «**chat**», alors l'IDF (ie $\log \{DF\}$) est donnée par le nombre total de documents (**10 000 000**) divisé par le nombre de documents contenant le terme «**chat**» (**300 000**).

$$\mathbf{IDF (\text{chat}) = \log (10\,000\,000/300\,000) = 1,52}$$

$$\mathbf{\therefore W \text{ chat} = (TF * IDF) \text{ chat} = 0,12 * 1,52 = 0,182}$$

3.8. Construction la matrice des occurrences :

La LSA utilise une matrice qui décrit l'occurrence de certains termes dans les documents. C'est une matrice creuse dont les lignes correspondent aux « termes » et dont les colonnes correspondent aux « documents ».

Les « termes » sont généralement des mots tronqués ou ramenés à leur radical, issus de l'ensemble du corpus. On a donc le nombre d'apparition d'un mot dans chaque document, et pour tous les mots. Ce nombre est normalisé en utilisant la pondération tf-idf, combinaison de deux techniques : un coefficient de la matrice est d'autant plus grand qu'il apparaît beaucoup dans un document, et qu'il est rare — pour les mettre en avant.

Cette matrice est courante dans les modèles sémantiques standards, comme le modèle vectoriel, quoique sa forme matricielle ne soit pas systématique, étant donné qu'on ne se sert que rarement des propriétés mathématiques des matrices.

La LSA transforme la matrice des occurrences en une « relation » entre les termes et des « concepts », et une relation entre ces concepts et les documents. On peut donc relier des documents entre eux. Soit X la matrice où l'élément (i, j) décrit les occurrences du terme i dans le document j — par exemple la fréquence. Alors X

$$\mathbf{t}_i^T \rightarrow \begin{matrix} & \mathbf{d}_j \\ & \downarrow \\ \left(\begin{array}{ccc} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{array} \right) \end{matrix} \quad (3)$$

aura cette allure :

Une ligne de cette matrice est ainsi un vecteur qui correspond à un terme, et dont les composantes donnent sa présence (ou plutôt, son importance) dans chaque document :

$$\mathbf{t}_i^T = (x_{i,1} \quad \dots \quad x_{i,n}) \quad (4)$$

De même, une colonne de cette matrice est un vecteur qui correspond à un document, et dont les composantes sont l'importance dans son propre contenu de chaque terme.

$$\mathbf{d}_j = \begin{pmatrix} x_{1,j} \\ \vdots \\ x_{m,j} \end{pmatrix}$$

(5)

Partie 2

Présentation les outils de programmation et l'implémentation

3.1 Présentation de Java

Java est un langage de programmation à usage général, évolué et orienté objet dont la syntaxe est proche du C. Il existe 2 types de programmes en Java : les applets et les applications. Une application autonome (stand alone program) est une application qui s'exécute sous le contrôle direct du système d'exploitation. Une applet est une application qui est chargée par un navigateur et qui est exécutée sous le contrôle d'un plug in de ce dernier. [39]

3.1.1 Les caractéristiques java

- Java est interprétée
- Java est portable : il est indépendant de toute plate-forme
- Java est orienté objet.
- Java est simple
- Java est fortement typé
- Java est sûr
- Java est économe
- Java est multitâche

3.2 Environnement d'exécution :

NetBeans est un environnement de développement intégré (EDI), placé en open source par Sun en juin 2000, En plus de Java, NetBeans permet la prise en charge native de divers langages tels le C, le C++, le JavaScript, le XML, le Groovy, le PHP et le HTML

Compilé en Java, NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X ou sous une version indépendante des systèmes d'exploitation (requérant une machine virtuelle Java). Un environnement Java Développment Kit JDK est requis pour les développements en Java. [38]

3.3 Présentation d'application **Analysis_Sentement**

L'application est nommée **Analysis_Sentiment**, son rôle est de classer automatiquement des sentiments dans les réseaux sociaux par LSA

3.3.1 Fonctionnement de l'application

Cette application est composée de trois classes principales :

- Classe **LatentSemanticAnalysis** pour extraction les tweets.
- Classe **Tf-idf** pour le calcul
- Classe **LatentSemanticAnalysisTest** pour classier les tweets

3.3.1.1 Classe **LatentSemanticAnalysis** :

Cette classe est composée à une fonction principale :

- La fonction prétraitement () le rôle de cette fonction traitement et nettoyé texte :
 - ✓ Supprimez toutes les URL (par exemple www.twitter.com)
 - ✓ Supprimez toutes les émoticônes par leur sentiment.
 - ✓ Supprimez toutes les ponctuations, symboles, numéros.
 - ✓ Supprimez les mots d'arrêt.
 - ✓ Expand Acronymes (on peut utiliser un dictionnaire d'acronymes).
 - ✓ Supprimez les Tweets non latine

3.3.1.2 Classe **TF-Idf**

Est utilisé pour peser un mot-clé dans un contenu et attribuer l'importance à ce mot-clé en fonction du nombre de fois qu'il apparaît dans le document. Plus important encore, il vérifie la pertinence du mot clé sur l'ensemble du Web, appelé corpus.

```
public void addWords(String[] words) {
    String key = Arrays.toString(words);
    for (String word : words) {
        addWord(key, word);
    }
}

public double getTFidf(String sentence, String word) {
    double tf = getTF(sentence, word);
    double idf = getIDF(word);
    return tf * idf;
}

public double getTFidf(FreqCount<String> freq, String word) {
    double tf = getTF(freq, word);
    double idf = getIDF(word);
    return tf * idf;
}

public double getTF(String sentence, String word) {
    FreqCount<String> freq = sentenceMap.get(sentence);
    return getTF(freq, word);
}
```

Figure 3.2 : les fonctions de la class TF-IDF

3.3.1.3. Classe LatentSemanticAnalysisTest :

Cette classe est composée d'une fonction de classification () le rôle de cette fonction :

- ✓ Classification les tweets (positive, négative ou neutre)

3.4 Les étapes de l'exécution :

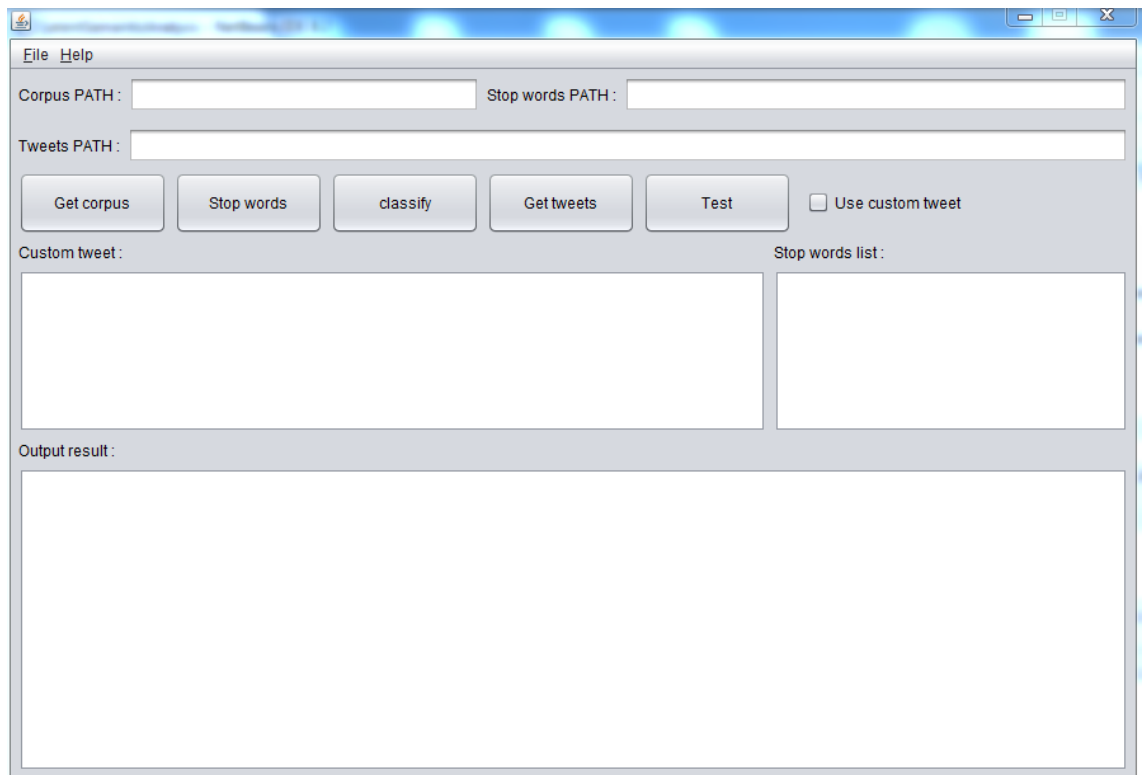


Figure 3.3: L'interface de l'application

- Cliquez sur le bouton **Get corpus** pour ouvrir le fichier qui contient le corpus de tweets traité.
- Cliquez sur le bouton **Stop words** pour ouvrir le fichier qui contient les mots vides.
- Cliquez sur le bouton **Classify** pour commencer à former l'algorithme LSA.

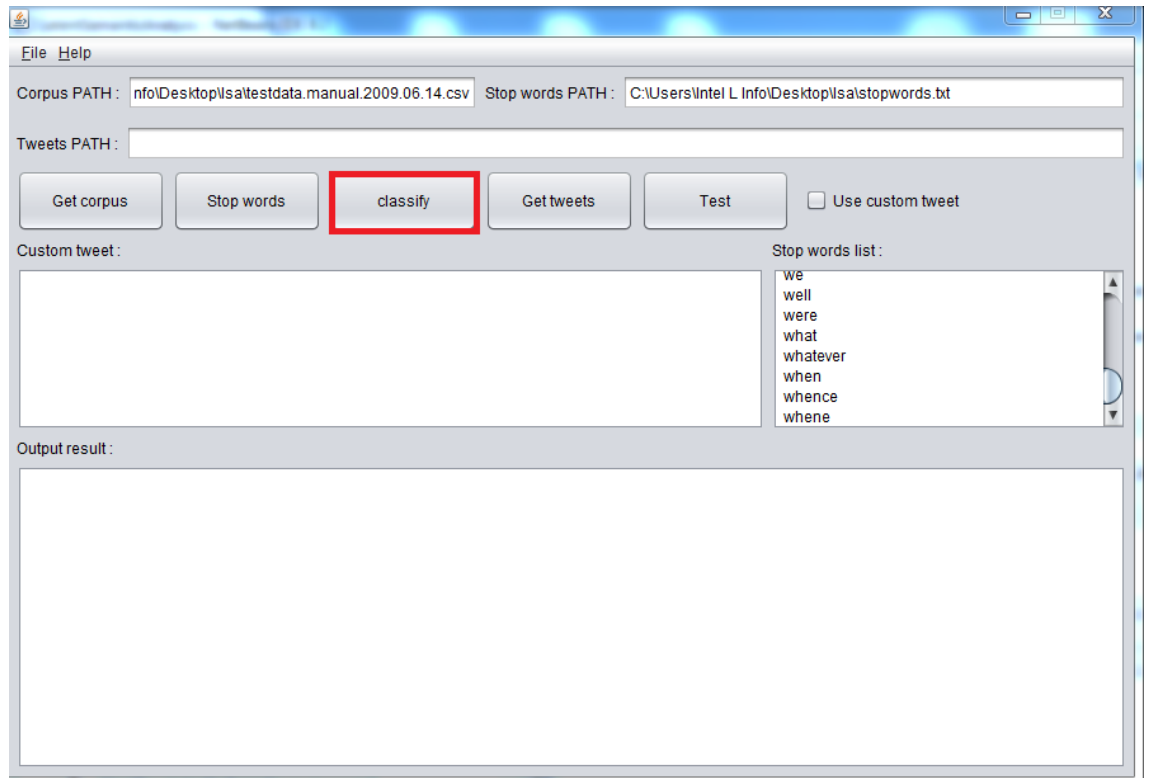


Figure 3.4: Training LSA

- Cliquez sur le bouton **Get tweets** pour ouvrir le fichier qui contient les tweets Ce que je veux classer.
- Cliquez sur le bouton **Test** pour faire la classification.

3.5 Résultat de l'exécution :

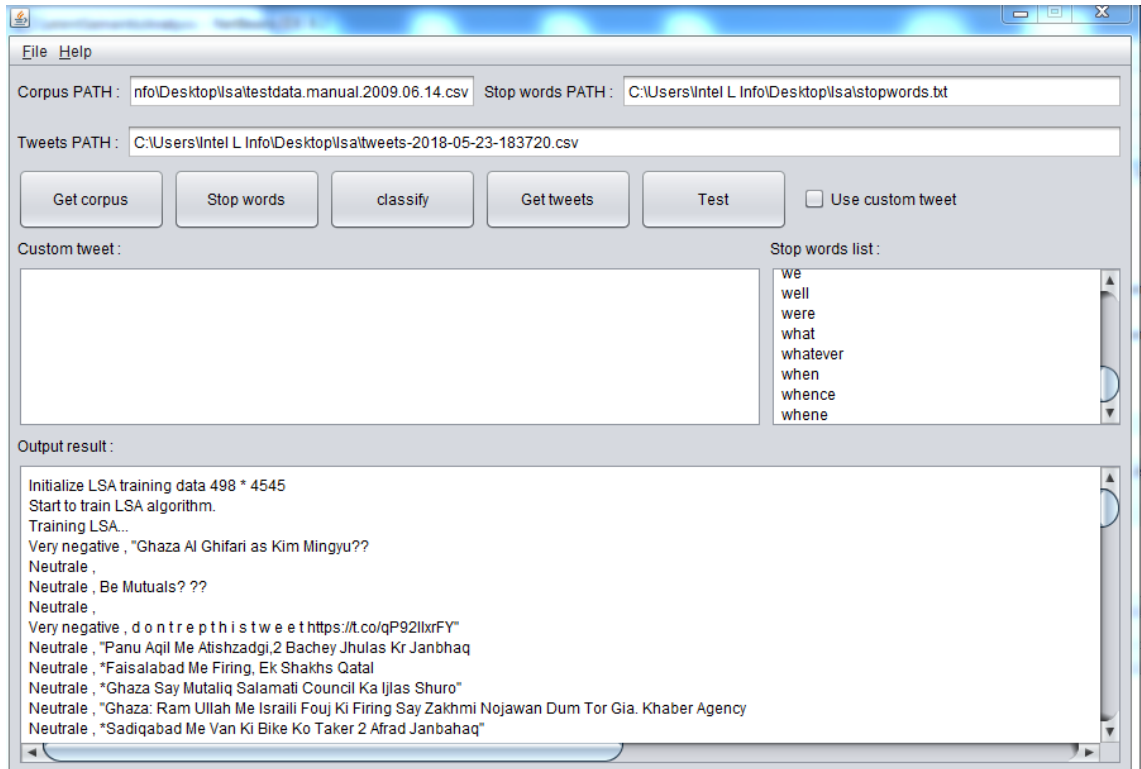


Figure 3.5: Résultat de test 1

Dans le cas où vous voulez écrire un commentaire et le classifier Nous faisons toutes les étapes **Get corpus** et **Stop words** et **Classify** puis Cliquez sur **Use custom tweets** ensuite **Test**

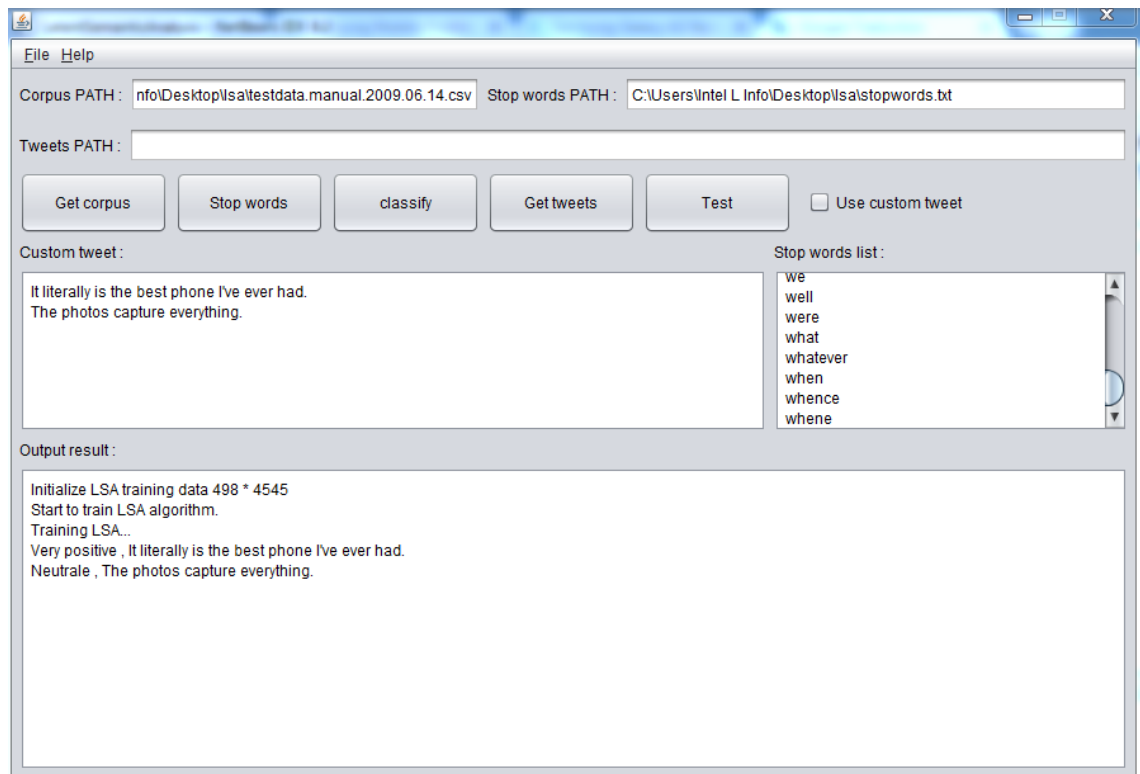


Figure 3.6: Résultat de test 2

3.6 Conclusion

Dans ce chapitre, nous avons présenté l'essentiel de notre travail qui consiste à créer un système d'analyse d'opinion pour détecter les sentiments dans les réseaux sociaux. Pour l'implémentation, nous avons utilisé l'une des méthodes de classification les plus connues; méthode LSA. Pour l'implémentation, nous avons choisi le réseau social Twitter en examinant les tweets pour les classer en: pour, contre ou neutre. Notre système s'intègre dans le domaine d'intelligence artificielle précisément "Machine Learning". Car la précision de la classification augmente à chaque fois quand exécute l'algorithme de classification.

Conclusion générale

L'analyse des sentiments et l'opinion Mining est un domaine émergent, dans ces dernières années plusieurs recherches s'intéressent à la tâche de l'analyse du sentiment, en particulier dans le domaine du micro-blogging.

Notre travail s'intègre dans ce même axe de recherche, nous proposons un système de classification subjective des opinions des utilisateur du réseau social twitter sur un produit ou un événement en trois catégories : positif, négatif et neutre.

Notre contribution consiste à intégrer la sémantique des langues naturels dans la phase de classification par l'emploi de la méthode de transformation du texte à base sémantique : Latent Semantic Analysis(LSA)

Notre système d'analyse de sentiment est encore en phase de développement et loin d'être complet. Pour l'amélioration, nous proposons l'utilisation des approches hybrides pour avoir des meilleurs résultats, et éventuellement élargir l'utilisation de ce travail vers d'autres objectifs comme l'analyse des tendances et l'extraction de connaissances à partir des réseaux sociaux

- [1] réseaux-sociaux-présentation L'Arobase, espace public numérique (EPN) de Pierrefitte-sur-Seine (Seine-Saint-Denis) .
- [2] Publié le 2 septembre 2012 par Azimo . <http://azimo.over-blog.fr/> .Consulté le 3 /6/2018
- [3] Marc Granovetter 24 Janvier 2016, 16:44pm . <http://nmstpe.over-blog.com>.Consulté le 25 /05/2018
- [4] www.alexa.com . Consulté le 25/05/2018 .
- [5] Facebook aligne deux milliards d'utilisateurs actifs. Quid des autres réseaux sociaux ? ,ITespresso.fr, 28 juin 2017 .
- [6] LinkedIn: Press Releases: LinkedIn Premium Services Finding Rapid Adoption .
- [7] GOOGLE+ 90 millions d'utilisateurs fantômes [archive], article sur <http://bigbrowser.blog.lemonde.fr> , daté du 28 février 2012 . Consulté le 25/05/2018 .
- [8] Instagram Direct, une messagerie pour envoyer des photos privés », sur Le Blog du Modérateur, 12 décembre 2013 .
- [9] Company Info : Twitter , daté du 5 mars 2017 .
- [10] Marc Granovetter Publié le 24 Janvier 2016, 16:44pm ,<http://nmstpe.over-blog.com> . Consulté le 25/05/2018 .
- [11] L'impact des réseaux sociaux sur notre vie Henri Michaud, rédacteur Canal Vie <http://www.canalvie.com> . Consulté le 25/05/2018 .
- [12] B.Bathelot, Publié le 2 février 2018 . www.definitions-marketing.com Consulté le 25/05/2018 .
- [13] • Liu, B. (2014) Sentiment analysis and Opinion Mining. Morgan & Claypool Publ.
- [14] "Full Circle Sentiment Analysis". Breakthrough Analysis. Publié le 2015-02-23.
- [15] Web mining , Publié par: Margaret Rouse , <https://searchcrm.techtarget.com> , Consulté le 02/06/2018 .
- [16] Introduction to Natural Language Processing (NLP), Publié le 11/08/2016 , <https://blog.algorithmia.com> ,Consulté le 02/06/2018 .

- [17] Eyrolles , Apprentissage automatique , Date de parution: 17/05/2018
- [18] Apprentissage automatique , <http://www.leparisien.fr/> , Consulté le 05/06/2018
- [19] Walaa Medhat , Ahmed Hassan , Hoda Korashy , Sentiment analysis algorithms and applications: A survey , Ain Shams Engineering Journal (2014) 5, 1093–1113
- [20] <https://werfamous.com/fr/> Consulté le 05/06/2018 .
- [21] <https://finnaarupnielsen.wordpress.com/> Consulté le 05/06/2018 .
- [22] <http://www.wjh.harvard.edu> Consulté le 05/06/2018 .
- [23] <http://www.sentic.net/> Consulté le 05/06/2018 .
- [24] <https://wordnet.princeton.edu/> Consulté le 05/06/2018 .
- [25] <http://www.sentiwordnet.isti.cnr.it/> Consulté le 05/06/2018 .
- [26] sentiment140. URL : <http://www.sentiment140.com> . Consulté le 05/06/2018 .
- [27] <http://www.tweetfeel.com> . Consulté le 05/06/2018 .
- [28] <http://twitrratr.com> . Consulté le 05/06/2018 .
- [29] <http://smm.streamcrab.com> . Consulté le 05/06/2018 .
- [30] These:Analyse des données de microblogs ,présenté par Housseem Eddine Dridi ,Université de Montréal .
- [31] These : Detection de racisme dans les réseaux sociaux ,présenté par Djalal Attallah, Université de M'sila .
- [32] www.growthaccelerationpartners.com Consulté le 05/05/2018 .
- [33] Neviarouskaya Alena, Prendinger Helmut, Ishizuka Mitsuru. Recognition of Affect, Judgment, and Appreciation in Text. In:Proceedings of the 23rd international conference on computational linguistics (Coling 2010), Beijing; 2010. p. 806–14.
- [34] Cao Qing, Duan Wenjing, Gan Qiwei. Exploring determinants of voting for the“helpfulness” of online user reviews: a text mining approach. Decis Support Syst 2011;50:511–21.

- [35] Chin Chen Chien, Tseng You-De. Quality evaluation of product reviews using an information quality framework. *Decis Support Syst* 2011;50:755–68.
- [36] Moreo A, Romero M, Castro JL, Zurita JM. Lexicon-based ,comments-oriented news sentiment analyzer system. *Expert Syst Appl* 2012;39:9166–80.
- [37] Chenlo J, Hogenboom A, Losada D. Sentiment-based ranking of blog posts using rhetorical structure theory. In: Presented at the 18th international conference on applications of Natural Language to Information Systems (NLDB'13); 2013.
- [38] <https://www.techno-science.net> consulté le 10/06/2018.
- [39] Jean Michel DOUDOUX , Développons en Java

الملخص:

ذا العمل يندرج ضمن مجال تحليل المشاعر والآراء، والغرض منه هو تصنيف مشاعر وآراء مستخدمي الشبكة الاجتماعية تويتر حول منتج أو حدث ما الى ثلاث فئات: إيجابية، سلبية ومحيدة. يأخذ نظام تصنيف المشاعر في الاعتبار دلالات اللغات، وهذا المعيار مكفول من خلال استخدام طريقة التحليل الكامن LSA

الكلمات المفتاحية: تحليل المشاعر والآراء، الشبكات الاجتماعية، تويتر، LSA

Résumé :

Ce travail s'intègre dans le domaine d'analyse des sentiments et l'opinion Mining, son objectif est de classer les sentiments et les avis des utilisateurs du réseau social twitter vers une cause, un événement ou un produit en trois catégories : positif, négatif et neutre. Notre système de classification de sentiment prend en compte la sémantique des langues naturelles, ce critère est assuré par l'emploi de la méthode Latent Semantic Analysis.

Mots clés : analyse des sentiments et l'opinion Mining, réseaux sociaux, Twitter, LSA.

Abstract :

This work fits into the field of sentiment analysis and opinion Mining, its purpose is to rank the feelings and opinions of users of the social network twitter to a cause, an event or a product in three categories: positive, negative and neutral. Our sentiment classification system takes into account the semantics of natural languages, this criterion is ensured by the use of the Latent Semantic Analysis method.

Key words: sentiment analysis and opinion Mining, social networks, Twitter, LSA.