



Rakmana

Revue d'études en sciences de l'information et de la communication

Volume 03 | Numéro 02
Juin 2023
Pages 186 - 200

ISSN-2773-4285
EISSN| 2830-8417
Dépôt légal|07/2021
Adresse|11, Route Dodo
Mokhtar, Ben Aknoun, Alger
Fax (023) 23 88 50|
Téléphone(0561) 62 29 75|

Reçu le 12/05/2023
Accepté le 16/05/2023
Publié le 27/06/2023

Auteur correspondant|
Ziani Ghaouti
Université Kasdi Merbah
Ouargla (Algérie)
ziani.ghaouti@univ-
ouargla.dz



Analyse des sentiments dans les médias en ligne Une approche Machine Learning

Sentiment analysis in online média A Machine Learning approach

Ziani Ghaouti
Université Kasdi Merbah Ouargla (Algérie)
ziani.ghaouti@univ-ouargla.dz

Abderrahim Mohammed El Amine
Université Kasdi Merbah Ouargla (Algérie)
abderrahim.mohammedlamine@univ-ouargla.dz

Résumé

Les médias en ligne offrent à leurs lecteurs la possibilité de donner leurs avis sur les thèmes publiés (commentaires). A cet effet, des outils de communication sont associés à chaque article mis en ligne. Ces commentaires sont pris au sérieux par les médias car ils contiennent des informations précieuses pour connaître les sentiments des internautes. Les réseaux sociaux génèrent des quantités gigantesques de données non structurées, du fait de leur utilisation par des milliers d'utilisateurs. Ainsi, notre étude a pour objectif de justifier et d'utiliser l'approche de Machine Learning (l'intelligence artificielle) pour analyser les sentiments des utilisateurs à travers les commentaires récoltés sur YouTube concernant des conférences de presse menées par le président de la république. Les résultats obtenus montrent la faisabilité de l'approche et suggèrent d'autres pistes pour son amélioration.

Mots-clés: Média en ligne, Intelligence artificielle, Machine Learning, Réseaux sociaux.

Abstract

Online media offer their readers the opportunity to give their opinions on published topics (comments). To this end, communication tools are associated with each article posted online. These comments are taken seriously by the media because they contain valuable information to know the sentiment of Internet users. Social networks generate gigantic amounts of data, due to their use by thousands of users. Thus, our study aims to justify and use the Machine Learning approach (Artificial intelligence) to analyze the sentiments of users through the comment collected on YouTube concerning press conferences led by the president of the republic. The results obtained show the feasibility of the approach and suggest other directions for its improvement..

Keywords: Online media, Artificial intelligence, Machine Learning, Social networks.

1. Introduction

Inventé en 1955 par John McCarthy, l'intelligence artificielle (IA) est considérée comme le domaine de la science et de l'ingénierie qui traite de la compréhension, à l'aide de l'ordinateur, du comportement intelligent et de la création de systèmes artificiels qui reproduisent ce comportement (Shapiro, 1992, cité dans Zouinar, 2020 :3). Ce qui signifie qu'il s'agit d'une simulation de l'intelligence humaine. L'IA est souvent liée au progrès humain et à la modernité (Stahl et al, 2021). L'IA imprègne plusieurs domaines, parmi lesquels celui du journalisme que ce soit dans l'écriture, l'enquête ou encore la vérification d'information. L'IA intervient dans l'analyse des commentaires postés sur un site web par des internautes sur un sujet précis. Ainsi, l'IA « s'impose également dans la relation qui unit un site d'information en ligne à ses destinataires et dans l'interprétation de ses modes de consultations et pratiques de lecture » (Béasse, 2022 :3). Dans le domaine du journalisme, les médias en ligne offrent à leurs lecteurs la possibilité de donner leurs avis sur les thèmes publiés. A cet effet, des outils de communication (forum de discussion, réseaux sociaux,...) sont associés à chaque article mis en ligne. Ces commentaires sont pris au sérieux par les médias car ils contiennent des informations précieuses dont l'analyse peut contribuer à la détermination des sentiments des internautes sur un thème mis en ligne. L'IA consiste à analyser des contenus textuels issus des échanges en ligne (Jain, Pamula, & Srivastava, 2021 ;), ou bien à comprendre les sentiments qui prévalent au sein des marchés en analysant des ensembles de données alternatives tels que des articles et des communiqués de presse, des publications de résultats trimestriels (Brancart, 2020 :22,23). Les outils de communication comme les réseaux sociaux génèrent des quantités gigantesques de données, du fait de leur utilisation par des milliers d'utilisateurs. Ces données à traiter sont de grande dimension et potentiellement non structurées. Pour analyser ces grandes quantités de données non structurées (commentaires postés sur des réseaux sociaux), il est très difficile d'utiliser les techniques traditionnelles (Suchdev et al, 2014 :36) comme celles des sciences de l'information et de la communication (SIC). Notre problématique relève donc exclusivement de l'analyse des matériaux textuels (commentaires) échangés entre lecteurs de médias en ligne. Notre objectif est d'apporter les éléments de réponse à la question suivante : pourquoi utiliser l'approche de machine Learning ? et comment l'utiliser dans l'analyse des sentiments d'internautes ?

Notre étude s'ancre dans une posture épistémologique dite interdisciplinaire qui consiste à importer un outil d'analyse de donnée de l'IA. Ensuite, nous aborderons la relation entre SIC et IA, puis nous définirons le ML en argumentant sur les raisons de son utilisation. Enfin, nous terminerons par l'application de ML dans l'analyse des sentiments dans le contexte de médias en ligne (cas du discours du président de la république sur YouTube).

2. Sciences de l'information et de la communication et IA

La relation entre les SIC et l'IA est complexe et se caractérise de trois façons. La première est que les deux disciplines ont pour socle commun l'information. Ainsi, les réseaux sociaux comme Facebook, Twitter, Youtube sont considérés comme des médias.

Selon Yves Jeanneret (2011 :63), le média « ne transmet pas de l'information mais crée des espaces de signification, au sein d'une culture de signes ». Ainsi, selon le même auteur, la notion de média renvoie à deux types d'informations : l'information au sens mathématique du terme (série de caractères, erratiques aux yeux d'un lecteur et font partie de la procédure technique) qui contribue à la transmission d'un certain nombre de propriétés du message et ne peut être interprétée et faire sens. Le deuxième type de l'information, constitue le message ou l'information sociale produite par l'utilisateur de média. En SIC, le terme information a notamment été fondé sur les données. Dans les champs qui traitent les données et l'information comme des entités, l'information se transcrit en donnée et signifiante. Selon cette définition, les données composent la « matière » de l'information (Floridi, 2005, cité dans Leleu-Merviel, et Useille, 2008 :36). Où l'information joue un rôle crucial dans la représentation de la réalité. Ainsi, l'opinion ou le point de vue désigne une attitude, une pensée ou un jugement inspiré par le sentiment. L'émotion est un sentiment raffiné : sensibilité délicate particulièrement exprimée dans une œuvre d'art, un sentiment romantique ou nostalgique frisant la sentimentalité (Pécore, 2019 :9). Pour Alain Kiyindou (2019 : 91), du fait que les SIC ont pour objet l'étude des propriétés générales de l'information (production, nature, circulation acteurs, effets...), il existe un lien tout tracé avec l'IA qui est aussi une technologie de l'information, qui effectue le traitement et la diffusion de l'information.

La deuxième façon, l'IA se préoccupe de certaines questions des SIC comme la manière d'informer et de communiquer, l'analyse de données publiées sur des réseaux sociaux, la recherche d'information, la création de contenu, etc. En ce sens, l'IA peut être considérée comme une nouvelle méthodologie en SIC.

La troisième façon, les SIC peuvent étudier l'usage social de l'IA et son rôle dans la société, et ainsi contribuer au développement des connaissances scientifiques, ses contributions dans la société, et l'éthique de son utilisation, et des méthodologies de conception de design d'interfaces homme-machine. Il s'agit d'un partenariat scientifique équilibré à construire entre les deux disciplines pour reprendre l'expression Laurent Petit (2021:7).

En IA, pour le traitement de toute information, on doit recourir à des ensembles massifs de données multi-variées et accumulées à des vitesses sans précédent sur les individus, et même sur les organisations, via le web, et les réseaux sociaux et qui sont appelés : Big Data (Teboul, et Amri, 2014 :2). Ainsi, comme le souligne Alexandre Gefen (2022) : « L'importance prise par l'IA impose de la comprendre non seulement comme un ensemble d'algorithmes, mais aussi comme une « technoculture », où les sciences impliquées (mathématiques, sciences de l'information et sciences cognitives) sont inscrites dans des institutions configurant les pratiques scientifiques, mais aussi dans des contextes culturels situés et dans des imaginaires impliquant des idéologies, des fictions, des représentations sensibles. ». Selon cet auteur, nous comprenons que l'IA n'est pas un domaine isolé, du fait qu'il mette en relation plusieurs disciplines telles que, l'informatique, les sciences cognitives, les neurosciences, les sciences humaines et sociales, la linguistique...

3. Qu'est ce que le Machine Learning?

Le ML est un sous-domaine de l'IA. Cette technique est utilisée pour le développement de méthodes et d'algorithmes permettant aux machines d'apprendre des données sans être explicitement programmées. Par ailleurs, elles sont capables de s'adapter et de s'améliorer automatiquement à partir de l'expérience, sans intervention humaine. L'apprentissage automatique ou ML « est la science (et l'art) de programmer les ordinateurs de sorte qu'ils puissent apprendre à partir de données »(Géron, 2019 :2). Il s'agit d'un ensemble de méthodes permettant d'établir, à partir de données, des modèles de prises de décisions, de prédictions ou de classifications (Mazieres, 2016). Cette nécessité est due au fait que « les sciences humaines et sociales ont autant à apprendre de l'IA que l'inverse. C'est donc un dialogue paritaire établi sur un pied d'égalité qu'il convient d'instaurer. » (Petit , 2021 : 8). Ainsi, pour réussir l'analyse des faits humains (messages postés sur Facebook) par les sciences du numérique (IA), il s'agit d'humaniser le numérique, au sens d'y introduire le recul et l'approche critique propre aux SHS (Ibid, 2021 :8). Le ML est utilisé dans de nombreux domaines, notamment : la reconnaissance de la parole, la vision par ordinateur, la traduction automatique, la classification et la prédiction de données, la recommandation de produits ou de contenus, la détection de fraudes, etc.

4. Pourquoi utiliser le Machine Learning?

Pendant les dix dernières années, le ML a envahi les systèmes informatiques des entreprises. Actuellement, sur les grandes firmes comme Google, ou d'autres sites web, plusieurs systèmes d'apprentissages s'activent à l'instant même où une recherche est lancée, ainsi certains s'assurent que vous n'êtes pas un robot (vous n'êtes pas un fraudeur ou hacker), d'autres analysent votre profil et vous proposent des publicités ou des contacts comme Facebook. Le ML intervient dans une multitude de domaines. Si les entreprises comme celles des médias en ligne n'utilisent pas le ML, elles risquent de ne pas être en phase avec le développement du domaine du journalisme (Mittal et Gujral, 2020 : 2789).

Les études effectuées sur le ML mettent en avant les arguments justifiant son utilisation comme suit :

a. Non structuration des données traitées

Le traitement des données extraites des réseaux sociaux est devenu très difficile à cause de la non structuration de ces données volumineuses (big data), et de leur subjectivité et hétérogénéité. Des techniques de ML ont été utilisées pour gérer cette difficulté dans des applications réelles (Jain, Pamula, & Srivastava, 2021 ; Brynjolfsson et McAfee, 2017 :3). Les données volumineuses et non structurées, appelées aussi big data, ont des caractéristiques bien définies : plus importantes que celles des bases de données habituelles, variées avec différents types et formats différents, arrivant à des rythmes très rapides...Les méthodes traditionnelles pour juger de la qualité, la pertinence et l'exactitude des données ne s'appliquent pas, sachant que la valeur de ces données est précieuse (Brancart, 220 :57).

b. Machine Learning vs programmation traditionnelle

L'application de la programmation traditionnelle dans le cas de l'analyse des sentiments (commentaires textuels) impose au programmeur de dresser une liste exhaustive de tous les cas possibles et envisageables dans les documents analysés. A cet effet, cet algorithme nécessite à chaque fois la mise à jour et l'ajout de tests sur les mots nouvellement détectés dans les commentaires et de ce fait il va contenir des milliers d'instructions (règles) complexes et difficiles à maintenir pour reprendre l'expression d'Aurélien Géron (2019 :3).

Par contre le ML, représente une approche fondamentalement différente à celle de l'approche traditionnelle : la machine apprend à partir d'exemples, plutôt que d'être explicitement programmée pour un résultat particulier (Brynjolfsson et McAfee, 2017 :3). L'analyseur des sentiments basé sur des techniques ML permet d'avoir des informations à l'avance sur les attitudes des lecteurs de médias en ligne et il apprend automatiquement quels sont les mots et les phrases qui constituent des bons prédicteurs.

5. Application de Machine Learning à l'analyse des commentaires dans les médias en ligne: cas de discours du président de la république

Dans un contexte de journalisme, nous avons choisi de mener une analyse des sentiments à travers les commentaires récoltés sur YouTube concernant des conférences de presse menées par le président de la république. L'objectif de cette analyse est d'obtenir des informations sur les opinions des internautes en lien avec le discours du président de la république en utilisant l'approche de ML, qui consiste à construire un modèle de l'analyse des sentiments. Ce modèle devra apprendre à partir de ces données récoltées (corpus de données d'apprentissage) et être capable de prédire les sentiments des internautes, c'est à dire classer les données invisibles en fonction du modèle formé pour reprendre l'expression de Khairnar Jayashri, & Kinikar Mayura (2013) . En se basant sur la méthode de développement d'un projet de ML proposée par Aurélien Géron (2019 : 283) et après l'avoir adaptée à nos besoins, nous avons retenu les étapes suivantes:

a. Cerner le problème

La construction d'un modèle de prédiction du sentiment en vers le discours du président de la république, servira d'une part, à démontrer que l'approche de ML est efficace pour les quantités gigantesque de données et peut être considérée comme une nouvelle méthodologie en SIC, d'autre part, ce modèle servira à alimenter un autre système de ML comme celui des élections présidentielles. Ce système utilisera l'information de prediction en entrée et d'autres informations pour déterminer s'il est utile pour le pronostic des résultats des élections présidentielles.

b. Extraction des données

Dans cette étape nous avons sélectionné les vidéos en relation avec les conférences de presse du président de la république et celles contenant un grand nombre d'interactions des usagers. L'extraction des commentaires se fait à partir du site YouTube par l'utilisation de

Python (langage de programmation le plus utilisé dans le domaine de l'IA) et de l'API Youtube (API : Application Programming Interface) : « googleapiclient.discovery ». Cette API est une interface de programmation d'application qui permet de «connecter» un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités (Gheorghin, 2022). La commande Python : « from googleapiclient.discovery import build » permet d'importer cette bibliothèque afin d'interroger et extraire les commentaires à partir du site YouTube. L'ensemble de données d'analyse se compose d'un total de 600 commentaires. Les commentaires sont enregistrés les uns à la suite des autres dans un même fichier Excel.

Après élimination des commentaires en double et les commentaires vides, nous avons obtenu 500 commentaires différents.

La tableau 1 présente des exemples de commentaires (voir Tableau 1) :

Tableau 1. Exemples de commentaires

Commentaires
"من ابن الدولة المسماة العميقة الابن المعارض لسياسة"
"أين نحن من هاده الوجود امام الله و الرسول والشعب"
"احبيك من اعماق قلبي نتا راجل وفحل ربي يوفتك لرئاسة الجمهورية"
"من الجنوب الجزائري الكبير رانا مع عبد المجيد تبون"

c. Annotation des commentaires

Pour chaque commentaire, nous avons vérifié manuellement que les commentaires récupérés étaient en lien avec le discours du président de la république. L'ensemble de données est étiqueté (classé) et, par conséquent, nous savons exactement quels commentaires exprimant une attitude positive (3), négative(2), neutre(1), et non pertinente(0). La tableau 2 présente les exemples de commentaires de la tableau 1 avec leurs annotations (voir Tableau 2) :

Tableau 2. Exemples de commentaires avec leurs annotations

Commentaires	Annotations
"من ابن الدولة المسماة العميقة الابن المعارض لسياسة"	2 (Négatif)
"أين نحن من هاده الوجود امام الله و الرسول والشعب"	2 (Négatif)
"احبيك من اعماق قلبي نتا راجل وفحل ربي يوفتك لرئاسة الجمهورية"	3 (Positif)
"من الجنوب الجزائري الكبير رانا مع عبد المجيد تبون"	3 (Positif)

En effectuant l'annotation de l'ensemble des commentaires, les distributions obtenues des différentes classes sont présentées dans la tableau 3 (voir tableau 3).

Tableau 3. Les distributions des commentaires annotés

Etiquette	Signification	Nombre de commentaires
0	Non pertinent	60
1	Neutre	28
2	Négatif	177
3	Positif	235

Du tableau 3 nous pouvons constater que le nombre de commentaires concernant les deux étiquettes « 0 » et « 1 » est très réduit devant les autres étiquettes, et par conséquent nous avons décidé de ne retenir que les commentaires ayant les étiquettes « 2 » et « 3 » pour réaliser notre expérimentation. Autrement notre experimentation proterait sur une classification binaire et nous cherchons à construire un classifieur binaire pour prédire la polarité d'un commentaire (positif, négatif) à propos des conférences de presse menées par le président de la république.

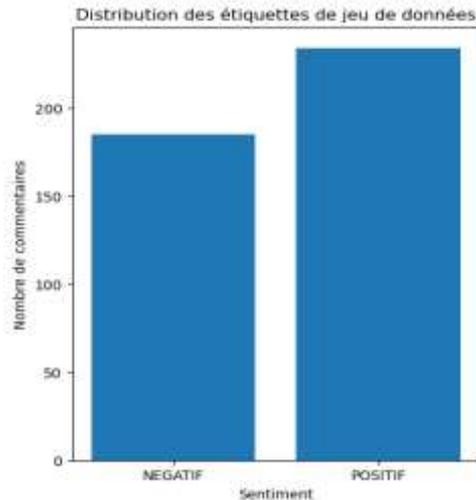
d. Prétraitement des commentaires

Le prétraitement est réalisé automatiquement à l'aide de la bibliothèque de programmes «NLTK », cette dernière est disponible gratuitement sur le net : «NLTK (Natural Language ToolKit), librairie de référence dans le domaine du NLP (Natural Language Processing), et qui est une bibliothèque logicielle en langage de programmation Python (utilisé dans le domaine de l'IA), permettant un traitement automatique du langage naturel (Hardeniya et al, 2016:4). Les différents prétraitements que nous avons opérés sont :

- Eliminer la ponctuation,
- supprimer les caractères spéciaux et laisser seulement les lettres,
- supprimer les signes diacritiques,
- supprimer les URLs,
- supprimer les nombres,
- supprimer les espaces supplémentaires,
- supprimer les mots outils ou les mots qui ne sont pas porteurs de sens (stop words),
- supprimer les emojis,
- supprimer les commentaires ne comportant qu'un mot (taille du commentaire=1)
- Tokenisation (découper le texte en unités fondamentales : les tokens)

A la suite de cette phase nous avons obtenu un corpus nettoyé annoté de taille 412 commentaires (235 Positif et 177 négatif). La figure 1 présente la répartition des étiquettes de notre corpus.

Figure 1. Distribution des étiquettes de jeu de données



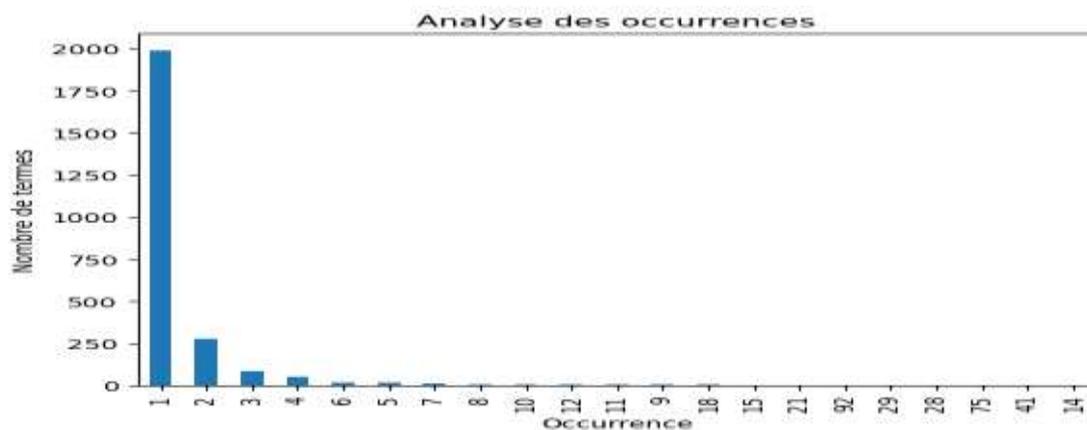
Ce corpus comprend un vocabulaire de 2479 termes différents. L'exemple suivant représente les 30 termes les plus fréquents de notre corpus (chaque terme est accompagné de sa fréquence):

(75, 'تبون'), (29, 'الجزائر'), (28, 'رئيس'), (21, 'الشعب'), (18, 'البلاد'), (18, 'العصابة'), (15, 'عبد'), (15, 'رجل'), (15, 'المجيد'), (12, 'السيد'), (11, 'يكون'), (11, 'تحيا'), (11, 'بيوفك'), (11, 'حناشي'), (10, 'وزير'), (10, 'خير'), (10, 'الرجل'), (8, 'البلاد'), (8, 'راك'), (8, 'التاريخ'), (8, 'سي'), (8, 'الخير'), (8, 'القايد'), (7, 'بوتفليقة'), (7, 'صادق'), (7, 'الحمد'), (7, 'انشاء'), (7, 'سيدي'), (7, 'الدولة'), (6, 'صالح'), (6, 'الرجل')

Nous pouvons constater que les termes (28, 'رئيس'), (29, 'الجزائر'), (75, 'تبون') sont les plus cités. Ce qui peut nous donner une première information concernant le contenu des commentaires analysés.

L'analyse fréquentielle nous a permis aussi de remarquer que la majorité des termes ne sont répétés qu'une seule fois. La figure 2 présente la distribution des occurrences dans notre corpus.

Figure 2. La distribution des occurrences dans les commentaires



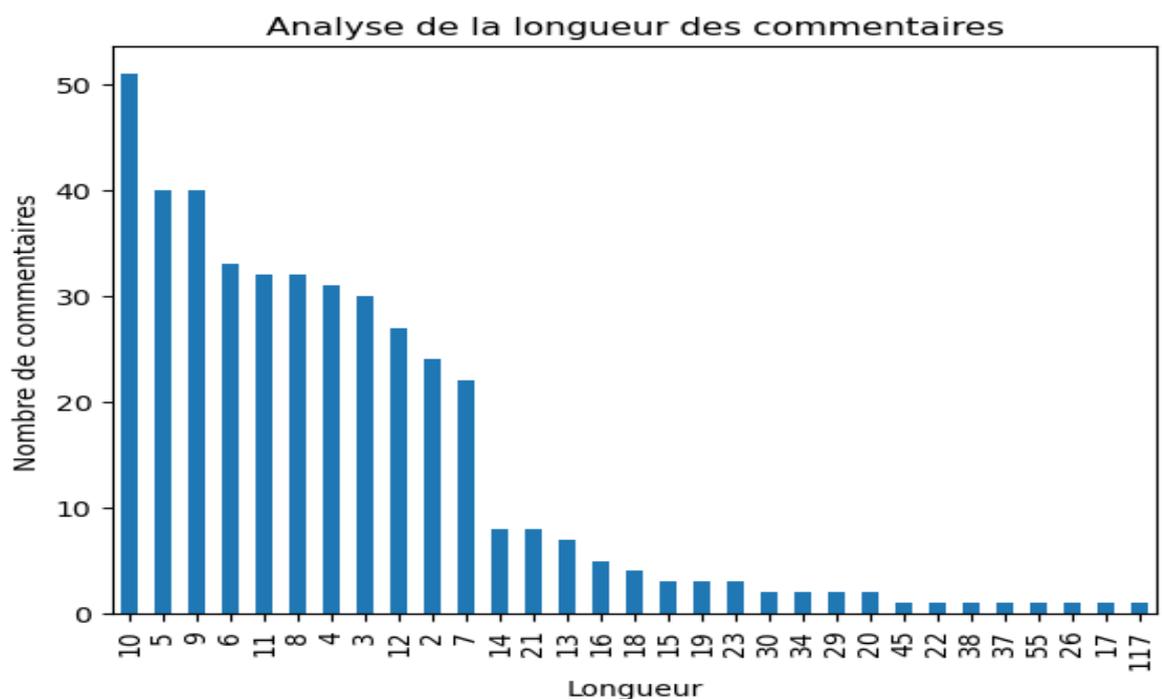
Le tableau 4 présente une indication sur la taille des commentaires en nombre de termes et de caractères.

Tableau 4. Taille des termes et des caractères dans les commentaires

Types de commentaires	Taille
Le plus long commentaire en nombre de termes	117
Le plus court commentaire en nombre de termes	2
Le plus long commentaire en nombre de caractères	697
Le plus court commentaire en nombre de caractères	9

La figure 3 montre la répartition des longueurs en fonction du nombre des commentaires. De cette figure nous pouvons constater que la majorité des commentaires sont courts avec une taille qui varie entre 2 et 12 termes. Pour les autres tailles (entre 13 et 117) le nombre de commentaires est toujours inférieur à 10. Autrement peu de commentaires, ne dépassant pas 10, ont une taille supérieure à 13 termes.

Figure 3. Répartition des longueurs en fonction du nombre de commentaires



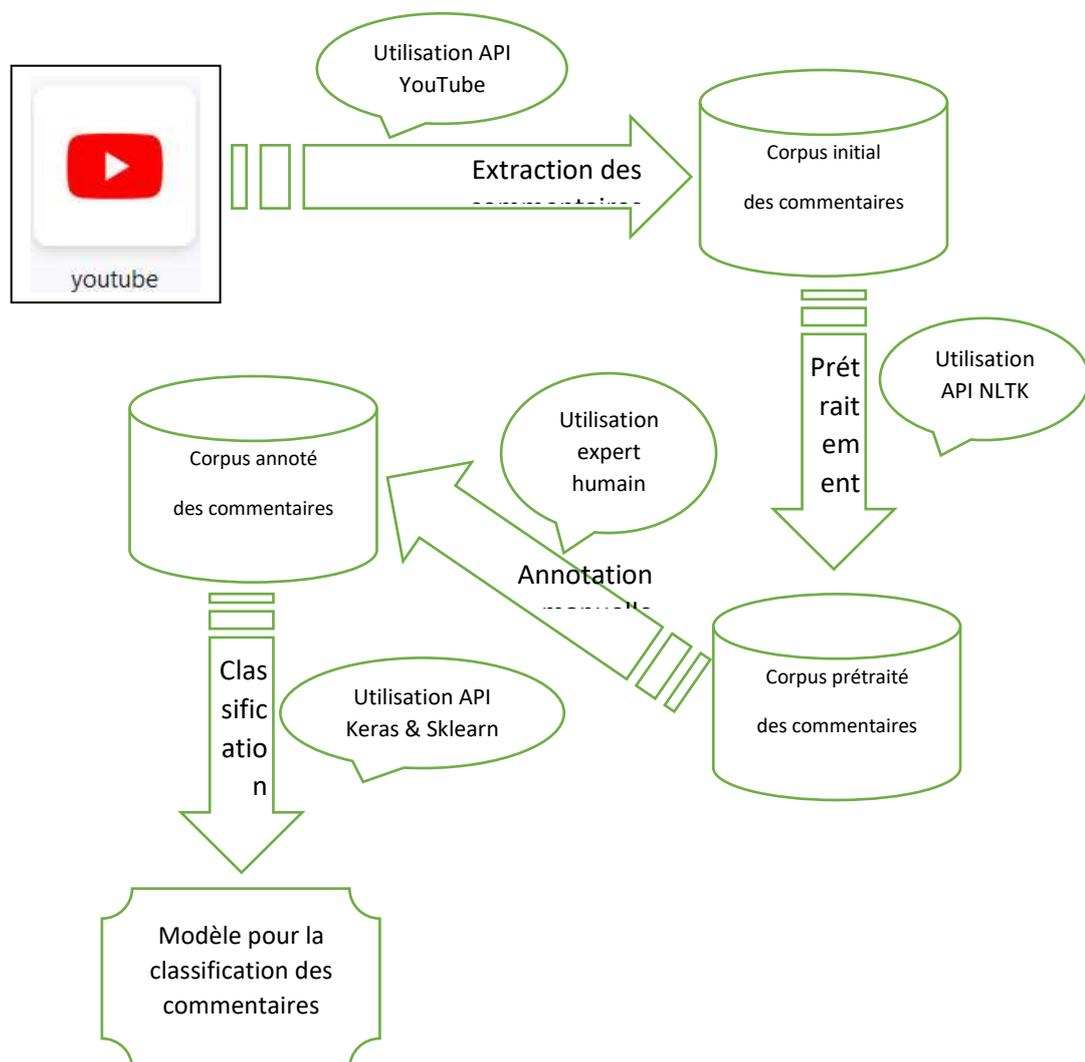
e. Classification et prédiction

Dans le cadre du ML, nous proposons, dans le contexte de cette communication, de tester l'approche supervisée (du moment que nous avons un corpus annoté). A cet effet plusieurs classifieurs (algorithmes de classement) existent (K plus proches voisins, Machines à support de vecteurs (ou SVM), Naïve Bayes, Arbres de décisions, Réseaux de neurones...). Pour notre part, nous avons testé plusieurs classifieurs afin de déterminer le meilleur. Pour ce faire nous avons utilisé deux bibliothèques de programmes Python à savoir : « keras », « sklearn ».

« Keras » est une bibliothèque open source pour la manipulation des réseaux de neurones et qui a pour objectif de permettre des expérimentations rapides sur les jeux de données (Atakpa, 2018 :20). « Sklearn » est une autre bibliothèque de ML écrite en Python. Elle est conçue pour fournir un recueil d'implémentation efficace d'algorithmes d'apprentissage automatiques (Géron, 2019 :62).

La figure 4 montre les étapes réalisées pour obtenir le modèle de la classification des commentaires.

Figure 4. Les étapes de la classification des commentaires



Pour la tâche de classification on doit suivre la démarche suivante :

- a) A partir de notre corpus, prétraité annoté, on génère trois sous corpus : un pour l'apprentissage (Train), un pour le test (Test) et un autre pour la validation (Validation),
- b) Définition du modèle (architecture du système intelligent ou classifieur)
- c) Encodages des commentaires,
- d) Préparation du vecteur des sentiments
- e) Evaluation du modèle.

Cette démarche est répétée plusieurs fois jusqu'à obtention du modèle testé et validé le plus pertinent. La figure 5 présente la démarche de classification. Par exemple, pour un modèle de réseaux de neurones, nous utilisons une seule couche cachée avec 50 neurones. La couche de sortie est un seul neurone pour prédire « 2 » pour les avis négatifs et « 3 » pour les avis positifs. Le réseau est entraîné par des paramètres adaptés aux problèmes de classification binaire. Nous garderons une trace de la précision lors de la formation et de l'évaluation du modèle. La figure 6 montre le modèle décrit de l'exemple précédent.

Figure 5. Démarche pour la classification des commentaires

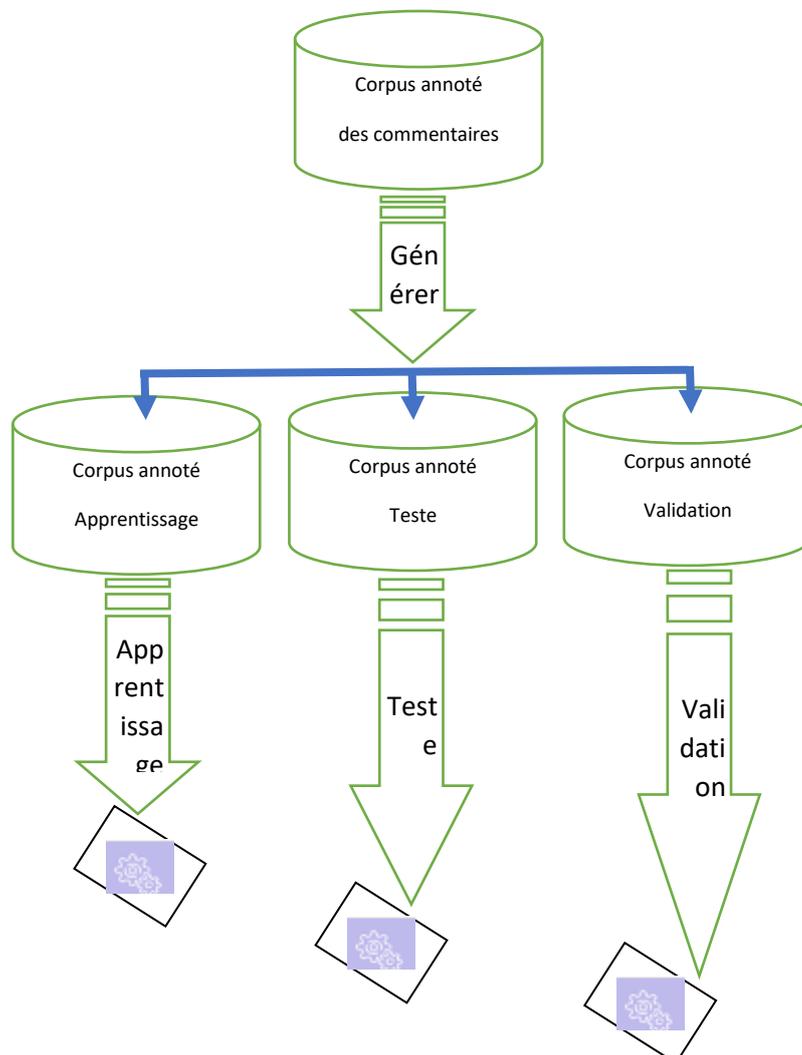
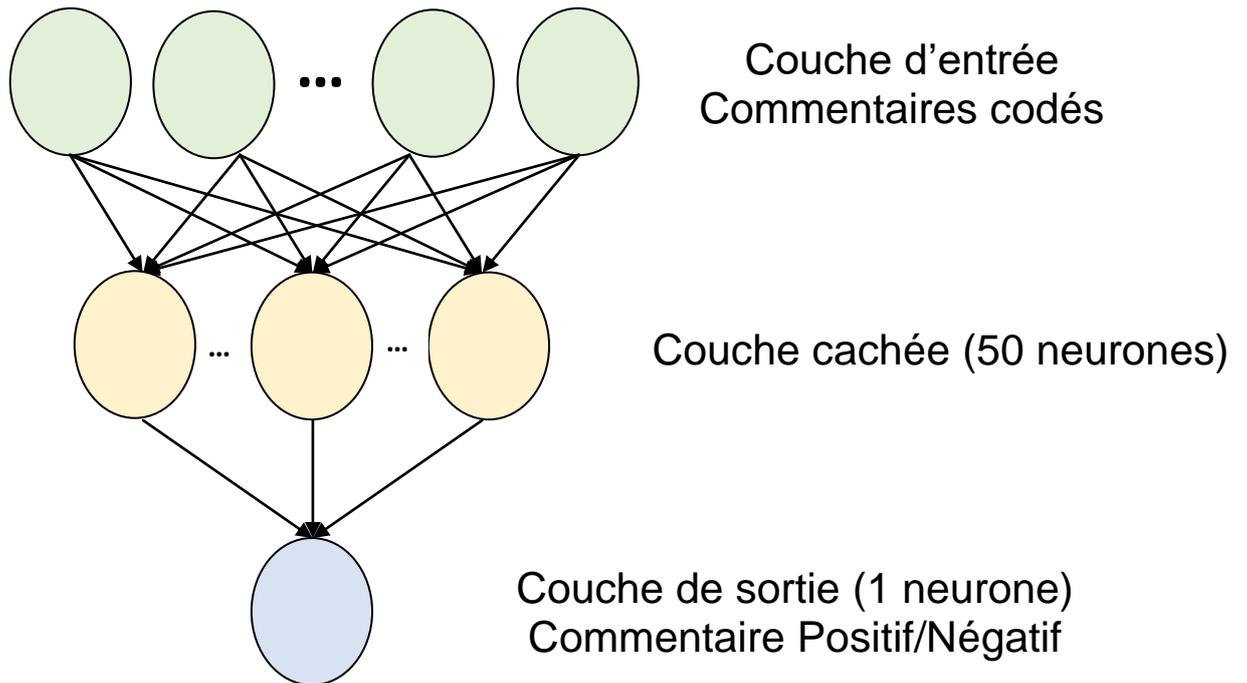
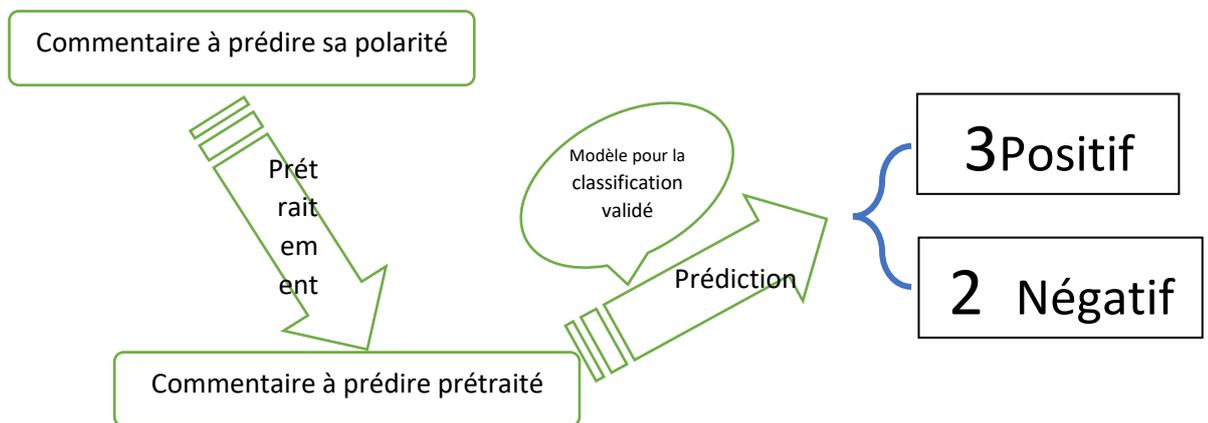


Figure 6. Exemple de modèle à base de réseaux de neurones pour la classification des commentaires



Pour la prédiction on introduit un commentaire pour le modèle construit lors de la phase précédente et le classifieur retourne le résultat de sa classification à savoir : Positif ou négatif. La figure 6 montre le schéma de la prédiction. La figure 6 montre l'étape de la prédiction de notre modèle.

Figure 6. Prédiction de la polarité d'un commentaire



6. Conclusion

Dans cet article nous avons montré la relation qui existe entre les SIC et l'IA, et nous avons aussi justifié l'utilisation de Machine Learning dans l'analyse des commentaires dans les médias en ligne au lieu d'utiliser les techniques traditionnelles des SIC. Cette étude s'est focalisée aussi sur une application de Machine Learning sur l'analyse des sentiments exprimés par les usagers des médias en ligne. D'un côté, nous avons fait une extraction de

600 commentaires de YouTube en utilisant une API gratuitement disponible sur internet et après avoir effectué le prétraitement et l'annotation de ces commentaires nous avons obtenus 500 commentaires annotés (235 positifs et 177 négatifs). D'un autre côté, nous avons utilisé des bibliothèques et des fonctions du langage de programmation Python pour construire un modèle pour l'apprentissage et la prédiction des sentiments en vers le discours du président de la république. Ce système sert à prédire la polarité des commentaires postés par les algériens dans les médias en ligne (Algérie).

Dans le cadre de cette communication nous soulignons qu'il reste d'autres modèles d'IA que nous n'avons pas testés. Nous proposons donc comme perspective de ce travail l'évaluation d'autres modèles notamment ceux de type « Transformers » qui sont parmi les modèles d'apprentissage profonds, et qui peuvent transformer un jeu de données et effectuer le classement de données comme le cas de l'analyse des sentiments (Géron, 2019:63), par exemple BERT (Bidirectional Encoder Representations from Transformers) qui semble donner de meilleurs résultats dans des tâches de classifications similaires à la nôtre.

7. Références

1. Atakpa Luc Germann Koba, *Mise en Oeuvre d'un Système de Traduction du Couple de Langue Français-Fongbé*, Université d'Abomey-Calavi, Benin, 2018, [en ligne], http://ratheil.info/myPublicThesis/bachelors/2018_AtakpaLuc--Houndji.pdf
2. Béasse Murie (2021), « IA et mesures d'audience du journalisme en ligne: pour quelle intelligence des publics? », *Les cahiers du journalisme-Recherches*, 2(7), R25-R38., <https://hal.science/hal-03709484/>;
3. Brancart Jérémy, *Construction d'un portefeuille d'actions sur base d'un modèle de Machine Learning*. Louvain School of Management, Université catholique de Louvain, Prom. : Vrins, Frédéric, Belgique, 2020, <http://hdl.handle.net/2078.1/thesis:25863>;
4. Brynjolfsson Eric, McAfee Andrew (2017), « The business of artificial intelligence », *Harvard Business Review*, pp.03-11, <https://starlab-alliance.com/wp-content/uploads/2017/09/AI-Article.pdf>;
5. Floridi Luciano (2005), « Semantic Conceptions of Information », *Stanford Encyclopedias of Philosophy*, [en ligne], <http://plato.stanford.edu/entries/information-semantic/>;
6. Gefen Alexandre (2022), « IA: pour une histoire culturelle. », *Revue d'histoire culturelle. XVIIIe-XXIe siècles*, (4)., [en ligne], <https://shs.hal.science/halshs-03913099/>;
7. Gheorghin Siméon (2022), *Etude de données Twitter en lien avec l'élection présidentielle française d'avril 2022* (Doctoral dissertation, Paris Sciences et Lettres), <https://hal.inria.fr/hal-03897939/document>
8. Géron Aurélien (2019), *Machine Learning avec Scikit-Learn : Mise en œuvre et cas concrets*, Malakoff, Dunod, France;
9. Hardeniya Nitin, Perkins Jacob, Chopra Deepti, Joshi Nisheeth, & Mathur Iti. (2016). *Natural language processing: python and NLTK*. Packt Publishing Ltd.,

https://books.google.com/books?hl=fr&lr=&id=0J_cDgAAQBAJ&oi=fnd&pg=PP1&dq=NLP+python&ots=lfwsz2rvWS&sig=w0JUZYgB_kW-oeYoNnYIDWbZXYo

10. Jain Praphula Kumar, Pamula, Rajendra (2021), « A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. », *Computer science review*, 41, 100413., <https://www.sciencedirect.com/science/article/pii/S1574013721000538>
11. Jeanneret Yve, (2011), *Y-a-t-il (vraiment) des technologies de l'information*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.
12. Khairnar Jayashri, & Kinikar Mayura (2013), « Machine learning algorithms for opinion mining and sentiment classification », *International Journal of Scientific and Research Publications*, 3(6), 1-6., [en ligne], <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=269d91e79049092bdf0651241d0d66830aa9fafc>
13. Kiyindou Alain (2019), *IA, pratiques et enjeux pour le développement*, Paris, L'Harmattan
14. Leleu-Merviel Sylvie, Useille, Philippe (2008), « Quelques révisions du concept d'information. », [en ligne], <https://hal.science/hal-00695777/>
15. Mazieres, Antoine, *Cartographie de l'apprentissage artificiel et de ses algorithmes.*, Manuscrit de thèse, Université Paris Diderot., (2016)
16. Mittal Neeti, Gujral, Harminder.Kaur (2020), « Realization of Artificial Intelligence in Human Resource Management Best Practices », *Mukt Shabd Journal*, Volume IX, Issue VII, https://www.researchgate.net/profile/Harminder-Gujral/publication/346082628_Realization_of_Artificial_Intelligence_in_Human_Resource_Management_Best_Practices/links/5fba8a0f458515b797626410/Realization-of-Artificial-Intelligence-in-Human-Resource-Management-Best-Practices.pdf
17. Petit Laurent (2021), « Les sciences humaines et sociales (SHS) et les SIC(SIC) aux défis de l'IA », *Communication, technologies et développement*, 10, [En ligne], <http://journals.openedition.org/ctd/4475>
18. Shapiro Stuart (1992), *Encyclopedia of Artificial Intelligence* (2nd Edition). New York : Wiley.
19. Stahl Bernd Carsten, Andreou Andreas, Brey Philip, Hatzakis Tally, Kirichenko Alexey., Macnish Kevin, & Wright David (2021), « Artificial intelligence for human flourishing–Beyond principles for machine learning. », *Journal of Business Research*, 124, 374-388., <https://www.sciencedirect.com/science/article/pii/S0148296320307839>
20. Suchdev Riya., Kotkar Pallavi., Ravindran Rahul., Swamy Sridhar (2014), « Twitter Sentiment Analysis using Machine Learning and Knowledge Approach », *International Journal of Computer Applications*, 4, (0975 – 8887) Volume 103, pp36-40, <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d5b7264fc852e65bfaecdde1d42e7af42d9deb06>
21. Teboul Bruno, Amri Taoufik (2014), « Les Machines pour le Big Data: Vers une Informatique Quantique et Cognitive », [en ligne], <https://hal.science/hal-01096689/>

22. Zouinar Moustapha (2020), « Évolutions de l'Intelligence Artificielle : quels enjeux pour l'activité humaine et la relation Humain-Machine au travail ? », *Activités* [En ligne], 17-1, mis en ligne le 15 avril 2020, <http://journals.openedition.org/activites/4941>