

Université de Lille,
Faculté des sciences économiques, sociales et des territoires

**Faculté des sciences
économiques, sociales
et des territoires**



**Opinion des utilisateurs concernant les vaccins contre Covid-19
sur Twitter
Et Anti-vaccination.**

Mémoire soutenu par Monsieur Nghiem Dinh Gia Nghi
Sous la direction de Monsieur Fabien Eloire

Mémoire réalisé dans le cadre du Master 1 de Sociologie,
parcours Études Numériques et Statistiques Publiques

Années universitaire 2021-2022

Remerciements

Je remercie M. Fabien Eloire d'avoir observé mon mémoire, m'ayant conseillé et orienté tout au long du processus.

Je remercie également M. Sébastien Delarre, responsable master 1ère année, enseignants en master 1ère année pour son aide, ainsi que ses conseils sur les connaissances.

Je remercie Mme. Ada Chmillevschi, M. Jean-Baptiste Richard, et M. Arnaud Gautier pour leurs soutiens dans la mise à disposition des données de Santé Publique.

Je remercie Le Khanh Trinh, Chu An Lai et Vu Van Nam, amis qui m'ont aidé sur les aspects techniques : méthodes d'apprentissage automatique, analyse de données utiles.

Résumé

La science sociale des émotions est l'un des aspects des sciences sociales. L'objectif est de comprendre le comportement et les sentiments des gens envers les événements sociaux, et leurs causes. Aujourd'hui, avec le développement de la science et de la technologie, il existe de nombreuses méthodes pour faciliter l'analyse des comportements émotionnels humains par les sociologues. L'un des environnements les plus appropriés pour l'analyse comportementale et émotionnelle sont les médias sociaux, où les nouvelles sont mises à jour quotidiennement, toutes les heures, et où les gens partagent leurs sentiments et leurs opinions sur l'actualité.

L'arrivée de l'épidémie de Covid-19 a changé la vie et les habitudes des gens, son impact sur la société a été extrêmement fort (Economie, Santé, Démographie). L'un des problèmes qui nous préoccupent le plus avec l'épidémie de Covid est de savoir comment la traiter et la prévenir. La vaccination est considérée comme la mesure la plus efficace pour prévenir la maladie, et ainsi les vaccins contre le Covid-19 sont devenus un sujet qui intéresse non seulement la communauté en ligne, mais aussi un sujet brûlant dans les débats politiques. Par les méthodes d'apprentissage automatique, d'apprentissage en profondeur dans l'analyse des attitudes et des émotions, apprenons ensemble les changements dans les émotions de la communauté en ligne, les problèmes cachés derrière les messages de l'utilisateur. Dans le même temps, la question de l'anti-vaccination est également une préoccupation majeure, car elle aura le lien principal avec la question de la vaccination contre le Covid-19 en France, alors que jusqu'à 16% des Français ne veulent pas utiliser les vaccins : des vaccins sont délivrés pour lutter contre le Covid-19.

Mots clés : Analyse des sentiments, apprentissage automatique, réseau social, Covid-19, vaccination, anti-vaccination.

Abstract

The social science of emotions is one aspect of social science. The aim is to understand people's behavior and feelings towards social events, and their causes. Today, with the development of science and technology, there are many methods to facilitate the analysis of human emotional behaviors by sociologists. One of the most appropriate environments for behavioral and emotional analysis is social media, where news is updated daily, hourly, and where people share their feelings and opinions on current events.

The arrival of the Covid-19 epidemic has changed people's lives and habits, its impact on society has been extremely strong (Economy, Health, Demography). One of the issues that concerns us most with the Covid epidemic is how to treat and prevent it. Vaccination is considered the most effective measure to prevent the disease, and thus Covid-19 vaccines have become a topic of interest not only in the online community, but also a hot topic in political debates. Through the methods of machine learning, deep learning in the analysis of attitudes and emotions, let's learn together the changes in the emotions of the online community, the problems hidden behind the user's messages. At the same time, the issue of anti-vaccination is also a major concern, as it will have the main link with the issue of vaccination against Covid-19 in France, when up to 16% of French people do not want not use vaccines: vaccines are issued to fight against Covid-19.

Keywords: Sentiment analysis, machine learning, social network, Covid-19, vaccination, anti-vaccination.

SOMMAIRE

I.INTRODUCTION	1
II.CONTEXTE	2
A. Crise Sanitaire	2
B. L'Épidémie Covid-19	3
C. Réseaux sociaux. La raison du choix de Twitter	10
III.CADRE THEORIQUE	12
A. Émotions, les mouvements des émotions, les sentiments	13
B. Réseaux sociaux	17
C. Analyse des sentiments (Approche de l'apprentissage de la machine).	19
D. Méthode d'analyse des sentiments.	22
E. Anti-vaccination.	25
IV.MÉTHODOLOGIE ET HYPOTHESES	28
A. Analyse des sentiments	28
B. Les Hypothèses	36
V.RESULTATS ET INTERPRETATIONS	38
A. Analyse des sentiments sur Twitter.	37
B. Analyse Baromètre de la Santé.	50
VI.CONCLUSION	54

I. Introduction.

Les réseaux sociaux se développent de jour en jour, et jouent un rôle important dans la transmission de l'information. Désormais, tout le monde peut participer à des sites de réseaux sociaux, se divertir, regarder des vidéos, voire commenter, partager des informations qui l'intéressent avec tout le monde autour, grâce à son "Réseau Social". En France, selon une étude, 59% des personnes interrogées sont membres d'un certain site de réseau social, si ce pourcentage correspond à l'ensemble de la population française, ce n'est pas un petit nombre. Le choix d'un site de réseautage social à rechercher est une question sérieuse et importante. Dans le contexte de la recherche sociale, Twitter est un site de réseautage social réputé pour la recherche, en raison de sa capacité à interagir facilement entre les individus, ainsi que de sa vitesse de transmission de l'information, ainsi que de la portée de la recherche. , un large public d' utilisateurs, répartis dans tous les âges, professions, et surtout de nombreuses grandes agences de presse, célébrités, personnes influentes, politiciens choisissent Twitter comme principal canal de parole.

Au cours des 2 dernières années, le mot qui est apparu dans presque tous les journaux, chaînes d'information, télévisées et réseaux sociaux a été : Covid-19. La pandémie de Covid-19 a balayé le monde entier, des pays les plus pauvres aux grandes puissances, des pays les plus arriérés à ceux aux économies les plus modernes et aux ceintures médicales durement touchées. Cette pandémie a changé, perturbé et affecté tous les aspects de nos vies. Les habitudes de vie changent, les habitudes de travail changent, la pensée épidémiologique change aussi.

La vaccination est le moyen le plus efficace de prévenir la maladie. Le vaccin Covid-19 a également été créé dans le but de revenir à un monde "normal" comme avant 2020. Mais la vaccination comporte aussi des risques, et une efficacité, donc il y a des idées. Les avis sont partagés sur la vaccination, surtout en cette période de pandémie, Le vaccin est presque la seule solution pour prévenir.

Mon mémoire souhaite analyser les attitudes et avis des internautes sur les réseaux sociaux et Twitter sur la question de la vaccination contre la pandémie de Covid-19 en France, et ainsi liée à la question de l'Anti-vaccination avant et pendant la période avant et pendant la pandémie. Y a-t-il un changement clair dans notre perception, à notre avis sur les vaccinations ? Ou serait-ce la détérioration de la confiance dans la question de la vaccination. Existe-t-il une différence dans notre attitude vis-à-vis d'un certain problème de la vie (dans ce cas, la vaccination) entre être sur les réseaux sociaux et être dans le monde réel.

Présentation des données empiriques

Mon mémoire est basé sur deux bases de données principales. Le premier est la base de données collectée sur Twitter, qui sont des tweets avec des mots-clés qui sont les noms de vaccins contre le Covid-19, la période allant de début 2020 à avril 2022, la zone de recherche sur le territoire France. Bien que je sache qu'en resserrant cette recherche il manquera des tweets sur le territoire français mais les utilisateurs n'activent pas la géolocalisation, mais dans un souci d'avoir les résultats les plus précis sur la recherche sur le territoire français, j'ai décidé d'ajouter le localisateur au filtre. La récupération des données de Twitter est autorisée

par Twitter lui-même et se fait via l'API Twitter. Les résultats incluent 8611 tweets écrits en français sur deux ans.

La seconde base de données est issue du Baromètre de Santé de 2010 à 2019 (2010-2014-2017-2019) émis par la Santé Publique. L'enquête est menée auprès de 15 000 personnes âgées de 15 à 75 ans par téléphone chaque année. Les numéros de téléphone sont tirés au sort parmi plus de 177 millions de numéros de téléphone fixe (01 à 05) et 118 millions de numéros de téléphone mobile (06 et 07). Le périmètre de l'enquête est la France métropolitaine. Dans cette enquête, nous nous intéresserons aux questions liées à la vaccination et aux caractéristiques sociales telles que l'âge et l'éducation.

Plan du mémoire

Le mémoire est divisé en six parties principales. La première partie est l'introduction, passant brièvement en revue le contexte, la raison pour laquelle vous avez choisi ce sujet. Ensuite, la section contextuelle traite des circonstances dans lesquelles la recherche a été menée. La situation pandémique est une situation particulière, mais ce n'est pas la première fois que l'humanité doit y faire face, alors quels grands ennemis nous ont testés ? La manière dont ces pandémies affectent l'humanité sera abordée dans la première section, Crise sanitaire. La deuxième section fait référence au contexte principal de cette étude, la pandémie de Covid-19, et ses effets sur la société. Et puis, nous apprendrons les caractéristiques du terrain de l'étude, le réseau social twitter, les raisons de choisir Twitter plutôt que facebook ou d'autres réseaux sociaux. La troisième partie est le cadre théorique, pour un étudiant qui n'est pas issu de la sociologie, c'est la partie la plus importante car elle me permet de comprendre beaucoup plus de définitions, ou de penser en termes de sociologie. Certains concepts du fait social ou de la sociologie des émotions seront abordés dans cette section. Une autre chose qui sera également mentionnée est la théorie liée à l'analyse sensorielle en termes d'apprentissage automatique. La quatrième partie est la méthodologie, où nous apprendrons comment les ordinateurs analysent et classent les sensations, les opinions, la structure des transformateurs, les méthodes BERT et CamemBERT sont les principaux sujets. La cinquième partie consiste à analyser les données et à interpréter et à faire des interprétations à partir des données. La dernière partie est la section des commentaires, nous combinons ce qui a été analysé ci-dessus pour avoir des commentaires détaillés sur les sentiments et les opinions de la communauté en ligne, ainsi que sur les comportements et les habitudes de leur connaissance.

II. Mise en contexte.

A. Crise sanitaire.

Le système de santé français a traversé trois grandes étapes depuis le XVIII^e siècle et la Révolution française. Les trois étapes correspondent à des crises de santé publique.

La deuxième pandémie de choléra est née en Inde en 1826. À Paris, la première infection a été détectée le 26 mars 1832, début avril de cette année-là, 1853 personnes étaient infectées¹. Pendant six mois, de mars à septembre 1832, une épidémie de choléra coûta la vie à près de 18 500 habitants de la capitale.² Le département Seine-et-Marne a perdu 2,5 % de sa population en 6 mois environ. L'épidémie s'est terminée en septembre-octobre de cette année-là, avec environ 100 000 victimes. Cela a eu un impact sur le gouvernement puis, des années après l'épidémie de choléra en 1832, le préfet de police Gabriel Delessert a pris des mesures draconiennes pour assainir les quartiers insalubres de Paris, qui avaient des conditions de vie extrêmement précaires, et améliora le réseau d'égouts à la fois. La pandémie de choléra de 1849 est considérée comme une épidémie mondiale. En France, peu de temps après l'arrivée des marins britanniques dans le port de Dunkerque le 20 octobre 1848, des épidémies ont commencé à se propager dans le nord de la France. La première vague de choléra a coûté la vie à 19 184 personnes dans la capitale, en France, il y a eu 100 661 victimes, la deuxième vague de choléra en 1853-1854, la France a fait 143 468 victimes. Les pandémies de choléra de 1832 et 1849 provoquent une prise de conscience politique en faveur de la loi de 1850 sur l'habitat insalubre et lancent le processus sanctionné par le vote de la grande loi de santé publique de 1902³.

La deuxième étape a été la période où le gouvernement français a construit un système de santé diversifié et relativement efficace. Ici aussi, nous avons une pandémie mondiale qui a coûté la vie à 50 à 100 millions de personnes, représentant entre 2,5 et 5 % de la population mondiale- la grippe espagnole (Johnson & Müller, 2003 ; Chandra, 2013). En France, le bilan est actualisé à 240 000 décès alors que la population est à la moitié de 2020. L'épidémie se divise en trois phases. La première phase s'est déroulée d'avril à juillet 1918 en France avec de nombreux cas, mais une faible létalité. À partir d'août 1918, de nombreux patients atteints de broncho pneumonie ou d'autres complications ont développé des symptômes de grippe en moins de deux jours. Le pic de l'épidémie est identifié en octobre 1918, selon les seules statistiques de l'époque, la ville de Paris et le département de Seine comptent 2 500 décès par semaine. Les activités économiques et sociales sont perturbées. Les écoles, les lieux publics tels que les théâtres, les cinémas sont fermés par mesure de précaution. La troisième vague de l'épidémie de grippe saisonnière débute au printemps 1919. Bien que la pandémie de grippe espagnole n'ait été observée que pendant 10 mois, ses conséquences sont très graves. Par rapport à la pandémie de Covid-19 (notre situation

¹ Jacques Boucher de Perthes (1863). *Sous dix rois: souvenirs de 1791 à 1860*, vol 5.

² Michelle Zancarini-Fournel (2016). *Les luttes et les rêves: Une histoire populaire de la France de 1685 à nos jours*, p.251, chap 7

³ Loi du 15 février 1902 relative à la protection de la santé publique, JO du 19 février 1902 ; Bourdelais P., *Les épidémies terrassées, une histoire de pays riches*, Éd. de la Martinière, 2003.

d'observation), en seulement 10 mois, elle a coûté la vie à 50 à 100 millions de personnes dans le monde, et le Covid-19, depuis début 2020, plus de 5 millions des gens sont morts de ce virus. Bien sûr, d'un point de vue humanitaire, 5 millions ou 50 millions de morts sont tous deux des chiffres extrêmement déchirants, mais en termes de gravité, l'épidémie de grippe espagnole a ensuite frappé un plus

fort que le Covid-19 à l'heure actuelle (d'abord en termes de démographie). Les principales raisons invoquées pour expliquer le manque de considération pour l'épidémie de grippe espagnole comprenaient alors la situation de guerre (Première Guerre mondiale), l'origine inconnue de la maladie et le retard de la réaction du gouvernement. Après des conséquences aussi graves, dès la fin de la première guerre mondiale, le premier ministère de la Santé est créé fin 1920⁴. Cette nouvelle administration tente de développer un système hospitalier sur l'ensemble du territoire français et accompagne le développement de la médecine municipale, quand les statuts de 1928-1930 et le plan de 1945 autorisent la mise en place d'un programme de sécurité sociale.

Enfin, une pandémie qui tue encore plus d'un million de personnes par an dans le monde, SIDA. Sida, le syndrome d'immunodéficience acquis a été découvert pour la première fois à la fin des années 1970, lorsque les médecins de San Francisco ont remarqué que leurs patients étaient déprimés, perdaient du poids et souffraient de cancers rares. La maladie a d'abord été connue sous le nom de "pneumonie gay" ou "cancer gay", lorsque l'incidence du sarcome de Kaposi chez les homosexuels était inhabituellement élevée⁵. Ces noms se sont avérés plus tard inappropriés car leur popularité ne s'est pas arrêtée aux homosexuels : à l'été 1982, l'utilisation de l'abréviation AIDS a commencé aux États-Unis. L'épidémie de sida et le scandale des empoisonnements sanguins de la fin des années 1980 ont été le résultat de l'absence d'une culture de la santé publique. Quand il est tout à fait possible d'éviter la propagation du virus. La France est un pays avec un taux d'infection après transfusion sanguine plus élevé que les autres pays⁶. La pandémie reflète aussi les carences du système de santé, alors qu'avant l'épidémie de sida, les gens ne s'intéressent qu'au traitement, mais oublient l'importance de la prévention. Avec l'émergence de la sécurité sanitaire, un système d'institutions et de réglementations s'est mis en place à une échelle sans précédent.

B. L'épidémie Covid-19.

1. *L'impact social.*

L'épidémie du Covid-19 creuse les distances sociales et les inégalités sociales. L'inégalité sociale de santé est renforcée par trois facteurs: les inégalités face au risque d'exposition, les

⁴ Documents parlementaires, Chambre, annexe n° 5131, séance du 31 octobre 1918, p.1793.

⁵ Lawrence K. Altman (consulté le 25 mai 2021), "RARE CANCER SEEN IN 141 HOMOSEXUALS", *The New York Times*, 3 juillet 1981; <https://www.nytimes.com/1981/07/03/us/rare-cancer-seen-in-41-homosexuals.html>

⁶ Morelle A. (1996), *La défaite de la santé publique*, p.24

différences de vulnérabilité face à la maladie et les écarts d'accès aux soins lors des épidémies⁷. Avec le covid, on peut ajouter les effets du confinement.

Le risque d'exposition est plus élevé pour certaines professionnelles.

Les différentiels d'exposition professionnelle semblent être une mesure importante dans plusieurs populations, qui peuvent être attribuées à l'exposition aux agents pathogènes, à la fréquence des contacts avec les autres sur le lieu de travail, à les modes de déplacement. L'étude 2020 d'Almagro sur les disparités d'exposition au Covid-19 à New York et dans les environs indique que le secteur professionnel est le facteur déterminant des disparités d'exposition. Les travailleurs de la santé, en particulier les travailleurs de première ligne, qui sont en contact direct avec des personnes malades et s'en occupent sont les plus exposés au risque d'infection. Outre les personnels soignants, deux populations sont plus à risque d'exposition ultérieure dans leurs professions, soit par des effets d'utilité sociale, soit à risque de perte de revenus du fait de l'épidémie. Dans ce deuxième cas, on peut dire que la population des personnes à faible revenu, ou sans emploi stable, est celle qui est durement touchée par l'épidémie⁸. On voit que les gens doivent choisir entre leur risque économique et leur risque sanitaire. De plus, le télétravail est moins développé dans les métiers non qualifiés que dans les métiers qualifiés, de sorte que le risque d'exposition dans ce groupe professionnel sera plus élevé⁹. Une enquête rapide de la DARES menée du 1er au 10 avril 2020 sur la continuité d'activité depuis le début de l'épidémie. Cela prouve que 27 % des salariés travaillent encore sur le site au 31 mars, et il y a une inégalité par les secteurs.¹⁰

Parmi les secteur où le plus de salariés travaillent encore sur le site, on a le secteur de fabrication de denrées alimentaire (plus de 50%), celui de la santé et l'action sociale (plus de 40%), des transport, de l'industrie d'énergie et du commerce (plus de 30%).

En particulier, la demande croissante de transports en commun à long terme, combinée à la contrainte de travailler au siège social, augmente le potentiel d'exposition, ce qui a des implications pour votre lieu de vie. Selon une étude de Mangeney 2020, qui montre que les personnes des secteurs clés qui ont été actives pendant les périodes de confinement en Ile de France, sont surreprésentées en Seine-Saint-Denis (12%) par rapport aux autres régions (7%). De plus, la moitié d'entre eux doivent quitter leur département pour travailler.

Le risque d'exposition lié avec la condition de la vie

Le risque d'exposition est également accru pour les personnes vivant dans des logements collectifs ou des établissements fermés, citant des espaces de vie partagés, et le mouvement des individus augmente la probabilité d'infection comme dans les Maisons de retraite (ou EHPAD); les établissement d'hébergement pour personnes handicapées type MAS/FAM; les foyers et résidences sociales (protection de l'enfance, hébergement social, établissement d'accueil des demandeurs d'asile, foyers de travailleurs migrants,...) et dans les prisons, les camps de réfugiés et centres de rétention administrative,...

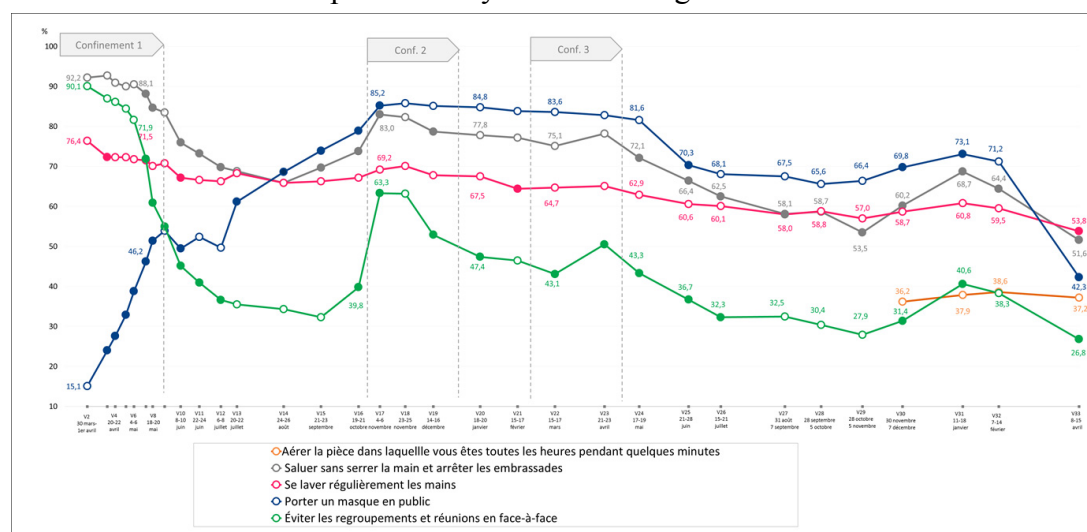
⁷ Blumenshine et al. (2008), *Sydenstricker, 1931*

⁸ Enrico Bergamini (2020), *How COVID-19 is laying bare inequality*

⁹ Dominique Mesda (2020), *France Culture*.

¹⁰https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/dares_acemo_covid19_synthese_17-04-2020.pdf

De la manière la plus générale, on peut dire que les conditions d'hébergement sont liées à l'infectiosité du virus. Quant à l'épidémie de Covid-19, nous savons tous que le virus est capable de se propager par voie aérienne, ce qui signifie que la transmission du virus chez les membres d'un même foyer est importante, la capacité à s'isoler ou à respecter la distance entre les personnes est presque impossible dans les endroits à forte densité de population et aux mauvaises conditions de vie. Et les conditions de vie peuvent également déterminer un autre problème, qui est l'accès aux mesures de protection, la capacité de protéger, comment prévenir les virus. Cet accès dépend des conditions matérielles pour se procurer des équipements de protection individuelle et comprendre les effets de chacune de ces mesures (masques, produits d'hygiène, gants, ...). À cet égard, les annonces, les articles doivent avoir des messages clairs et faciles à comprendre, les destinataires des informations doivent avoir un niveau suffisant de compréhension de lecture et de compréhension des textes sanitaires et médicaux, une conscience du niveau de risque et de danger de la maladie. L'inégalité sociale à cet égard est probablement une conséquence inévitable de toutes les sociétés hiérarchisées, quand on a des différences dans les conditions de vie, les niveaux d'éducation, etc., donc la réception de l'information Il y a aussi des différences de croyances ou de mesures préventives au sein de la société, entre riches et pauvres, entre des personnes ayant des niveaux d'éducation différents et entre des personnes ayant des idéologies différentes.



le sentiment de distance est sous-estimé dans l'intervalle entre les deux premières détentions, suivi d'une baisse des autres mesures de prévention, à l'exception du port du masque. Selon le tableau ci-dessus, nous pouvons également voir que le port d'un masque est passé de la mesure la plus répréhensible à devenir une mesure que presque tout le monde utilise, même jusqu'à l'heure actuelle lorsque le port d'un masque Les masques sont devenus facultatifs à l'extérieur, certaines personnes encore utilisent cette mesure lorsqu' ils sortent. Des conditions de travail et de vie différentes exposent les gens à des risques d'infection différents. Les facteurs ci-dessus ont créé une inégalité dans le niveau de vulnérabilité à la pandémie de Covid-19

L'inégalité de la vulnérabilité face au Covid-19

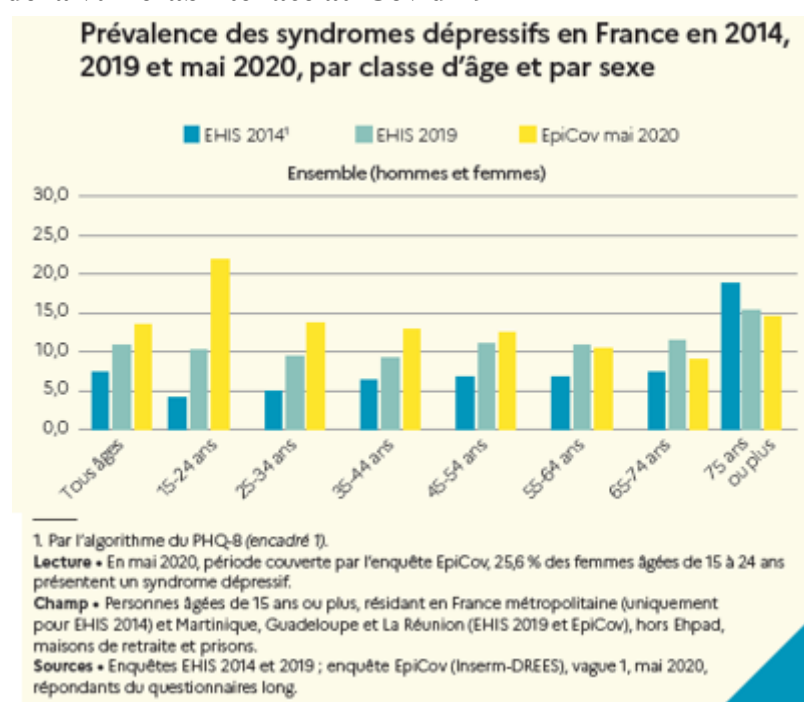


Figure 2

Lorsque l'épidémie de Covid-19 est arrivée, les déplacements entre les zones étaient limités ou interdits, seuls quelques cas étaient autorisés à se déplacer alors que la distanciation sociale était en cours. Cela rend difficile et empêche les gens de communiquer avec les autres. Limiter les contacts et les personnes restant à la maison peut aider à limiter la propagation de l'infection, mais cela ne peut pas effacer l'injustice de la vulnérabilité. Il est clair que les personnes âgées sont les plus vulnérables aux effets de Covid-19, à la fois physiquement et mentalement. Les personnes âgées sont celles qui sont porteuses de nombreuses maladies sous-jacentes ou qui ont une résistance plus faible que le reste des âges, de sorte que les personnes de ce groupe d'âge sont plus sensibles à l'infection par Covid-19 que les autres âges. Mentalement, ne pas pouvoir sortir, ne pas être en contact avec les gens va frapper un gros coup psychologique, surtout pour ceux qui sont seuls. Selon la Drees, 15 % des personnes âgées interrogées ont présenté des signes de dépression en période

épidémique. Toujours selon la DREES, il est à noter que les personnes les plus touchées mentalement par la pandémie de COVID-19 ne sont pas les plus de 65 ans, mais les 15-24 ans, cela paraît absurde, car les jeunes sont généralement plus en forme et avoir plus de formes de divertissement. Cependant, parce que l'habitude des jeunes Français est de sortir chaque semaine pour s'amuser et rencontrer des amis, lorsque la pandémie frappe, ils doivent quitter leurs routines hebdomadaires et rester à la maison et se lier d'amitié avec des équipements technologiques. En conséquence, 22 % des répondants âgés de 15 à 24 ans ont déclaré avoir ressenti des symptômes dépressifs pendant le premier confinement, soit plus du double de l'EHIS de 2019.

Nous avons donc une liste des personnes les plus vulnérables au Covid-19. D'une part, selon les données de la littérature sur le Covid-19, les personnes âgées de 65 ans et plus (même si les personnes âgées de 50 ans à 65 ans doivent être surveillées de façon plus rapprochée); des personnes avec antécédents (ATCD) cardiovasculaires : hypertension artérielle compliquée (avec complications cardiaques, rénales et vasculo-cérébrales), ATCD d'accident vasculaire cérébral ou de coronaropathie, de chirurgie cardiaque, insuffisance cardiaque stade NYHA III ou IV; les diabétiques, non équilibrés ou présentant des complications; les personnes ayant une pathologie chronique respiratoire susceptible de décompenser lors d'une infection virale (broncho pneumopathie obstructive, asthme sévère, fibrose pulmonaire, syndrome d'apnées du sommeil, mucoviscidose notamment) ; les patients ayant une insuffisance rénale chronique dialysée ; les malades atteints de cancer évolutif sous traitement (hors hormonothérapie); et les personnes présentant une obésité. D'autre part, en raison d'un risque présumé de Covid-19 grave: les personnes avec une immunodépression congénitale ou acquise : médicamenteuse : chimiothérapie anticancéreuse, traitement immunosuppresseur, biothérapie et/ou corticothérapie à dose immunosuppressive ; infection à VIH non contrôlée ou avec des CD4 <200/mm³ ; consécutive à une greffe d'organe solide ou de cellules-souches hématopoïétiques ; liée à une hémopathie maligne en cours de traitement ; les malades atteints de cirrhose au stade B du score de Child Pugh au moins ; les personnes présentant un syndrome drépanocytaire majeur ou ayant un antécédent de splénectomie ; les femmes enceintes, au troisième trimestre de la grossesse, compte tenu des données disponibles et considérant qu'elles sont très limitées.

Covid-19 a changé l'emploi français.

Le Covid a eu un fort impact sur les activités des entreprises. Les activités "non essentielles" sont restreintes ou interdites, ce qui entraîne la fermeture de nombreux magasins, types de services ou pénurie de personnel en raison des premières distanciations sociales. Selon une enquête Acemo, dans les entreprises non agricoles de 10 membres ou plus, jusqu'à 19 % des employés ont démissionné, 30 % ont réduit leur temps de travail d'au moins la moitié. Malgré un retour en force après s'être adapté à la situation épidémique, au premier trimestre 2020, 715 000 emplois ont été détruits, la plupart fin mars, les 2 premières semaines de l'ordre de distanciation sociale. Après avoir été durement touchés fin mars, les emplois se sont légèrement redressés au deuxième trimestre, bien qu'un quart des emplois temporaires permanents ne soient pas revenus à la fin du deuxième trimestre. Les entreprises privées ont largement contribué au nombre d'emplois détruits (-650 000 emplois). Les jeunes sont

également durement touchés par l'épidémie, 9% des jeunes de 15-24 ans ont perdu leur emploi. Au troisième trimestre, l'emploi a fortement rebondi, avec plus de 344 000 emplois créés, dont 136 000 emplois temporaires¹¹.

Il y a eu un paradoxe durant cette période, lorsque le taux de chômage tel que défini par le BIT a baissé. Cette baisse est en réalité due à l'épidémie, qui a provoqué un certain nombre de chômeurs qui remplissent encore les critères du BIT en raison de difficultés à postuler et à travailler.

Les cadres semblent avoir échappé à la crise de l'emploi puisque 86 % d'entre eux continuent de travailler, dont les deux tiers télé-travaillent. La pandémie a créé une nouvelle façon de travailler, celle du télétravail. Cela crée également une inégalité au travail. 67 % des cadres travaillent à domicile, tandis que dans les emplois en milieu de carrière, 53 % des employés sont des hommes et 51 % des employées travaillent à domicile. Ce chiffre pour les employés et les ouvriers n'est que de 27% et 3%. La majorité des travailleurs doivent encore travailler hors site, ce qui, comme mentionné ci-dessus, entraîne un risque d'infection plus élevé. Par rapport à la même période en 2019, le pourcentage de personnes travaillant à domicile a doublé au premier moment de la distanciation sociale (48 % contre 22 %). En ce qui concerne le sexe, les femmes ont un taux de travail à domicile plus élevé que les hommes (51-43). Les moins de 25 ans, c'est le pourcentage de personnes travaillant à domicile qui a le plus augmenté, multiplié par 6 sur la même période en 2019 (5% vs 30%). En termes d'occupation, les cadres ont connu l'augmentation la plus spectaculaire de 45 % à 81 %.

2. *L'impact démographique.*

La mortalité a fortement augmenté en mars-avril 2020, mais l'inégalité entre les départements. En janvier et février 2020, alors que le nombre d'infections au Covid-19 n'est pas très élevé, le nombre de décès par jour est de 1700-1800 personnes. Ce nombre de décès tend à diminuer par rapport aux années précédentes, où le nombre moyen de décès sur la même période des années précédentes était de 1900 personnes par jour. Ce nombre est resté jusqu'à la première quinzaine de mars 2020, juste avant le premier confinement. Mais depuis le 15 mars 2020, date à laquelle le nombre de cas contaminé confirmé en France a commencé à grimper à 4 chiffres par jour, le nombre de décès par jour a commencé à montrer des signes d'augmentation significative, et c'est aussi le moment où l'on sait que le Covid-19 a un réel impact sur nos vies. Le 16, le nombre de morts a atteint 2000 personnes par jour, et s'est continuellement maintenu à plus de 2000 morts par jour pendant un mois. Le pic a été le 1er avril, le nombre de morts a atteint 2809 (en hausse de 66% par rapport au même jour en 2019), le plus élevé depuis 15 ans. La dernière fois que le nombre de morts a été plus élevé, c'était les 11-12-13 août 2003, à cause de la canicule. Dans les jours suivants, il a commencé à décliner, mais est resté au-dessus de 2 200 décès par jour. Dans la période de mai à septembre 2020, la mortalité est revenue à des niveaux normaux par rapport aux années précédentes. Au total, pour la période de mars et avril 2020, le nombre de décès a augmenté de 27 000 personnes (+27%) sur un an, et une augmentation de 26% par rapport à la moyenne

¹¹ Thierry Mainaud (2020), *Les inégalités sociales à l'épreuve de la crise sanitaire*, p.18

de la même période des cinq années précédentes. Avec la restriction des déplacements ou des contacts pendant la période du confinement, le nombre de décès dus à des causes telles que les accidents de la circulation, les accidents du travail doit évidemment être réduit. Et la cause de cette augmentation soudaine du nombre de morts s'explique par la première vague de Covid-19 qui a frappé la France. Selon le site international de statistiques Worldometer, durant les mois de mars et avril 2020, le nombre de décès dus au Covid-19 en France était de 24 350 personnes, représentant 90 % de l'augmentation des décès durant cette période.

L'Ile-de-France a été la région plus touchée par la première vague : +91% de décès en mars-avril de 2020 par rapport à la même période en 2019 (Hauts-de-Seine +112% ; Seine-Saint-Denis +123,4%). On peut voir que la plus forte de décès se situe aux départements du nord de la région. La deuxième région plus touchée par Covid-19 est le Grand-Est (+55%) dans laquelle Haut-Rhin a augmenté 115,9%. Les Hauts-de-France est dans la même situation (+28%) dans laquelle l'Oise a été touchée dès la première semaine de la première vague.

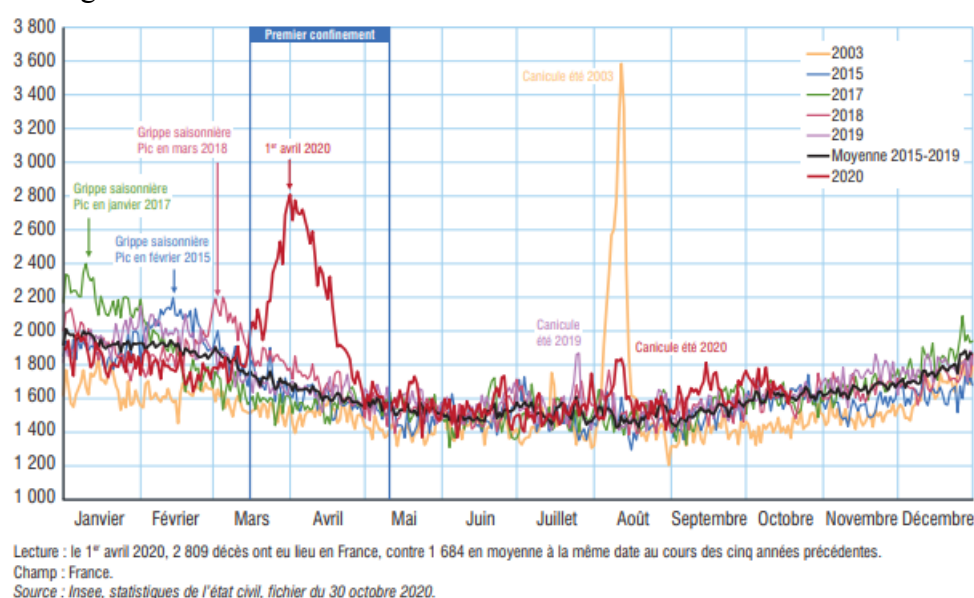


Figure 4

L'épidémie touche fortement les personnes âgées, les étrangers et les régions pauvres. Il y a de la différence de l'augmentation de la mortalité selon la sexe, d'après l'INSEE. Le nombre de décès des hommes augmente de 27% contre 25% des femmes. Le nombre d'hommes de plus de 85 ans a augmenté de 35%, contre 29% des personnes âgées. Les chiffres pour les 75-84 ans sont respectivement de 33% et 26%. L'augmentation de la mortalité a tendance à diminuer aux âges plus jeunes. Même en dessous de 25 ans, le nombre de décès a nettement diminué (18,4 % chez les hommes, 8 % chez les femmes).

Il n'est pas possible de savoir s'il existe ou non un lien entre la mortalité et la profession, ainsi que le niveau de vie actuel. Mais on voit bien que, il y a une inégalité sociale, où la mortalité des étrangers, des immigrés d'Afrique, d'Asie est plus élevée que celle des Français, surtout qu'on a observable en Seine-Saint-Denis. Entre mars et avril 2020, la

mortalité des personnes nées en France a augmenté de 22 %, le nombre de jours des personnes nées au Maghreb de 54 % et dans les autres pays d'Afrique de 114 %. Autre particularité dans ce groupe de personnes, l'épidémie est plus touchée chez les moins de 65 ans que chez les personnes nées en France (+96% pour ceux nés en Afrique, hors Maghreb). En particulier, la Seine-Saint-Denis est frappée par le plus fort excédent de mortalité en le premier confinement (+123%). La mortalité des africains a augmenté de 191% pour le Maghreb et 368% pour les autres pays d'Afrique. D'après l'INSEE, la Seine-Saint-Denis est le deuxième département plus pauvres de France en 2020, et premier de l'Ile-de-France.

C. Réseaux sociaux, la raison du choix Twitter.

1. Utilisateurs de Twitter en France :

Population : Les utilisateurs actifs mensuels sont de 16,8 millions (4,5 millions quotidiens). En 2020, les utilisateurs de Twitter en France sont 60% des hommes et 40% des femmes. On peut voir que les hommes sont majoritaires à utiliser Twitter plutôt que les femmes. Les utilisateurs par âge se découpent en cinq tranches : 22% des utilisateurs en France sont de 18-24 ans, 25% des utilisateurs en France sont de 25-34 ans, 24% des utilisateurs sont de 35-44 ans, 20 % sont de 45-54 ans, et que 9% sont de 55-64 ans. On peut voir qu'il y a une distribution uniforme entre le nombre des utilisateurs de 18-54 ans (trois tranches).

Caractéristiques : D'Après l'enquête #MoiJeune en 2020 du pouvoir sur l'information de 20 minutes, 52% des 18-30 ans choisissent Twitter qui est le réseau social ayant le plus de pouvoir, la deuxième est Facebook (32%). Avec le site de réseau social le plus puissant, Twitter est influent dans la transmission d'informations sur la vie sociale et politique aux utilisateurs de sites de réseautage social. Surtout pendant la période la plus stressante de l'épidémie de Covid 19. 62% des utilisateurs en France pensent que l'information y est diffusée plus rapidement que les autres sites. Pendant le confinement, 41% des internautes de la génération Z (15-24 ans) utilisent Twitter chaque mois, il augmente de 3% avec 38% avant le confinement.

2. La raison du choix de Twitter.

Twitter et ses avantages

Twitter est l'une des plus grandes plateformes de réseautage social sur Internet. L'adhésion est gratuite et très simple en créant un compte. Twitter est très simple d'utilisation, il vous suffit de "tweeter" ce que vous pensez, qui sera largement partagé par tout le monde. Vous pouvez tweeter ou « retweeter » ou mentionner une personne en particulier. En plus de son interface simple, Twitter a un avantage sur ses concurrents dans la transmission d'informations de contenu :

Fil d'actualité en temps réel.

Twitter est réputé pour diffuser des informations en direct d'une manière qu'aucun autre site de réseautage social n'a jamais réalisée. Les annonceurs, les célébrités ou les chaînes médiatiques transmettent le message en créant des tweets, et ils peuvent rester informés en tweetant directement ci-dessous. Plus que cela, Twitter permet d'informer les gens de ce qui

se passe dans n'importe quel endroit, ville, communauté ou portée mondiale, avec la date et l'heure exactes. Cela permettra de transmettre plus facilement l'information à tout le monde et plus rapidement à ceux qui s'en soucient.

Public plus diversifié.

L'un des principaux objectifs de l'utilisation de Twitter est de socialiser en se connectant avec des amis, des membres de la famille et en obtenant des informations qui vous intéressent. C'est la meilleure façon de garder un réseau social ensemble. Twitter est considéré comme le réseau social le plus puissant de France en 2020, ce qui signifie que les informations sur Twitter ont une plus grande influence sur les utilisateurs que les autres réseaux sociaux. De plus, les personnes à revenu élevé choisissent souvent Twitter plutôt que Facebook ou Instagram, de même que les politiciens ou les grandes chaînes d'information choisissent également Twitter comme premier canal pour transmettre des informations.

Twitter : un outil efficace de partage de la santé publique

Histoire de la relation entre les médias numériques et les soins de santé.

L'utilisation passée des médias numériques a été à l'intersection des médias sociaux et de la médecine, et elle renforce cette intersection pour le moment. L'utilisation des médias numériques en médecine ou dans les soins de santé n'est pas rare, comme l'explique le Dr Gerald Weissman dans son livre *Médecine digitale* publié au début de 2000, suite à la popularité généralisée de la télévision par câble et Internet.¹²

Dans les soins de santé.

De nombreux médecins considèrent la communication numérique comme un outil efficace pour impliquer davantage les patients dans les soins de santé et éliminer les obstacles à la compréhension du traitement médical¹³. Bien que la télévision permette aux téléspectateurs de participer, il s'agit d'un type de participation passive, où l'interaction entre les médecins, les professionnels de la santé et les téléspectateurs est à sens unique. Les médias sociaux, en particulier Twitter, sont différents des programmes de télévision traditionnels en ce sens que, que vous soyez un patient, un médecin, un spécialiste ou simplement un « passant », vous pouvez participer à la conférence ; Toute personne possédant un compte Twitter peut publier un message public de 280 caractères ou moins. Au lieu d'informations provenant d'un côté, où les diffuseurs doivent maintenir leur audience, Twitter permet à tout professionnel de la santé de partager son expertise en créant un compte¹⁴. De nombreux comptes fournissent des informations médicales, alors que seuls quelques programmes de santé sont diffusés chaque semaine. Une étude récente a confirmé qu'il y a plus de 2000 médecins sur Twitter sur la base d'identifiants. Tous les médecins tweetent plus d'une fois par jour et ont tous plus de 300 abonnés.

La première chose qui rend Twitter de plus en plus populaire dans l'industrie médicale est que Twitter permet aux professionnels de la santé d'atteindre un large éventail de publics,

¹² Weissmann (2012), *G. Epigenetics in the Age of Twitter: Pop Culture and Modern Science*

¹³ Cayton, H. The alienating language of health care (2006), *J. R Soc. Med*

¹⁴ Mesko, B. Conclusion. (2013) In *Social Media in Clinical Practice*, p.139-141

des autres médecins aux patients. De plus, sur Twitter, les gens ne reçoivent pas d'informations de manière passive, ils peuvent partager leurs opinions en répondant ou en "retweet", ou en "aimant" des informations jugées utiles et exactes. Ces réponses aident les experts à savoir quelles informations les gens s'intéressent ou à gagner plus de confiance de la part des gens. De plus, Twitter fournit un outil appelé "hashtag", qui désigne un sujet sur lequel les utilisateurs se concentrent pour discuter. Twitter aide les professionnels de la santé à se connecter avec des collègues et des patients pour partager de nouveaux traitements passionnants et des connaissances spécialisées.

Réduire les barrières en médecine est également très précieux, car une coopération constructive peut améliorer considérablement les méthodes de traitement et accroître la coopération entre de nombreux domaines tels que l'informatique, la technologie, la biologie, etc. Twitter sert de grand terrain de jeu, sur lequel tout le monde, tout expert peut "jouer", mais surtout, tout le monde peut se vérifier, vérifier les informations, les connaissances. La formule nouvellement publiée est correcte ou non. Tweeter un article, partager votre point de vue avec tout le monde sera plus facile que de publier des commentaires sur cette page de journal.

Twitter aide également les professionnels de la santé à se tenir au courant des progrès de la médecine. Le langage de la médecine n'étant pas uniforme à l'échelle mondiale, Twitter sert à pousser ces avancées scientifiques aussi loin qu'il peut aller, il suffit d'internet et d'un compte créé en moins d'une minute jusqu'à 3 minutes, même si vous êtes en France, aux Pays-Bas ou en Chine, vous pouvez toujours accéder aux informations depuis les États-Unis. Twitter fournit une communication bidirectionnelle entre sa source et son destinataire. Cette disposition donne la parole aux parties prenantes, aidant les personnes ayant un accès limité à la santé traditionnelle à recevoir ouvertement des informations sur les traitements.

III. Cadre théorique.

Dans ce chapitre, nous étudierons ensemble la théorie des problèmes abordés dans le mémoire. Des documents, des études, des livres de sociologues, de psychologues ou de scientifiques seront mentionnés dans ce chapitre comme guides pour mon mémoire.

Mon mémoire comporte deux parties distinctes, à savoir la sociologie et l'apprentissage automatique. Ces deux parties ne sont pas complètement séparées, mais elles se complètent. D'une part, si la sociologie est considérée comme l'épine dorsale, à partir d'elle, la partie apprentissage automatique prend tout son sens, d'autre part, l'apprentissage automatique permet de prouver ou d'infirmer les arguments tirés de la sociologie.

L'aspect principal qui nous intéresse est les émotions, les sentiments. Alors en sociologie, comment les émotions et les sentiments sont-ils définis, et que signifient-ils ? La première partie nous aidera à répondre à cette question.

A. Émotions, mouvements des émotions, et sentiments.

D'après Nathalie Vallerand, les émotions sont des réactions spontanées à une situation. Elles peuvent être présentes par les manifestations physiques ou psychologiques. La joie, la peur et la colère sont des exemples.¹⁵

Julien Bernard dit que les émotions sont un phénomène que nous connaissons tous, mais qu'il est difficile à définir. Leurs définitions soulèvent des questions dans des facteurs différents (physiologie, psychologie, sociologie, etc.). Les émotions sont un phénomène affectif se produisant souvent par des parties du corps, des pensées, plus ou moins contrôlables.

Qu'en sont-ils les sentiments ? D'après Nathalie Vallerand, les sentiments sont un état affectif plus durable qui évolue avec le temps. Cela signifie que les sentiments naissent, grandissent, puis peuvent se terminer et disparaître avec le temps. L'amour, la haine, la confiance, le doute, le bonheur sont les exemples les plus visibles des sentiments dans la vie. Bien qu'ils soient différents par définition, les émotions et les sentiments sont étroitement liés. Les sentiments rendent toutes les émotions révélées, dans le sens inverse, les émotions peuvent aussi créer des sentiments, le mouvement d'émotions est aussi l'expression d'un certain sentiment. Les émotions sont considérées comme un phénomène individuel, parfois "subjectif", tandis que les sentiments sont parfois considérés éthiquement en raison de leur stabilité.

En raison de la constitution des sentiments, au lieu de traiter de l'analyse des sentiments, ou de la sociologie des sentiments, les sociologues étudient davantage la sociologie des émotions, et l'analyse du changement ou du mouvement des émotions. Cette approche facilitera l'analyse des sociologues, au lieu d'analyser un sentiment complexe, on pourra décomposer chaque émotion, puis analyser l'enchaînement des émotions. Par exemple, définir le sentiment d'amour sera très vague, mais si nous abordons avec des facteurs émotionnels tels que l'attention, la confusion et que nous passons ensuite à l'excitation, la joie, cela nous aidera à déterminer plus facilement le type d'amour.

La sociologie des émotions reconnaît la dimension expressive de l'expérience personnelle ainsi que la dimension subjective de la perception, de la pensée ou de la moralité qui peut en être son origine, mais se distingue des autres sciences qui étudient les émotions, la sociologie des émotions reconnaît l'impact de la dimension historique, culturelle ou politique sur un individu, un groupe social ou une société. Et si les émotions sont affectées par des facteurs externes tels que la société, la culture, la politique, alors on peut aussi s'intéresser à la position de l'individu dans la société. Déterminer la place des individus ayant ce type d'émotion dans la société peut expliquer pourquoi ils ont cette émotion, et de plus, cela peut aussi aider les sociologues à comprendre le statut, en pensant à un groupe d'individus, une classe dans la société. Selon la définition d'Emile Durkheim, les manières de faire qui s'imposent à un individu et lui sont extérieures, ou en d'autres termes, les facteurs sociaux qui affectent l'individu s'appellent "faits sociaux".

¹⁵ Nathalie Vallerand (2014), *Magazine Naitre et grandir*.

A.1 Faits sociaux.

Selon Durkheim, la notion de faits sociaux est au cœur du concept de la sociologie. Il définissait, dans les Règles de la méthode sociologique, les faits sociaux comme toute voie, fixe ou non, susceptible de lier un individu à l'extérieur de cet individu, d'autre part : dans une société donnée, la sociale est sa propre existence, indépendante du niveau individuel. Durkheim décrit la société non seulement comme la somme des individus, mais comme une société existant par elle-même, sui generis. Les gens en société vivent dans le cadre que le fait social impose aux gens, laissant de côté certains intérêts personnels et visant le bien commun, permettant aux individus d'une société de vivre ensemble. Le sociologue face aux faits sociaux comme n'importe quel scientifique face aux problèmes pratiques qu'ils étudient. Inspiré par la vision des faits sociaux comme une chose de Claude Bernard, Durkheim dit qu'il est nécessaire de mettre de côté les "prénotions" dans l'étude des faits sociaux¹⁶. Voici une règle importante qu'il faut observer, selon Durkheim : un fait social ne peut s'expliquer que par un autre. La sociologie doit être basée sur un raisonnement sur d'autres faits sociaux. Il est clair que la relation entre les faits sociaux ne peut pas être simplement spéculative mais doit être fondée sur des preuves. Au lieu de mener des recherches en laboratoire comme les autres sciences, Durkheim compense en mettant en évidence les relations statistiques entre les faits de société. En statistique, cette relation entre les faits sociaux est appelée corrélation. Ces corrélations prouvent en outre qu'il existe une relation de causalité entre les faits sociaux. C'est cette approche de Durkheim qui démontre l'importance des statistiques en sociologie.

Les faits sociaux ont plusieurs caractéristiques. D'abord, ils sont hors de l'individu, il existe hors de l'individu, son origine est dans la société. Ces faits sociaux sont indépendants des comportements de l'individu, les individus peuvent ou non leur obéir, violer ces faits sociaux comporte le risque de sanction. Deuxièmement, c'est la contrainte de faits sociaux sur l'individu, et enfin la répétition, qui se répète dans le temps.

A.2 Les poids des facteurs sociaux sur les émotions, les sentiments.

Comme nous l'avons mentionné ci-dessus, les faits sociaux sont les façons dont les pensées, les actions et les sentiments sont affectés et les liens de la société avec l'individu sont exprimés. Et selon Durkheim, la société existe selon elle-même, donc elle aura des mécanismes fonctionnant selon les caractéristiques de chaque société. Il existe deux mécanismes généraux qui expliquent la fréquence à laquelle nous exprimons des émotions (autrement dit : les mécanismes sociaux qui affectent nos sentiments émotionnels) par la société dans laquelle nous faisons une partie.

La socialisation est le premier mécanisme. La socialisation est le processus d'interaction sociale continue par lequel les individus développent des capacités humaines et apprennent leurs modèles culturels.¹⁷ En d'autres termes, c'est le processus par lequel une personne acquiert continuellement des connaissances, la culture et la moralité afin que cet individu devienne membre de la société dans laquelle il vit. La socialisation existe deux façons pour les individus de s'intégrer dans la société dans laquelle ils vivent. Premièrement,

¹⁶ Massimo Borlandi (2017), *Sociologie et histoire : la contribution de Durkheim*, p.83-104

¹⁷ Macdonald John, J. (1987), *Sociologie*, p.154

la société transmet des normes obligatoires à chaque individu. La tâche des individus est de l'apprendre et de la pratiquer dans la vie. Par exemple, la connaissance à l'école, nous apprenons constamment à pouvoir l'appliquer à la vie ou au travail (Mécanisme institutionnel). Deuxièmement, les individus acquièrent naturellement des connaissances. Ce mécanisme est mis en œuvre par deux voies : l'imitation et la diffusion (Mécanisme non institutionnel). L'imitation est la reproduction, la répétition d'actions, de comportements, de façons de penser et de se comporter d'un individu ou d'un groupe d'individus. C'est l'un des moyens d'acquérir des expériences sociales, les individus utilisent ces modes pour choisir les actions, les connaissances et les comportements qu'ils jugent justes et intéressés. La diffusion est le processus par lequel les comportements sociaux passent naturellement d'une personne à une autre. La diffusion diffère de l'imitation en ce que les comportements sociaux se propagent même lorsque les individus n'imitent pas ou n'apprennent pas intentionnellement. La socialisation crée la personnalité, faisant en sorte que chaque individu ait des valeurs sociales et le respect des normes.

Les groupes sociaux créent des visions du monde basées sur les émotions et la morale afin de maintenir l'attachement entre les individus. L'attachement affectif est la prémisse de la solidarité. C'est aussi la raison pour laquelle le mécanisme de socialisation doit accompagner ce deuxième mécanisme, le contrôle social. Le contrôle social a été mentionné pour la première fois par Albion Woodbury Small et George Edgar Vincent en 1894¹⁸, mais à l'époque, les sociologues ne s'intéressaient pas vraiment au sujet. Quelques ouvrages typiques sur le sujet sont *Leviathan* de Thomas Hobbes, où il traite de l'ordre social et de la manière dont l'État utilise le pouvoir civil et militaire ; ou *On Crimes and Punishments* de Cesare Beccaria soutient que les gens éviteront de commettre un crime s'il était soumis à une punition plus sévère, démontrant que les punitions jouent un rôle dans le contrôle social¹⁹. Durkheim aborde également cette question dans son livre *De la Division du travail social* lorsqu'il évalue que le contrôle social est ce qui nous rend respectueux des lois dès en premier temps. La société applique certaines sanctions pour faire respecter une norme de comportement qu'elle considère comme juste.²⁰ Les individus et les organisations utilisent le contrôle social pour établir des normes et des règles sociales, et ceux qui les appliquent incluent les collègues, les amis, la famille, les institutions gouvernementales, les religions et les croyants, ou au sein de l'école ou de l'entreprise. Le but du contrôle social est de créer et de maintenir l'ordre social, d'éliminer les idées déviantes et de fournir des interventions appropriées à ceux qui ont des pensées déviantes. Les formes de contrôle social peuvent être divisées en deux types de base (mesures de contrôle informelles et formelles). Le mécanisme de contrôle social informel consiste à utiliser le processus de socialisation pour former des normes et des valeurs sociales chez les individus. ²¹Cette normalisation limitera le comportement de l'individu à un cadre socialement acceptable. Les mesures formelles de contrôle social sont des mesures punitives prises par des organisations puissantes de la

¹⁸ Hollingshead et al. (1941), *The Concept of Social Control*. *American Sociological Review*, p217-224.

¹⁹ Carmichael, Jason (2012), *Social Control*.

²⁰ Conley, Dalton (2017), *You may ask yourself: An Introduction to thinking like a sociologist*, p.197

²¹ Lindzey, Gardner (Ed) (1954) *Handbook of Social psychology. I. Theory and method. II. Special fields and applications*

société (gouvernement) pour empêcher le chaos dans la société. Selon Emile Durkheim, il considère ces mesures comme réglementaires. Les deux mesures ci-dessus constituent des niveaux de contrôle différents. Le sociologue Edward A. Ross soutient que les systèmes de croyance exercent un plus grand contrôle sur le comportement humain que les lois imposées par le gouvernement, quelle que soit cette croyance ou religion. Nous sommes passés par deux mesures de base du contrôle social, alors quelle en est la définition ? À l'origine, le concept de contrôle social faisait simplement référence à la capacité de la société à s'autoréguler depuis que la société s'est organisée. Dans les années 1930, cependant, le terme a pris un sens plus moderne et les chercheurs ont commencé à étudier la théorie du contrôle social comme un domaine distinct. Selon Merriam Webster, le contrôle social apparaît comme une certaine règle et norme dans la société.

La socialisation et le contrôle social visent à mettre les émotions et les sentiments d'un individu en harmonie avec les émotions appropriées dans la société. Cependant, il ne s'agit pas d'une liaison à sens unique. Nous contribuons et reproduisons également les schémas émotionnels de la société.²² Cette contribution est spontanée, mais aussi forcée. Les émotions et les sentiments deviennent une forme de langage social, sa fonction est d'aider les individus à communiquer avec les autres et avec eux-mêmes. Les facteurs sociaux qui imposent et lient les individus ont tous leurs propres buts et intentions, ou en d'autres termes, les facteurs sociaux ne portent pas un état neutre. Cela affecte l'individu, rendant l'initialisation initiale de l'émotion non neutre. En fait, nous voyons également que les deux mécanismes de socialisation et de contrôle social ont des éléments polarisants ou ont leurs propres intentions sur la façon de gérer et d'établir l'ordre social.

On voit donc ici l'empreinte d'événements culturels, de messages, d'articles publiés dans les médias, de rapports, de projets de lois du gouvernement ou encore de conférences sur Internet dans les écoles, tous ces facteurs contribuent au processus de socialisation par le biais de la transmission de contenus. Ces événements affectent les individus dans la société, créant en eux leurs propres pensées, opinions et sentiments. Exemple : Un article qui rejoue un souvenir douloureux ou haineux.

Maintenant que nous connaissons l'origine des émotions, examinons également l'histoire des émotions en termes sociologiques.

A.3 Histoire des émotions et des sentiments en sociologie.

Dans l'aspect sociologique, les émotions, les sentiments sont liés à la sensibilité d'un groupe social à changer dans le temps. Cela nous intéresse également à la distribution et à la variation des émotions dans les groupes sociaux au niveau *macrosociologique*.

La macrosociologie est le domaine d'étude qui se concentre sur la structure sociale à plus grande échelle. Habituellement, la macrosociologie analyse le système social dans son ensemble et se concentre également sur la population dans son ensemble. Grâce à la macrosociologie, nous pouvons entrer dans des concepts plus larges, contrairement à la microsociologie, et les études macro peuvent également être appliquées aux individus. C'est-à-dire que la macrosociologie est parfois l'étude de phénomènes individuels parce que

²² Mauss M (1950). *Rapport réels et pratiques de la psychologie et de la sociologie*

les individus et leurs interactions font partie d'un système social plus large. La macrosociologie traite de l'analyse statistique et intègre également des études empiriques pour arriver à des conclusions. De plus, les études macrosociologiques se concentrent davantage sur de vastes domaines, mais ces résultats peuvent également être appliqués à de petits phénomènes.

Les travaux d'Alexis de Tocqueville et d'Edward P. Thompson sur les émeutes de la faim et l'économie morale de la foule anglaise aux XVIIIe et XIXe siècles montrent les origines des affects et la réponse sociale face aux mouvements sociaux.

Alexis de Tocqueville, dans son ouvrage, a tenté de décrire la vie du peuple français aux XVIIIe et XIXe siècles, et l'influence d'institutions puissantes sur les émotions et les sentiments sociaux. Il a constaté que l'ancienne institution se caractérisait par une stabilité socio-émotionnelle, alors qu'en démocratie, il y avait une variabilité socio-émotionnelle marquée. Cela témoigne du lien entre le contexte historique, les institutions politiques et les sensibilités émotionnelles dans la société. De plus, les émotions sont aussi la cause de changements dans la société, généralement des périodes de rébellion (l'ancien français est lié à terme émotionnel émeute). Hypothèse : Les gens ne se sont pas révoltés alors que leur situation était vraiment mauvaise en 1787, mais lorsqu'ils ont comparé leur situation avec ce qu'ils avaient prévu, ils se sont sentis frustrés par le système en place, alors, il est nécessaire de récupérer leurs pouvoirs.

Dans son travail, Thompson dit que les émotions sont sous-tendues par des systèmes de valeurs émotionnelles. Cela est évident dans les situations où la justice sociale, la justice économique est menacée lors d'émeutes ou de mouvements de protestation. Selon Thompson, la transition du sentiment à l'émotion est due à la dimension énergétique du sentiment. Les sentiments forment alors un lien étroit entre contenus culturels et relations sociales. Les sentiments sont l'aspect profondément intériorisé de l'action, là où la culture bat son plein. Selon Karl Marx, après avoir observé la Révolution française de 1848, les révolutionnaires étaient animés d'une émotion forte, passionnée mais spontanée. Ces émotions sont le résultat d'un système social inégalement réparti.

L'histoire de la sociologie des émotions a été largement esquissée par les sociologues à partir des grandes mutations sociales. Car lorsque les émotions sont les plus intenses, c'est à ce moment-là qu'on va les laisser s'exprimer en paroles et en actes (les Révolutions sont un exemple de résurgence de la frustration).

La sociologie des émotions est un aspect complexe de la sociologie, l'étude des émotions, le mouvement des émotions et les résultats de ces émotions. Max Weber a essayé un jour de retirer l'élément émotionnel de la sphère d'influence sur l'action sociale. Il ne considère pas l'émotion comme un concept central de sa sociologie même s'il n'ignore pas ses manifestations dans la vie religieuse et politique. Cela montre la confusion dans ses arguments, alors que même si les émotions sont la dernière limite pour juger d'une action sociale, les émotions se glissent encore dans les arguments sur l'action sociale de Weber.

La sociologie des émotions examine les activités et les relations entre la vie sociale et la dynamique des émotions, ainsi que les conséquences de l'expression émotionnelle (reconstruction, changement social). Et l'analyse des émotions et des sentiments nous aide à

comprendre le lien entre les normes sociales et les valeurs, comprenant ainsi la situation sociale. L'application de méthodes d'analyse émotionnelle simplifiera ce processus d'analyse complexe. Nous allons d'abord découvrir l'environnement dans lequel nous exprimons nos sentiments et nos émotions, le réseau social.

B. Réseaux Sociaux.

Les réseaux sociaux, selon la manifestation directe de la réalité, regroupent les sites d'information électronique numérisés (weblogs, blogs, sites web) ou « sites web » en abrégé (en ligne), les journaux en ligne, les archives de données d'information, etc. ou organisations (y compris l'État), nationales ou internationales, est reçue et transmise sur le système mondial de lignes de transmission à fibre optique (Internet).

Normalement, dans la conscience quotidienne de la majorité des gens, les réseaux sociaux sont compris comme un environnement d'information et de communication nouveau, moderne et spécifique. Il comprend canaux, flux d'informations, nouvelles spécifiques sur tous les domaines de la vie sociale en général. Ce contenu d'information est basé sur et associé à la plupart des installations médiatiques, physiques - techniques et technologiques détermination respective. L'information et la communication des réseaux sociaux sont différentes des canaux, des flux d'information et des médias, des bases physiques - techniques et technologiques des médias de masse traditionnels, orthodoxes et officiels, tant en termes d'information que de communication. Le contenu de l'information reflète - la propagande, à la fois sur les médias et les moyens de communication.

Les sites Internet, les journaux en ligne, ... sont majoritairement extérieurs au système d'information - communication animé, dirigé et régulièrement géré par les partis et l'Etat. Par conséquent, bien sûr, ils ont un contenu qui n'est pas seulement simplement positif, mais aussi négatif. L'article se concentre sur la mention et la considération principalement de cet aspect des réseaux sociaux.

Les réseaux sociaux ont un très haut degré d'indépendance, car en termes de science et de technologie informatiques, art - technologie de communication, il est associé au système Internet international et mondial.

Par conséquent, même dans les cas où la gestion étatique a été menée au maximum, la plus drastique est le dépistage régulier, ou l'élimination complète de sa présence au niveau national, il existe toujours objectivement à l'échelle internationale et mondiale. De plus, avec l'élan de développement et les progrès de la science de l'information et des communications modernes dans la révolution technologique 4.0 d'aujourd'hui, le temps viendra bientôt pour que l'internet sans fil (c'est-à-dire le wifi, même à partir de satellites en orbite autour de la planète) devienne exactement comme la radio micro-ondes. C'est-à-dire que les informations sur les réseaux sociaux seront fournies, couvertes par une gamme suffisamment large et suffisamment bon marché au coût civil. A cette époque, sauf perturbations, ingérences locales, le fonctionnement des réseaux sociaux sera transnational, supranational, liberté quasi

absolue, chacun pourra librement accéder, poursuivre l'approche ainsi que vice versa consiste à poster, publier et diffuser des informations.

La grande indépendance des réseaux sociaux est également due au nombre extrêmement élevé de sujets sources d'information - individus ou organisations participant à la diffusion du réseau. Leur adresse (admin) se déforme, se transforme, se décale, change souvent. Cela découle également directement des fonctionnalités, des effets et des utilités particulièrement élevés du système Internet mentionné ci-dessus. Par conséquent, la surveillance, l'examen, la censure, le blocage ou la gestion générale du sujet et du contenu informatif d'un réseau social nécessite la mise en place d'énormes ressources humaines et matérielles. Particulièrement pour les sites d'information et les journaux

Réseau de langue français, ou de l'étranger en général, en pratique, le contrôler, l'éliminer complètement et complètement est clairement impossible. Le concept général d'un réseau social peut comprendre trois aspects principaux comme suit : Premièrement, sur la base technique - technologie de communication, il s'agit d'un moyen de communication de masse général, polyvalent, multi-usagers, avancé, moderne. Deuxièmement, à propos du contenu de l'information qui reflète - propage, il s'agit d'un canal de communication de masse spécial et spécial mixte, complexe, pluraliste et multidimensionnel. Troisièmement, à propos du sujet social, ce n'est pas un environnement purement information-communication "sans propriétaire", ou "humanité commune", de toute la communauté, est objective, impartiale, ouverte, fournissant des informations au service de tous, pour tous. En fait, le sujet des réseaux sociaux est une multitude non seulement d'individus, mais aussi d'organisations nationales et internationales très spécifiques. Et ils ont toujours leurs propres buts, objectifs, intentions et tâches dans de nombreux aspects (y compris la politique).

En tirant parti de la faculté d'informatique et de l'industrie de l'ingénierie - technologie de communication moderne, les réseaux sociaux concentrent, synthétisent, intègrent et simultanément exécutent et pratiquent les fonctions de presque tous les autres médias et médias civils disponibles :

- Comme un type de journal imprimé qui peut mettre à jour de nouvelles informations instantanément, en continu, d'innombrables fois dans n'importe quel terme ; transmet et libère des informations à des vitesses quasi instantanées de la lumière, à des distances maximales à mi-chemin autour du monde et une couverture mondiale maximale avec des hauteurs arbitraires de demande normalement raisonnable (plafond plafonnier de l'aviation civile).
- Comme le dit un type de journal, à la fois la transmission filaire et la transmission radio audio avec une vitesse, une distance, un emplacement et une capacité de diffusion d'informations illimités. Mais elle est aussi supérieure aux radios conventionnelles en ce qu'elle offre en même temps aux auditeurs des "cassettes et disques originaux" qui peuvent être stockés et reproduits à volonté, facilement et gratuitement.
- En tant que type de journal, à la fois fixe (photo) et animé (vidéo) peuvent être obtenus, en particulier pour la distribution en ligne, simplement et rapidement ; peuvent être stockées et reproduites un nombre incalculable de fois, instantanément et facilement.

- En tant que bibliothèque, ou centre d'information généraliste avec un contenu informatif très riche, diversifié et actualisé ; être arrangé, arrangé avec souplesse, systématiquement ; peuvent être consultés et exploités facilement, fréquemment, rapidement et facilement stockés et reproduits. En tant que référentiel d'informations doté d'une capacité, d'une étendue et d'une profondeur exceptionnelles ; peut être consulté, utilisé couramment, souvent.
- En tant qu'éditeur, éditeur et commerçant de livres électroniques indépendant, autogéré et indépendant ; a un nombre illimité de pages, d'impressions et de titres. En tant que salle de rédaction, ou le seul comité de rédaction pouvant assurer la fonction d'échange en ligne, c'est-à-dire l'interaction directe, immédiate, d'information publique entre le journal et n'importe qui, et devant tout le monde.

Les réseaux sociaux contribuent au développement de la conscience, de la réflexion et des compétences de vie des personnes :

Les réseaux sociaux deviennent de plus en plus un lieu d'information et de connaissance sur tous les domaines de la vie sociale. En quelques étapes simples, les utilisateurs recevront toujours des mises à jour opportunes sur les domaines et les problèmes qu'ils souhaitent surveiller. Les aidant ainsi à saisir les tendances de la vie pour leur travail et leur vie. De plus, sur les réseaux sociaux, il existe de nombreuses pages qui enseignent des compétences de vie telles que les langues étrangères, la cuisine, la réparation, la communication, la psychologie, le sport, etc., aidant les utilisateurs à acquérir sans problème les compétences de base nécessaires à la vie moderne. Suivre des cours ou payer des frais de scolarité.

Les réseaux sociaux contribuent activement au développement de la culture communautaire. La culture des médias sociaux fait partie de la culture communautaire et a une influence croissante sur la culture communautaire. Grâce à l'application des progrès de la science et de la technologie, les réseaux sociaux permettent aux utilisateurs de se connecter et d'interagir avec leurs amis, leur famille et la communauté de manière de plus en plus pratique. Les utilisateurs peuvent facilement partager leurs sentiments, leurs joies, leurs peines... avec la communauté. La participation individuelle aux affaires communautaires est également encouragée. Les ressources internes de la communauté sont valorisées plus efficacement dans le développement socio-économique. Les formes de commerce en ligne sur les réseaux sociaux des particuliers et des entreprises sont de plus en plus développées et professionnalisées.

C. Analyse des sentiments (Approche de l'apprentissage de la machine).

Comprendre quelqu'un, comprendre que le sens de ses phrases lors de la communication fait partie des questions qui existent depuis longtemps, et de nombreux travaux de recherche font référence à cette question. Surtout le développement des sites internet 2.0, il est plus facile de communiquer en société, d'analyser et de comprendre l'intention de la phrase, ou l'attitude de l'adversaire devient encore plus utile.

L'interaction croissante dans la société conduit à un lien étroit entre les sciences sociales telles que les sciences sociales, l'informatique et les mathématiques. Avec le développement de la science, des codes open source sont développés, les données textuelles sont facilement enregistrées sur le réseau. Avec ce développement, les sociologues disposent également d'outils et de moyens pour faciliter la recherche analytique. L'analyse des sentiments est une méthode automatique de traitement du langage qui tente d'analyser, de formater et de catégoriser les émotions et les attitudes d'un texte. Cependant, les avantages de cet outil ne s'arrêtent pas là. De nombreux scientifiques utilisent l'analyse des émotions, l'observant en termes de densité et de répétition des sentiments pour soutenir d'autres disciplines²³.

Cela signifie également que les données traitées doivent avoir une certaine opinion sur un certain sujet ²⁴. Mais en fait, c'est plus difficile que de distinguer le nombre de phrases ou le but fondamental et pur d'une phrase, comme une déclaration sarcastique, en particulier pour les ordinateurs. L'analyse des sentiments est une combinaison de traitement du langage naturel et de compréhension de ce langage en se concentrant sur la sémantique et les émotions actuelles.²⁵

Goncalves et al (2013) ont montré qu'il existe actuellement deux manières principales d'analyser les sentiments des textes. ²⁶La première, la plus ancienne, s'appuie sur l'approche lexicale. Le principal problème avec cette approche est de savoir si la liste de vocabulaire que nous générons n'apparaît dans aucun contexte, ce qui peut créer des problèmes lorsque nous rencontrons un mot ambigu. L'analyse des sentiments basée sur le glossaire montre également des difficultés lors de la rencontre de mots avec des connotations ironiques. Cette méthode est basée sur une liste de mots et les combine ensemble.

La deuxième méthode, qui sera implémentée dans ce mémoire, est basée sur l'approche par corpus et est principalement appliquée en apprentissage automatique. Cette méthode se concentre principalement sur le développement d'un système capable de classification sans intervention humaine. Ensuite, ce système utilise généralement une validation binaire qui indique si les données sont positives ou négatives. Le but ultime de ce processus est la capacité de traiter rapidement d'énormes volumes de données, ce qui laisse au chercheur plus de temps pour d'autres activités. La caractéristique remarquable de cette méthode est le développement du système dans un certain contexte. Pour ce faire, il faut d'abord trouver un échantillon de source de données qui est lié au bloc de données textuelles à analyser comme des sites de données spécialisés, ou la source de données est sur la même plateforme que les données à analyser. À partir de là, les sources identifiées lors du processus de sourcing sont examinées et les documents jugés pertinents sont compilés dans un entrepôt de données. Enfin, ce corpus est utilisé pour entraîner un système automatisé qui contient des documents dans le même domaine. Nous pouvons voir qu'avec des procédures aussi

²³Mauss M (1950). *Rapport réels et pratiques de la psychologie et de la sociologie*

²⁴ Pang, B. and Lee, L. (2008) *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval

²⁵ Cambria, E., and White, B. (2014). *Jumping NLP curves: A review of natural language processing research*. IEEE Computational Intelligence Magazine, p48-57

²⁶ Goncalves et al., (2013). *Comparing and combining sentiment analysis methods*. In Proceedings of the first ACM conference on Online social networks (pp.27-38)

rigoureuses, cette méthode donnera des résultats plus précis que l'approche par champ lexical. Presque toutes les données d'origine sont brutes, sans annotations ni étiquettes. Par conséquent, de nombreuses données doivent être étiquetées ou annotées manuellement. Cela dit, nous n'aurons presque jamais les évaluations de données les plus complètes et les plus précises ²⁷.

C.1 Classifier les données.

Pour pouvoir classifier les nouvelles données, nous devons identifier le type de sentiment actuel. Cela se fait d'abord manuellement. La personne analyse la phrase et lui attribue une étiquette qu'il juge appropriée. Ces personnes peuvent également choisir des fonctionnalités spécifiques en fonction de la situation ou du contexte que la personne juge raisonnable. Ces données étiquetées ont été analysées à trois niveaux différents (Benamara et al., 2017) : Déterminer le sentiment dans son ensemble à travers le document, au niveau de la phrase, et enfin au niveau de l'aspect bord. Dans ce cas, un système est conçu pour tenter de déterminer le sentiment associé à un mot, un groupe de mots ou une phrase. Cette approche aidera à identifier les facteurs négatifs au niveau mondial plutôt que les facteurs positifs. Autrement dit, lors de l'évaluation d'une phrase avec des connotations négatives, le système évalue en fonction du contexte, trouvera le contexte sémantique de phrases similaires qui ont été étiquetées. Cela aidera le système à MOINS omettre les phrases sarcastiques.

Avec l'exemple suivant, nous observerons l'évaluation du système de classification :

Ex1 : *J'ai reçu trois doses du vaccin Covid, ce qui est une chose merveilleuse. Bien qu'il y ait une légère douleur dans la main et de la fatigue pendant 4 jours, cela affecte le travail, maintenant ça va. J'en suis très satisfait.*

Certains systèmes évaluent cela comme une phrase positive avec deux points positifs et deux points négatifs. Cependant, j'aimerais développer un système qui prendra en compte tous les aspects, par exemple ici, un système d'analyse d'aspect peut être complètement étiqueté négatif, car la vaccination fatigue le tweeter et affecte sa vie quotidienne, en particulier son travail. Le système évalue que les douleurs aux mains, la fatigue, affectant le travail sont extrêmement importantes et il est déraisonnable pour le commentaire de prétendre qu'il est très satisfait de ce vaccin. Ou le système juge cela comme une phrase sarcastique.

En fait, les études dans ce domaine ont montré qu'il existe de nombreuses possibilités dans l'analyse et la sélection des caractéristiques. Cela signifie qu'il existe de nombreuses méthodes pour analyser le sentiment. L'une des façons consiste à regrouper des mots avec la même caractéristique, puis à les encoder en vecteurs avec des caractéristiques similaires. Ou il existe une méthode utilisant un classificateur linéaire comme Bayésien, un classificateur avec le principe d'entropie maximale et une machine à vecteurs de support (SVM). Et la méthode du sac de mots est souvent combinée avec le support de la machine vectorielle. Il y a quelques chercheurs qui ont mis au point des méthodes tout à fait remarquables. Pang et Lee, 2004 ont utilisé la méthode des coupes minimales (minimum cuts) pour analyser le sentiment dans le texte. Cette méthode aide les chercheurs à découper chaque couche sémantique de

²⁷ Boullier, D., & Lohard, A. (2012). *Opinion mining et Sentiment analysis*. S.l. : OpenEdition Press

mots, à la recherche de relations entre les mots ou les phrases dans le texte. La méthode est efficace pour analyser des textes dans différents contextes.

C.2 Applications de l'analyse de sentiments.

L'arrivée du Web 2.0 a changé le monde des médias sociaux. Non seulement en termes de médias sociaux en ligne, de partage d'informations et d'opinions, mais aussi en termes d'affaires, de commerce et d'études de marché. Le nombre d'utilisateurs de médias sociaux augmente de jour en jour et, selon les estimations de 2019, il y a 2,77 milliards d'utilisateurs de médias sociaux dans le monde. Sur les réseaux sociaux, les utilisateurs partagent toutes sortes d'informations telles que du texte, des images, des vidéos, des sons. Les médias sociaux contiennent une grande quantité de données précieuses pour d'autres domaines, et pour trouver et comprendre ces valeurs, l'analyse des sentiments devient un outil puissant pour aider les professions à prendre des décisions juste dans la stratégie à l'avenir.

L'application de l'analyse des sentiments est extrêmement large, des affaires, du commerce, de la politique à la santé et aux affaires publiques. Il fournit des applications qui aident à la prise de décision.

L'analyse des sentiments peut être appliquée à des événements mondiaux tels qu'un événement mondial majeur, une activité sportive ou une catastrophe en cours. Avec de tels problèmes publics, l'analyse des sentiments nous aide à observer l'étendue de l'impact de l'événement sur la société, ou vice versa, le niveau d'intérêt social, ou les idées et opinions diffèrent surgissent dans la société par rapport à cet événement. Cela conduit également à une analyse des opinions politiques d'un individu, d'un groupe d'individus ou d'une société. Le dernier article de recherche d'Amanda Jackson, publié sur Medium sur le conflit entre la Russie et l'Ukraine, a de nombreuses opinions sur la légitimité et le soutien des deux parties. Selon l'étude, plus de tweets soutiennent l'Ukraine que la Russie, et la majorité des tweets proviennent des pays occidentaux, qui sont considérés comme les plus durement touchés par cette guerre. L'analyse des sentiments nous a également aidé à voir les opinions politiques divergentes qui sont présentes dans les pays occidentaux, lorsque des messages pro-russes apparaissent.

À partir de l'analyse des sentiments, nous pouvons également évaluer la capacité ou le niveau de sensibilisation de la société à un certain problème, tel que la sécurité ou la sécurité des informations des utilisateurs. Cela aide les agences concernées, ou plus largement le gouvernement, à prendre des mesures pour s'adapter aux changements de la société et de la technologie, ainsi qu'à remodeler les normes et les perceptions sociales si nécessaire.

Dans le domaine des soins et de la santé publique, l'analyse des sentiments est utilisée comme cadre de service et d'évaluation. Par exemple, pendant une période épidémique, l'analyse de la densité des émotions et des sentiments nous renseignera sur la situation de cette zone, ou mesurera la probabilité et le risque d'une épidémie de cette zone.

Enfin dans le domaine du business et du commerce, l'analyse des sentiments aide à étudier les commentaires des clients sur les produits, à partir desquels les entreprises peuvent mettre à jour l'état actuel des produits, de la qualité des produits au stock de produits, c'est dans quelques magasins. Cela a été montré dans une étude, qui a conclu des opinions et des expériences sur les médicaments et les produits cosmétiques sur les réseaux sociaux. A partir

de là, les entreprises peuvent réfléchir à la création d'une application téléphonique, ou d'une plateforme e-commerce qui leur est propre pour optimiser la qualité des produits, ainsi que les étapes allant de la production à l'expérience utilisateur, en utilisant la technologie API open source, collecter des informations, expérimenter, les avis des clients, puis analyser leurs sentiments et leurs attitudes. Les entreprises d'ici peuvent créer leur propre écosystème.

D. Méthode d'analyse des sentiments.

Comme nous l'avons mentionné dans la section A, il existe de nombreuses méthodes pour aborder, traiter et analyser le sentiment. Nous comprenons tous que l'objet principal de l'analyse des sentiments, c'est nous, et comme d'autres sciences à facteurs humains comme la sociologie, la médecine, le but principal est de nous aider à mieux nous comprendre, satisfaisant ainsi notre curiosité et nous aidant à avoir une vision plus confortable. la vie. L'analyse des sentiments est donc un domaine de recherche complexe, c'est une combinaison de la théorie, des résultats de la sociologie et de la mise en œuvre de l'apprentissage automatique, du traitement du langage naturel au-delà de l'intelligence artificielle. Et pour mieux appréhender cette méthode complexe, nous apprendrons tour à tour la NLP, la méthode d'analyse des sentiments ABSA, et l'outil d'analyse des sentiments CamemBERT de la langue française.

D.1 NLP (Natural Language Processing).²⁸

L'une des volontés fortes, apparue très tôt chez les informaticiens en général et l'intelligence artificielle en particulier, est de réussir à construire des systèmes et des programmes informatiques capables de communiquer avec les gens à travers le langage naturel, c'est-à-dire le langage que les gens utilisent pour communiquer au quotidien, pas un langage de programmation ou un langage machine de bas niveau. Le traitement du langage naturel est une branche de la recherche sur l'intelligence artificielle qui développe des algorithmes et construit des programmes d'apprentissage automatique capables d'analyser, de traiter et de comprendre le langage humain.

Le traitement du langage naturel se concentre sur l'étude des interactions entre les ordinateurs et le langage naturel humain, sous forme de parole ou de texte. L'objectif de ce domaine est d'aider les ordinateurs à comprendre et à effectuer efficacement des tâches liées au langage humain telles que : l'interaction homme-machine, améliorer l'efficacité de la communication homme-machine, ou simplement améliorer l'efficacité du traitement et de l'analyse informatique.

Le traitement du langage naturel est apparu dans les années 40 du siècle dernier, en passant par de nombreuses méthodes différentes telles que l'utilisation de modèles d'automatisation et probabilistes, les méthodes basées sur des symboles, les méthodes stochastiques, l'utilisation de l'apprentissage automatique traditionnel, et surtout l'apprentissage en profondeur au cours de la dernière décennie.

²⁸ Christopher Manning and Hinrich Schütze (1999). MIT Press. *Foundation of Statistical Natural Language Processing*.

Le traitement du langage naturel se divise en deux grandes branches, ces deux branches ne sont pas totalement indépendantes, le traitement de la parole et le traitement du texte. Le traitement de la parole se concentre sur le développement et la recherche d'algorithmes et de programmes informatiques qui traitent le langage sous forme parlée, c'est-à-dire, y compris la reconnaissance vocale, la synthèse vocale et la conversion parole-texte, et vice versa.

Le traitement de texte se concentre sur l'analyse de données textuelles. Les applications importantes du traitement de texte comprennent la recherche, la recherche d'informations, la traduction automatique, la synthèse automatique de texte, la vérification des erreurs et l'analyse du contenu du texte. Dans le traitement de texte, nous pouvons diviser en deux tâches principales : la compréhension de texte et la génération de texte. Alors que la compréhension de texte a pour tâche l'analyse, la génération de texte traite de la création de nouveaux textes tels que des traductions ou des résumés de contenu.

Le traitement de texte comporte quatre étapes. La première étape est l'analyse morphométrique. Dans cette étape, la tâche principale de l'apprentissage automatique est de reconnaître, d'analyser et de décrire la structure d'un morphème de langue donné tel que les mots racines, les prépositions, les affixes, les types de mots, etc. segmentation et étiquetage des parties du discours. La deuxième étape - l'analyse - est le processus d'analyse d'une séquence de symboles, soit en langage naturel, soit en langage informatique selon une grammaire formelle. La grammaire formelle comprend : la grammaire sans contexte, la grammaire catégorielle combinatoire et la grammaire des dépendances. L'entrée de l'analyse est une phrase composée d'une séquence de mots et des étiquettes de leurs types, et la sortie est un arbre d'analyse représentant la structure syntaxique de la phrase. La troisième étape est l'analyse sémantique. Cette étape associe des structures sémantiques, du niveau du mot, de la phrase, de la clause, de la phrase et du texte entier, à leurs significations indépendantes. Il s'agit de trouver la sémantique du mot d'entrée, et comporte deux niveaux : la sémantique lexicale, en distinguant la sémantique des mots ; La sémantique de la composition concerne la manière dont les mots se lient pour former des significations plus larges. La dernière étape est l'analyse du discours, qui considère la relation entre la langue et le contexte d'utilisation. L'analyse du discours est donc effectuée sur la portée du passage, et sur l'ensemble du texte. Comme nous l'avons vu, avec tant de tâches, d'étapes et de sous-domaines, le traitement du langage naturel est un domaine complexe. Par exemple : Vaste vocabulaire, régulièrement mis à jour, structure grammaticale flexible et parfois tout le monde n'utilise pas la grammaire correcte, langage exprimant le sens et les sentiments de l'écrivain.

L'analyse des sentiments du texte, comme mentionné ci-dessus, est également l'une des tâches importantes du traitement du langage naturel. Il s'agit d'une tâche complexe car elle implique des émotions, ce qui conduit à la complexité de la sémantique du contexte. L'une des solutions inventées par les scientifiques est l'ABSA (Aspect based Sentiment Analysis).

D.2 Comprendre la méthode ABSA.²⁹

²⁹ Abderrahman Essebbbar et al. (2021), *Aspect Based Sentiment Analysis using French Pre-Trained Models*

ABSA-Aspect Based Sentiment Analysis, Analyse des sentiments basée sur les aspects en français- est une méthode d'analyse des attitudes et émotions des clients largement utilisée ces dernières années. Il permet d'indiquer les aspects de satisfaction et les aspects à améliorer en fonction de l'expérience client (Pang et Lee, 2008).

ABSA fait partie du traitement du langage naturel et excelle à fournir des informations plus contextuelles qu'une analyse de sentiment typique. Une chose à noter dans ce domaine est que les gens peuvent porter des jugements et des commentaires sur différents aspects de leur public cible. Par exemple, lors de l'évaluation d'un restaurant, les gens peuvent évaluer de nombreux critères dans ce restaurant, tels que la qualité de la nourriture, l'attitude du service, l'espace, etc. Ainsi, une analyse de sentiment normale ne peut pas répondre à l'analyse dans de nombreux aspects et contextes.

ABSA nous aide à séparer le texte en différents aspects, puis à donner une évaluation de chaque aspect sur les niveaux de sentiment : Positif, Négatif ou Neutre.

Un exemple du fonctionnement d'ABSA : Un commentaire de critique de restaurant : "Pourtant les plats sont bons et la déco est sympa". ABSA sépare deux aspects du commentaire, la nourriture et l'espace.

[Catégorie : Plats#Qualité, étiquette : Plats, label : Positive]

[Catégorie : Espace#General, étiquette : déco, label : Positive]

ABSA assume différents rôles dans le traitement de texte et l'analyse des sentiments. Le premier est l'extraction de la cible d'opinion pour trouver d'anciens objectifs, c'est-à-dire qu'ABSA trouvera des mots qui reflètent la même opinion que le mot donné. Deuxièmement, la détection des catégories de facettes vise à détecter différentes catégories apparaissant dans le texte. Enfin, il y a la classification des sentiments d'aspect (The Aspect Sentiment Classification). Dans le cadre du mémoire, nous nous intéressons à la classification des sentiments d'aspect.

Le modèle que j'ai choisi d'analyser ABSA est PTM (Pre-Trained Model, le modèle de Pré-Formé en français) qui est un type de modèle d'apprentissage automatique couramment utilisé aujourd'hui. Le modèle pré-formé est un modèle qui a été formé précédemment avec un grand ensemble de données ou avec des méthodes modernes qui réduisent l'effort de formation du modèle à partir de zéro. Le modèle est ensuite formé pour s'adapter à l'ensemble de données réel ou utilisé directement dans l'apprentissage automatique. Avec le développement des représentations de mots utilisées dans le traitement du langage naturel, la tendance actuelle est aux grands modèles de langage pré-entraînés tels que les informations complémentaires bidirectionnelles des représentations d'encodeurs bidirectionnels à partir de transformateurs (BERT).). PTM fournit un contexte pour les mots précédemment appris et réapprend les mots à partir de données non annotées. BERT est un modèle de langue anglaise pré-formé qui examine le contexte des mots. Et CamemBERT est la version française de BERT. Nous allons plonger dans le fonctionnement de PTM dans la section de Méthodologie.

Analyser les sentiments des statuts publiés sur Twitter à propos des vaccins Covid est notre objectif. Non seulement les lignes de statut positif méritent d'être notées, mais les mots négatifs sont toujours présents dans tout, et en particulier sur les réseaux sociaux où nous

pouvons complètement cacher nos identités. Ceux qui ne sont pas favorables à la vaccination Covid, ou plus largement, à la vaccination sont présents. L'anti-vaccination peut provenir d'une perte de confiance dans les vaccins, ou plus profondément d'une raison politique, mais elle existait aussi avant que l'épidémie de Covid-19 n'éclate.

E. Anti-vaccination.

Depuis la publication de la séquence génétique du coronavirus 2 qui cause le syndrome respiratoire aigu sévère, les scientifiques du monde entier ont travaillé plus vite que jamais pour développer un vaccin capable de lutter contre l'entrée et l'effet du coronavirus 2 sur notre corps. Comme d'habitude, il faut 10 à 15 ans pour faire passer un vaccin du laboratoire aux centres de vaccination. Pour un vaccin contre le Covid-19, il n'a fallu que 18 mois aux scientifiques pour achever ce processus. C'est une bonne nouvelle, mais cela soulève aussi des doutes et des inquiétudes. La rumeur veut que la pandémie soit un canular pour vendre des vaccins dans les médias. La pandémie de Covid-19 a soulevé l'importance de comprendre les facteurs politiques et sociaux et l'acceptation par le public des recommandations et réglementations gouvernementales strictes pour minimiser la propagation du virus. L'anti-vaccination n'est pas un phénomène nouveau, comme on l'a dit, l'idée est née d'un scepticisme à l'égard du premier programme de vaccination contre la variole. À l'échelle mondiale, cependant, de plus en plus de parents choisissent de retarder ou de refuser tout ou partie des vaccins de leurs enfants. Ceci est responsable de la réduction de l'immunité collective contre certaines maladies. Selon l'Organisation mondiale de la santé, l'hésitation à se faire vacciner ou le fait d'être forcé de recevoir une offre de partage de vaccins est l'une des dix menaces à la sécurité sanitaire mondiale. Par exemple, le programme de vaccination humaine contre le VPH a été remis en question en raison de rapports faisant état de divers symptômes après la vaccination dans les médias au Danemark et en Irlande. Cela a réduit le taux de vaccination contre le VPH de 85 % à moins de 40 % dans ces deux pays. Les médias peuvent diffuser intentionnellement/non intentionnellement de fausses informations qui font douter les parents de l'innocuité du vaccin contre le VPH.

Cependant, il faut faire la distinction entre le terme "anti-vaccination" et l'hésitation ou la suspicion vaccinale, car le niveau de confiance n'est pas le même.³⁰ Le concept de réticence à la vaccination représente un changement d'attitude de la confiance au doute, ou vice versa, à propos de la vaccination. Le groupe d'individus qui correspondent à la vaccination un groupe social est hétérogène en termes d'attitudes. Une personne qui hésite à se faire vacciner peut tout de même accepter ou refuser de se faire vacciner, ou refuser un vaccin A mais accepter un vaccin B. Cette situation traduit très logiquement une suspicion et des inquiétudes légitimes vis-à-vis de la vaccination. Par conséquent, l'hésitation à la vaccination est assez différente de l'activité anti-vaccination (contre toute forme de vaccination). La réalité a montré que nous nous retrouvons souvent à cheval sur plus de deux pôles d'attitudes vis-à-vis de la vaccination, tout le monde n'ose pas admettre qu'il est un anti-vaccin complet, et tout le monde ne le fera pas, dit-il avoir pleinement confiance dans

³⁰ MacDonald NE (2015). *Vaccine hesitancy: definition, scope and determinants*

l'innocuité de tous les vaccins. Le fait que le taux de vaccination général soit élevé ne signifie pas que la confiance dans le vaccin est élevée, en particulier dans les situations où il est impératif. Et bien sûr, accepter le vaccin ne signifie pas consommer le vaccin. L'acceptation d'un vaccin est définie comme l'individu ou le groupe qui décide d'accepter ou de refuser, lorsqu'il a la possibilité de vacciner, le public est informé des avantages et des besoins (ou éventuellement des effets secondaires). L'anti-vaccination fait également référence à des attitudes et des comportements très spécifiques d'un individu ou d'un groupe d'individus.

Le mouvement de l'anti-vaccination

Le mouvement anti-vaccination est souvent utilisé pour désigner ceux qui consacrent une partie de leur temps, de leurs efforts et de leurs ressources à exprimer leur inquiétude au sujet des vaccins, pour attirer d'autres personnes à rejoindre le mouvement avec le désir qu'il y ait des changements sur la question, ou la réglementation des vaccinations.³¹ De nombreux sociologues jugent le terme "anti-vaccination" trop négatif, car il suggère que les militants anti-vaccination rejettent la vaccination en général. Ils ont promu le terme de critiques de la vaccination, laissant entendre que les militants ne faisaient que critiquer les ingrédients du vaccin. Autre problème lié au mouvement anti-vaccination, on s'attarde souvent sur ceux qui critiquent les vaccins, mais qui agissent contre la vaccination, oubliant qu'il y a des partis politiques qui le soutiennent à d'autres fins (politique, groupe d'intérêts, ...) Peut-être que les vaccins ne sont pas ce dont ils se soucient vraiment, mais c'est un outil pour eux d'exprimer leurs opinions, d'attirer les gens au moyen de son argumentation.

Par conséquent, il est faux de considérer ceux qui suivent le mouvement anti-vaccination comme n'ayant aucune connaissance et des connaissances limitées.

IV. Méthodologie et Hypothèse.

A. Méthodologie de l'analyse des sentiments.³²

BERT signifie Représentations d'encodeurs bidirectionnels à partir de Transformers, qui est compris comme un modèle pré-formé, qui apprend des vecteurs de représentation contextuelle bidimensionnelle de mots, qui sont utilisés pour transférer des problèmes sur le traitement du langage naturel. Si PTM est un modèle d'analyse des sentiments linguistiques, la partie la plus importante de celui-ci est la façon dont le langage est traité, Transformers en fait partie. Si comparer PTM est une maison, les transformateurs ou d'autres méthodes de traitement du langage sont le cadre structurel de la maison PTM.

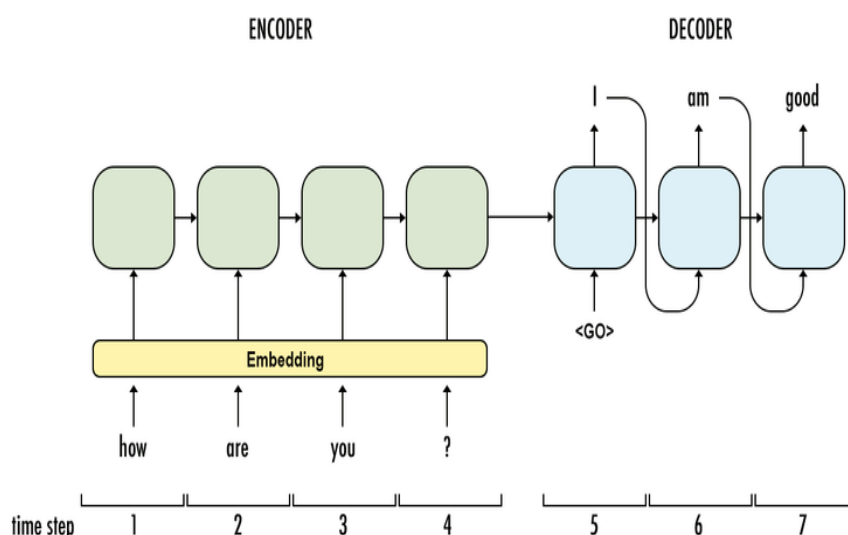
Avant que Google ne publie l'article Transformers, la plupart des tâches de traitement du langage naturel, en particulier la traduction automatique, utilisaient l'architecture des Recurrent Neural Networks (RNNs). La faiblesse de cette méthode est la lenteur du traitement et la difficulté à détecter le lien entre les mots dans la phrase. Les transformateurs sont nés pour résoudre ces problèmes, et la variante BERT qui est un state-of-the-art pour

³¹ Hobson-West (2007). *Trusting blindly can be the biggest risk of all: organized resistant to childhood vaccination in the UK.*

³² Données référencées d'*Attention Is All You Need* (2018) <http://nlp.seas.harvard.edu/2018/04/03/attention.html>

les tâches liées au traitement du langage naturel. D'abord nous découvrons comment fonctionnent les RNNs afin de comprendre la supériorité des Transformers.

A.1 Sequence-to-Sequence, le modèle utilise RNNs.

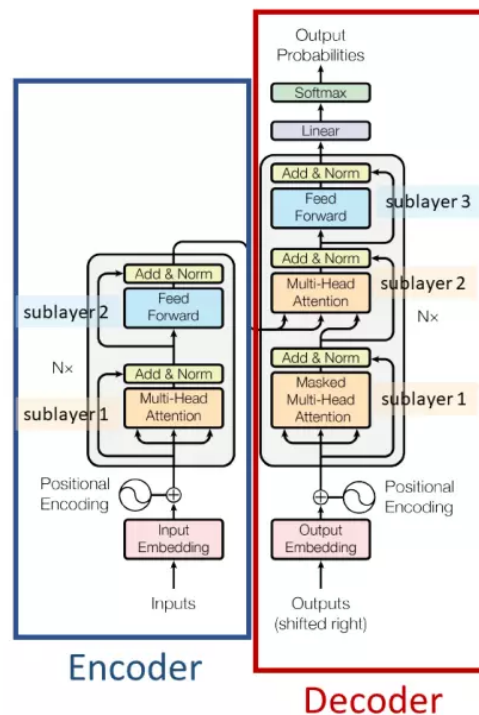


Le modèle Sequence-to-Sequence prend une entrée sous forme de séquence et renvoie une sortie sous forme de séquence. Par exemple, dans le problème Q&A ci-dessus, l'entrée est la question "comment allez-vous" et la sortie est la réponse "je vais bien". La méthode traditionnelle utilise des RNNs pour l'encodage et le décodage. La première faiblesse est la vitesse d'entraînement très lente. La deuxième faiblesse est qu'il ne gère pas bien les phrases longues en raison du phénomène de Gradient Vanishing / Exploding. Au fur et à mesure que le nombre d'unités augmente, le gradient diminue progressivement à la fin des unités, entraînant une perte d'informations et un relâchement des dépendances entre les unités. Explication : Qu'est-ce qu'un dégradé ? Le gradient est la dérivée d'une fonction utilisée pour représenter le taux de variation d'une fonction à un certain point. Le phénomène de Gradient Vanishing / Exploding se produit lorsqu'une longue phrase se compose de plusieurs unités, ce qui rend le calcul imprécis. Gradient Vanishing est le phénomène où les valeurs de poids ne convergent pas, et Gradient Exploding est le phénomène où au lieu de converger les résultats, ils se désintègrent.

En 1991, la mémoire à long terme (LSTM) a été introduite pour faire face aux phénomènes de gradient de fuite, lorsque le modèle a été ajouté à un panier pour stocker des informations, aidant à conserver les informations pour les longues phrases. Cependant, l'ajout de ce panier au modèle rend également le LSTM plus complexe et, bien sûr, la vitesse d'entraînement sera beaucoup plus lente que les RNNs.

En étant capable de tirer parti du parallélisme GPU pour accélérer la formation des modèles de traitement du langage, tout en surmontant les faiblesses du traitement des phrases longues, Transformers devient un état de l'art de l'industrie du NLP.

A.2 Transformers.



L'architecture des transformateurs utilise également deux parties, l'encodage et le décodage, qui sont assez similaires aux RNNs. La différence est qu'au lieu d'introduire lentement les données, avec les transformateurs, les entrées seront alimentées en même temps, et nous n'avons donc plus le concept de pas de temps comme dans les RNN. Alors, quel est le mécanisme alternatif pour la méthode de régression des RNN ? C'est l'auto-attention. Nous allons passer en revue cette architecture étape par étape avec la première partie, l'encodage.

Encodage.

1. Input Embedding

Les ordinateurs ne comprennent pas les mots, mais seulement les nombres et les vecteurs. Par conséquent, l'étape de conversion de la forme du mot en nombre, vecteur s'appelle Input Embedding. Cela garantit que les mots dont le sens est proche ont à peu près le même vecteur.

Input Embedding



2. Positional Encoding

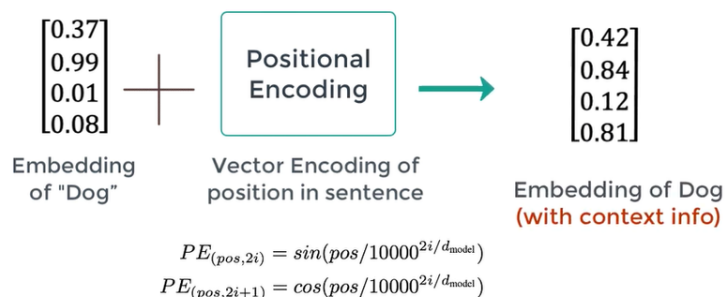
La conversion de mots en nombres et en vecteurs nous aide à représenter en quelque part la sémantique de ce mot. Cependant, le même mot mais à des positions différentes dans la phrase peut avoir des sens différents, par exemple le mot “grand” avant et après un nom peut avoir deux sens différents. Transformers a ajouté une partie Positional Encoding pour traiter plus d'informations sur la position d'un mot. On a la formule pour calculer le Positional Encoding d'un mot dans une phrase.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}})$$

Où, pos est la position du mot dans la phrase, PE est la valeur du i -ème élément dans le Embedding de longueur d_{model} . Et après, on ajoute le vecteur PE avec le vecteur Embedding.

Comme nous pouvons le voir, le même mot, mais si différentes positions dans une phrase, produiront des vecteurs différents, cette différence peut être définie par la sémantique de ce mot dans différents contextes.



3. Self-Attention

L'auto-attention (Self-Attention) est un mécanisme qui aide les transformateurs à comprendre l'association entre les mots d'une phrase. Par exemple, comment le mot « prends » dans la phrase « je prends la balle » se rapporte à d'autres mots. Nous voyons que le mot "prends" est lié au sujet "je", et bien sûr il est fortement lié à "prends" et "ombre" lui-même.

Avec Self-Attention, les entrées sont des vecteurs, il y a 3 matrices principales Querys (Q), Clés (K) et Valeurs (V). A partir de ces trois vecteurs, nous allons calculer le vecteur

d'attention Z , qui est le vecteur qui calcule la relation entre les mots. D'abord, pour obtenir les trois vecteurs Q , K , V , "input embedding" est multipliée par les trois vecteurs de poids correspondantes (générées lors de l'apprentissage) q (query), k (key) avec la dimension d_k ; v (value) avec la dimension d_v .

Calculez les scores d'alignement en multipliant l'ensemble des requêtes emballées dans la matrice, Q , avec les clés de la matrice, K . Si matrice, Q , est de taille $m \times d_k$ et matrice, K , est de taille, $n \times d_k$, alors la matrice résultante sera de taille $m \times n$. En conséquence, plus deux mots sont liés, plus le résultat sera grand, et vice versa:

$$QK^T = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1n} \\ e_{21} & e_{22} & \dots & e_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ e_{m1} & e_{m2} & \dots & e_{mn} \end{bmatrix}$$

Après, on met à l'échelle chacun des scores par $\frac{1}{\sqrt{d_k}}$:

$$\frac{QK^T}{\sqrt{d_k}} = \begin{bmatrix} \frac{e_{11}}{\sqrt{d_k}} & \frac{e_{12}}{\sqrt{d_k}} & \dots & \frac{e_{1n}}{\sqrt{d_k}} \\ \frac{e_{21}}{\sqrt{d_k}} & \frac{e_{22}}{\sqrt{d_k}} & \dots & \frac{e_{2n}}{\sqrt{d_k}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{e_{m1}}{\sqrt{d_k}} & \frac{e_{m2}}{\sqrt{d_k}} & \dots & \frac{e_{mn}}{\sqrt{d_k}} \end{bmatrix}$$

Ensuite, nous appliquons la fonction softmax pour obtenir l'ensemble des poids. Cela nous donnera une distribution de probabilité sur les mots.

$$\text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) = \begin{bmatrix} \text{softmax} \left(\frac{e_{11}}{\sqrt{d_k}} & \frac{e_{12}}{\sqrt{d_k}} & \dots & \frac{e_{1n}}{\sqrt{d_k}} \right) \\ \text{softmax} \left(\frac{e_{21}}{\sqrt{d_k}} & \frac{e_{22}}{\sqrt{d_k}} & \dots & \frac{e_{2n}}{\sqrt{d_k}} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \text{softmax} \left(\frac{e_{m1}}{\sqrt{d_k}} & \frac{e_{m2}}{\sqrt{d_k}} & \dots & \frac{e_{mn}}{\sqrt{d_k}} \right) \end{bmatrix}$$

Enfin, nous multiplions cette distribution de probabilité par V pour supprimer les mots inutiles et conserver les mots importants. Dans la dernière étape, nous additionnons les vecteurs V qui ont été multipliés par la sortie softmax, créant un vecteur Z d'attention pour un mot. En répétant ce processus pour tous les mots, nous obtenons la matrice d'attention pour une phrase.

$$\text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \cdot V = \begin{bmatrix} \text{softmax} \left(\frac{e_{11}}{\sqrt{d_k}} & \frac{e_{12}}{\sqrt{d_k}} & \dots & \frac{e_{1n}}{\sqrt{d_k}} \right) \\ \text{softmax} \left(\frac{e_{21}}{\sqrt{d_k}} & \frac{e_{22}}{\sqrt{d_k}} & \dots & \frac{e_{2n}}{\sqrt{d_k}} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \text{softmax} \left(\frac{e_{m1}}{\sqrt{d_k}} & \frac{e_{m2}}{\sqrt{d_k}} & \dots & \frac{e_{mn}}{\sqrt{d_k}} \right) \end{bmatrix} \cdot \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1d_v} \\ V_{21} & V_{22} & \dots & V_{2d_v} \\ \vdots & \vdots & \ddots & \vdots \\ V_{n1} & V_{n2} & \dots & V_{nd_v} \end{bmatrix}$$

4. Multi-head Attention

Le problème avec Self-Attention est que l'attention d'un mot sera toujours fortement liée à celui-ci. Cela est logique, car il est évident que le mot a à voir avec lui-même le plus.

Mais nous ne voulons pas cela, notre objectif lorsque nous essayons de comprendre et de créer ces formules complexes est de trouver des interactions entre différents mots dans une phrase. Par conséquent, au lieu d'utiliser un seul triplet, nous utilisons plus d'un triplet en même temps, et peut-être qu'avec plus de triplets d'attention, nous analyserons plus de parties différentes d'une phrase.

5. Feed Forward

Après normalisation, les vecteurs Z sont mis à travers un réseau entièrement connecté. Un réseau entièrement connecté est un réseau composé de plusieurs layer entièrement connectées. Les données normalisées seront ensuite entièrement connectées pour optimiser la précision des couches d'analyse. Ici, nous poussons les vecteurs Z pour optimiser la précision de la corrélation entre les mots dans la phrase.

Décodage et résultats.

Si le cryptage consiste à transformer des mots en un langage lisible par l'homme, le décryptage consiste à transformer des séquences de nombres en résultat souhaité. Notre problème ici est d'analyser les attitudes des utilisateurs (positifs ou négatifs aux vaccins contre le Covid). Ainsi, grâce au modèle PTM, nous avons les modèles de phrase et l'extension de chaque phrase disponibles, passons par le processus d'encodage, puis l'apprentissage automatique comme décrit ci-dessus, le processus de décodage est quand à partir des vecteurs numériques de ce modèle de phrase, il va trouver la relation entre les phrases qui ont souvent des éléments positifs (ou négatifs) et entraînez-vous à l'étiqueter. Ici, j'utilise 5000 échantillons de tweet couramment utilisés sur Twitter, avec 80% pour l'apprentissage automatique et 20% pour les tests, le résultat de précision est de 81% (un résultat acceptable en apprentissage automatique).

Ensuite, la question est de savoir pourquoi il n'y a que deux sortes d'opinions, des sentiments (positifs ou négatifs) sur le vaccin covid et non une opinion neutre. L'explication à cela est que lorsqu'une personne montre une attitude neutre envers le vaccin covid, elle n'aura que deux options pour la vaccination, l'une est de suivre l'ordre du gouvernement de vacciner, l'autre est non vacciné. Il ressort clairement des actions que l'on peut constater une polarisation des opinions individuelles envers les vaccins contre le Covid-19.

A.3 BERT et CamemBERT³³

CamemBERT est basé sur RoBERTa, qui s'améliore par rapport au BERT d'origine en identifiant les hyper paramètres clés pour des performances plus robustes.

Comme nous l'avons dit plus haut, le problème de mémoire que nous devons résoudre est de classer les messages Twitter en deux catégories selon les attitudes et les sentiments. Et peut-être qu'avec l'explication de Transformers, nous ne sommes toujours pas satisfaits car l'explication de la dernière partie : comment classer les textes n'est pas claire. Transformers est une structure fondamentale, et à partir de là, de nombreuses structures avec différentes tâches, et BERT en fait partie.

BERT signifie Représentations d'encodeurs bidirectionnels à partir de transformateurs, une nouvelle construction pour la classe de problèmes de représentation et de classification du langage. BERT est conçu pour former des vecteurs représentant des langages textuels à travers leurs contextes bidimensionnels (gauche et droite). En

³³ Jacob Devlin et al, BERT: *Pre-training of Deep Bidirectional Transformers for Understanding*

conséquence, le vecteur représentatif généré à partir du modèle BERT affiné avec des couches de sortie supplémentaires a abouti à de nombreuses architectures qui améliorent considérablement les tâches de traitement du langage naturel sans trop modifier l'ancienne architecture.

Avant d'aborder les mécanismes du BERT, il existe deux stratégies d'utilisation des représentations linguistiques pré-entraînées que nous devons connaître : feature-based et fine-tuning.

Feature-based est une approche qui crée des contextes de mots dans une phrase, puis construit un PTM.

Fine-tuning consiste à prendre un modèle pré-formé, en utilisant une partie des calques pour créer un nouveau modèle. Par exemple, si vous souhaitez catégoriser l'ambiance des critiques de films sur un site Web, vous pouvez obtenir des critiques de produits musicaux, puis modifier les "layers".

Architecture de BERT.

L'architecture du modèle BERT est une architecture de Transformers multicouches (section précédente).

Appelons L le nombre de couches Transformers utilisées avec des couches cachées de taille H et des têtes à la couche Attention à A. Dans tous les cas, la taille du filtre est toujours égale à 4H. Cela signifie que lorsque H=768 la taille du filtre est 3072 ou lorsque H=1024 la taille du filtre=4096. Ce résultat est principalement basé sur 2 tailles de modèles :

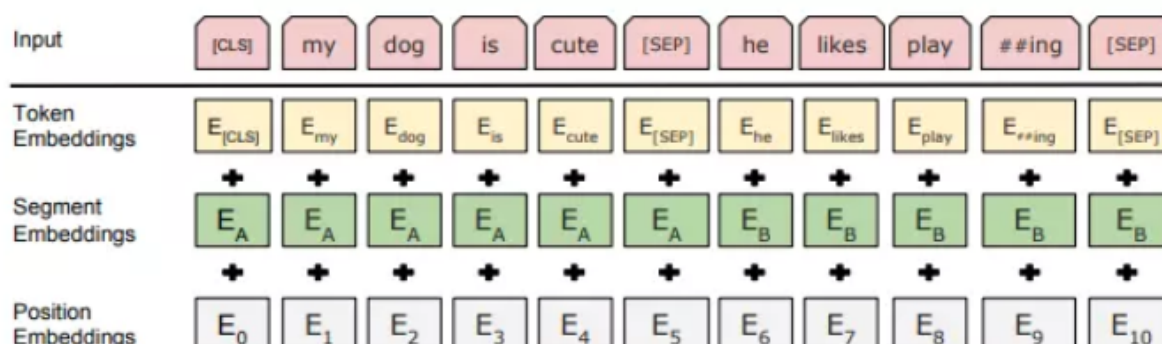
***BERT_{BASE}*: L=12, H=768, A=12, Total Parameters=110M**

***BERT_{LARGE}*: L=24, H=1024, A=16, Total Parameters=340M**

Input Representation.

Voyons maintenant comment l'entrée de BERT est représentée.

Notre exemple d'entrée est une phrase de texte définie sur une chaîne composée de mots. En tant que chaîne d'entrée, notre représentation est construite en additionnant ces jetons (tokens) avec le vecteur de segment, la position correspondante des mots dans la chaîne (phrase):



Nous voyons que la gestion des entrées de BERT est la même que pour Transformers. Il y a quelque chose de spécial à propos du jeton de valeur [CLS]. La sortie des transformateurs correspondant à ce jeton sera utilisée comme représentation des phrases dans les tâches de classification. Dans le cas de paires de phrases réunies en une chaîne, nous distinguons les phrases de deux manières. Tout d'abord, nous les séparons par un jeton spécial [SEP]. Deuxièmement, nous ajoutons un segment d'intégration pour la phrase A et un autre segment d'intégration pour la phrase B. Lorsqu'il n'y a qu'une seule phrase, nous n'aurons qu'un seul segment d'intégration.

Pre-training Tasks

Le modèle de formation pour BERT est basé sur deux tâches de prédiction non supervisées appelées Masked LM et Next Sentence Prediction.

Un modèle d'apprentissage en profondeur qui est appris sur la base d'un contexte bidimensionnel est beaucoup plus naturel et puissant qu'un modèle qui n'utilise qu'un contexte unidimensionnel (de gauche à droite ou vice versa). Auparavant, les modèles de langage ne pouvaient s'entraîner que de gauche à droite ou de droite à gauche. La raison en est que l'utilisation d'un contexte bidimensionnel crée le paradoxe qu'un mot peut se voir indirectement dans un contexte multicouche. Pour former un modèle de représentation basé sur un contexte bidimensionnel, nous masquons certains jetons d'entrée de manière aléatoire, puis nous prédisons et appelons ces jetons cachés, la tâche étant un "Masked LM" (modèle de langage masqué). Chez Google, les chercheurs ont caché 15% des jetons appartenant au dictionnaire WordPiece dans une séquence aléatoire. Ensuite, au lieu d'utiliser l'auto-encodeur, ils prédisaient que 15 % des mots cachés au lieu de reconstruire l'intégralité de l'entrée.

Bien qu'elle permette d'envisager un contexte bidimensionnel, deux inconvénients subsistent. Premièrement, étant donné que les mots seront masqués, il y aura une déconnexion entre le pré-entraînement et le réglage fin car les jetons marqués ne sont pas affichés lors du raffinement du modèle. Pour améliorer et éviter que cela ne se produise, nous n'affichons pas toujours les mots cachés avec le jeton [MASK]. Au lieu de cela, le générateur de données d'entraînement choisit 15 % de jetons au hasard et effectue les étapes suivantes : Exemples de phrases : "Il fait beau aujourd'hui", le mot choisi pour cacher est le mot "beau". Remplacez 80 % des mots sélectionnés dans les données d'apprentissage par le jeton [MASK]. 10% des mots sélectionnés seront remplacés par un mot au hasard, la phrase peut devenir "Il fait la table aujourd'hui". Les 10% restants sont conservés "Il fait beau aujourd'hui".

Encodage de Transformers aucun mot n'a été remplacé dans l'entrée, il doit donc représenter le contexte du jeton d'entrée. De plus, puisque seulement 15 % de tous les jetons sont remplacés par un mot aléatoire, cela n'affecte pas la capacité du modèle à comprendre la langue.

Le deuxième inconvénient est que seulement 15 % des jetons sont prédits à chaque tour, ce qui signifie que nous avons besoin d'un modèle de pré-entraînement différent pour rendre l'ensemble du modèle plus efficace.

Next Sentence Prediction

De nombreuses tâches importantes dans le traitement du langage naturel telles que la réponse aux questions nécessitent une compréhension de la relation entre deux phrases de texte. Pour former un modèle qui comprend cette relation, nous construisons un modèle qui prédit la phrase suivante en fonction de la phrase actuelle. Nous choisissons la phrase A et la phrase B pour la pratique, dans laquelle 50% de la phrase B est la phrase suivante de la phrase A, et les 50% restants sont une phrase aléatoire.

Exemple :

Input : [CLS] Cet homme est [MASK] dans le magasin [SEP] il est [MASK] et sympa [SEP]

Label : Vrai

Input : [CLS] Cet homme est [MASK] dans le magasin [SEP] elle a un poisson dans le sac [SEP]

Label : Faux

Pre-training Procedure

Les scientifiques ont utilisé deux ensembles de données pour le processus de formation, BookCorpus (800 millions de mots) et Wikipedia anglais (2 500 millions de mots). Pour Wikipédia, nous ne sélectionnons que des paragraphes de texte. En fait, le but ici est de sélectionner des paragraphes de texte (mots, phrases dans un certain ordre).

Fine-tuning BERT

Pour les tâches de classification des phrases, BERT est très simple à affiner. Pour obtenir la représentation d'une séquence d'entrée avec un nombre fixe de dimensions, il suffit d'obtenir l'état caché dans la dernière couche, c'est-à-dire la sortie de la classe Transformers pour le premier jeton (jeton spécial [CLS] construit pour l'en-tête de chaîne).

A.4 Test du résultat

Une fois que nous en avons fini avec l'étiquetage automatique de tous les tweets, vérifions ensemble la précision de la méthode d'apprentissage automatique. Selon le calculateur, la précision mesurée est de 78 %. On va tester 3 fois, avec la méthode de prendre au hasard 30, 40 et 50 tweets parmi 8611 tweets, puis on va taper manuellement les labels. Enfin, comparez les deux résultats et calculez le pourcentage correct de la méthode d'étiquetage automatique.

Premier test : 80 % de réponses correctes (24 tweets correctement étiquetés sur 30). Les phrases mal étiquetées sont principalement dues à l'apprentissage automatique qui interprète mal les mots de liaison dans des phrases composées comme "même si", ou des phrases avec des emojis ajoutés lorsque l'apprentissage automatique n'a pas réussi à apprendre ces symboles.

Par exemple, la phrase : des scientifiques découvrent que le virus contient des séquences et brevetées par moderna trois ans avant le début de l'épidémie... collez ça sous le pif des lobotomisés qui usent de morales bas de rayon 🤪

Cette machine à phrases apprend l'étiquetage positif. L'ordinateur comprend qu'il s'agit d'une affirmation selon laquelle les scientifiques ont découvert l'ADN du virus il y a trois ans, mais il ne comprend pas le symbole de la colère à la fin de la phrase.

2ème test : 87% de taux de précision, assez surpris que cette fois, le taux d'étiquetage correct soit de 87%, il y a une note dans ce test quand l'ordinateur comprend que lors de l'utilisation du gros mot, il y a un avis négatif, car dans un cas dans ce test lorsqu'un compte a utilisé des mots grossiers pour décrire des sentiments positifs à propos du vaccin Covid-19, l'ordinateur l'a qualifié de négatif.

3ème test : Dans ce dernier test, avec 50 tweets, le taux de précision est de 86%. La caractéristique de ce troisième test est que l'apprentissage automatique interprète souvent à tort le sens des phrases négatives en positif. Cependant, ce n'est qu'une minorité, le taux de précision est encore très élevé.

En moyenne, le taux de précision de 3 tests est de 84%, 5% supérieur au score calculé par l'ordinateur. Cela démontre l'efficacité de la labellisation automatique de la méthode CamemBERT, et l'utilisation du corpus de 5 millions de tweets sur Twitter pour entraîner l'ordinateur.

B. Les Hypothèses

Les hypothèses de la première partie: données de Twitter

La majorité des utilisateurs ont exprimé leur opinion positive/neutre/négative

(Hypothèse de la majorité) :

Sur les plateformes de réseaux sociaux numériques comme aujourd'hui, nous communiquons les uns avec les autres via des comptes virtuels, cachant parfois nos vrais noms, cachant des informations personnelles. Cela nous permet de révéler plus facilement nos émotions, notre vrai moi que dans le monde réel. Dans la vraie vie, vous pouvez être une personne timide, ayant peur de communiquer, de discuter ou de donner pas souvent des opinions personnelles, sur les réseaux sociaux numériques cela peut être complètement le contraire, car c'est un monde virtuel, un monde plus plat où les droits de tous sont les mêmes, nous avons le même droit de parole là-haut. À cause de cela, il y a aussi beaucoup d'insatisfaction et de colère que nous publions sur les réseaux sociaux. La frustration, l'insatisfaction ou les émotions négatives apparaissent de plus en plus sur les réseaux sociaux numériques. Selon les dernières études, les émotions négatives, la colère sont les émotions les plus virales sur les réseaux sociaux en Chine. Ou encore, des recherches menées par des étudiants de l'Université de Yale montrent également que les médias sociaux nous motivent à exprimer extérieurement des émotions négatives telles que la colère ou la cruauté. En France, la violence via les plateformes numériques apparaît de plus en plus. Il y a eu entre 2017 et 2018 une

augmentation de 66,6% du nombre de commentaires violents sur ces mêmes réseaux. Au total, ils représentent 9,5% de l'ensemble des commentaires sur les réseaux sociaux. Et cette progression fulgurante s'explique en partie par ce chiffre : les auteurs d'insultes sur le web étaient 10,3% de plus en 2018 qu'en 2017.

Méthodologie : En utilisant la méthode de CamemBERT, observer la proportion des opinions, on peut montrer cet avis ?

À l'aide du modèle CamemBERT, analysez le contenu des tweets. À partir de là, CamemBERT calcule le score de ce "tweet" et l'organisera en 3 groupes : Positif, Négatif

Il y a des différences d'opinion selon les caractéristiques sociales des utilisateurs (Hypothèse de la différenciation)

Avec l'aide de CamemBERT, on peut voir si ce sont les femmes ou les hommes (jeunes ou âgés) qui sont ceux qui réagissent le plus fortement à la question des vaccins contre le Covid-19. Et avec le temps, si ce ratio changera ou non.

Méthodologie : La méthode appliquée dans cette section sera d'utiliser une tabulation croisée entre des groupes de tweets positifs ou négatifs selon le genre (femmes ou hommes). Utilisez ensuite un autre tableau croisé avec des colonnes de chronologies divisées pour observer le changement d'attitude des utilisateurs. Ceci est destiné à répondre à quelques questions telles que : Y a-t-il un changement marqué dans les attitudes sur la question des vaccins Covid-19 selon le sexe. Ou : Face aux décisions de la partie principale, ou aux articles d'intérêt, qui sont les plus sensibles.

Influence entre individus dans un réseau social (Hypothèse de l'influence dans le réseau social). (en temps de réflexion)

Degenne et Forsé (1994 : 216) parlent de cercle social lorsque les individus «se reconnaissent comme faisant partie d'un ensemble et que cette appartenance influence certaines de leurs conduites ». Si l'on accepte cette définition et que l'on envisage le cercle comme le lieu d'un code, se pose assez naturellement la question des mécanismes ou des moyens de promotion de ce code. C'est dans cette optique que nous poserons les questions du possible influence de la perception subjective du contrôle social du réseau d'une part (deuxième question), de la structure du réseau social (taille et nombre de liens, densité des relations, composition sociale, inscription des partenaires dans le réseau) d'autre part (troisième question), sur la correspondance entre les normes idéales des acteurs et les normes pratiques du réseau. Ces questions représentent la face inverse de celle relative à la marge de manœuvre de l'acteur social : « Dans quelle mesure la norme du réseau s'impose-t-elle à lui ? ».

Méthodologie : Notre hypothèse est qu'il existe une influence comportementale chez les utilisateurs d'un même cercle social.

Politique dans les tweets de l'utilisateur.

Les tweets des utilisateurs expriment-ils simplement leurs opinions ou y a-t-il d'autres objectifs politiques ? Utilisent-ils Twitter pour lutter dans un but politique, pour manipuler, influencer d'autres groupes sociaux ?

Méthodologie: Observez les 20 utilisateurs de Twitter les plus actifs sur la question du vaccin Covid-19, puis recherchez les opinions politiques, le lien entre leurs tweets et les opinions politiques.

Les habitudes de comportement et l'utilisation de Twitter sont présentes.

L'analyse des habitudes de comportement ou d'utilisation d'un site de réseautage social nous aidera à prédire les caractéristiques de ce groupe social. De l'heure de publication, du nombre de publications par jour, etc.

Méthodologie : En utilisant la méthode Time Series en Python, analysez-la et affichez-la avec un graphique, à partir duquel il est possible d'évaluer en fonction des critères ci-dessus par stationnarité, saisonnalité

Les hypothèses de la deuxième partie: anti-vaccination

Définition : Le phénomène de résistance à un, plusieurs ou tous les vaccins, et un manque de confiance dans l'efficacité et la sécurité de la vaccination.

Pourquoi le refus de la vaccination ?

Beaucoup de gens ne réalisent pas l'efficacité des vaccins. Une étude britannique portant sur 14578 enfants a révélé qu'environ 75 % des refus de vaccination des parents n'étaient pas intentionnels. Par conséquent, le manque d'information n'est pas le principal facteur conduisant au refus de vacciner les enfants.

La principale raison de la vaccination est l'opinion d'un professionnel de la santé, le motif étant l'autoprotection. La principale cause d'hésitation est l'inquiétude quant à l'innocuité du vaccin.

L'hypothèse est que l'anti-vaccination en France existe depuis longtemps, et pendant la pandémie de Covid-19, ce phénomène est devenu de plus en plus grave.

Méthodologie : Analyse du Baromètre de Santé.

On a de la différenciation entre les caractéristiques sociales dans l'anti vaccination.

On utilise les chiffres de la baromètre de santé en 2019, on fait le tableau de tris croisés entre les deux variables :

L'âge -Q1/Q2/Q3 : nous montre le pourcentage de personnes qui sont d'accord, pas d'accord avec la question de la vaccination. A partir de là, on peut voir à quel âge le pourcentage de personnes qui ne veulent absolument pas se faire vacciner est le plus élevé.

Si les moins de 40 ans, la tranche d'âge la plus exposée et influencée par les médias sociaux, ont les taux d'anti-vaccin les plus élevés.

L'utilisation des chiffres de 2010, nous aide à voir s'il y a un changement dans la perception des gens sur cette question. Si la proportion de personnes opposées à la vaccination a augmenté ou diminué au fil du temps. La raison de ce changement.

Diplôme-Q1/Q2/Q3: nous montre s'il existe une différence de perception de cette question parmi les personnes instruites. Existe-t-il un lien entre le niveau d'éducation et l'anti vaccination ?

Et si ce ratio change avec le temps, quelle est la raison pour laquelle cela se produit.
Je n'ai pas encore trouvé d'article qui parle de ce problème.
A partir de ces données, nous observons le processus de changement dans la société concernant la vaccination, s'il y a un changement. Ceci combiné avec les données ci-dessus, pouvons-nous dessiner quoi que ce soit.

V. Résultat et interprétation.

A. Analyse des sentiments sur Twitter.

A.1 Inégalité entre les sexes dans la publication de statuts sur Twitter.

Les données analysées comprennent 8611 tweets issus de 3135 comptes observés, il y a une inégalité entre les sexes dans l'expression du statut et des émotions sur Twitter pour les questions liées aux vaccins contre le Covid.-19. En conséquence, les hommes représentaient une proportion plus élevée d'opinions personnelles ou de commentaires sur les problèmes de vaccin Covid, avec 1960 (62,5%) personnes au lieu de 1089 femmes (34,7%). Le reste compte 86 individus de sexe indéterminé, soit environ 2 %. La raison pour laquelle il y a des individus non identifiables est des comptes d'organisations, d'associations ou de comptes qui ne peuvent pas être sexués du propriétaire. Alors, avec ces chiffres, peut-on conclure que les hommes français sont plus intéressés par les vaccins Covid-19 que les femmes ?

Tout d'abord, examinons les données démographiques des utilisateurs de Twitter. En France, Twitter compte 16,8 millions d'utilisateurs mensuels (4,5 millions d'utilisateurs quotidiens), dont 60% d'hommes et 40% de femmes, ce chiffre ne comprend pas les comptes qui ne permettent pas d'identifier le sexe. D'après ces chiffres, on constate que les hommes français ont tendance à utiliser Twitter plus que les femmes. D'autre part, Twitter est aussi un réseau social capable de transmettre rapidement des informations officielles, quand en France de nombreux canaux de communication du gouvernement, ou de grandes agences de presse, choisissent Twitter comme canal pour mettre à jour les premières informations. Par conséquent, les gens ont tendance à choisir Twitter pour mettre à jour les actualités plutôt que pour les loisirs. Alors, avec ces chiffres, pouvons-nous confirmer que les hommes se soucient des problèmes sociaux ou lisent plus d'actualités que les femmes ? Selon [journaldunet](#), les femmes utilisent davantage Internet que les hommes, représentant 53 % des internautes, mais elles sont plus intéressées par la connexion avec des amis, la cuisine, les achats et les questions liées à la santé. Vaccin contre le Covid-19 ? Je comprends ce que vous pensez, bien sûr, c'est un problème lié à la santé. Alors, les femmes participent-elles plus fréquemment aux débats que les hommes ?

	name	username
135	Amor LOUHICHI	338
784	Ep	153
2665	jean pierre grosse	140
2636	heautontimoroumenos	90
3072	Colonel Kurtz	68
...
1258	Jul Lazzarini	1
1259	Jules Koch ID 3/3	1
1260	Jules d'Oc FR	1
1261	Julia Dumont	1
3111	💧💙#Jessy🌟🌟FR	1

Amor LOUHICHI est le participant le plus actif dans les commentaires, les posts exprimant son attitude vis-à-vis des Vaccins Anti-Covid, et c'est un homme. La particularité, c'est qu'il n'y a pas de femmes dans le top 5 des personnes les plus actives. Les hommes submergent complètement les femmes à cet égard. L'aspect de la santé, ou un peu lié à la santé, mais les hommes dominent complètement en termes d'expression d'opinions. C'est donc que le vaccin contre le Covid-19 n'est pas simplement une question de santé, mais peut être plus large (éventuellement politique), nous en reparlerons plus tard.

Mais que nous soyons des hommes ou des femmes, il nous est plus facile de partager des informations ou nos opinions sur les réseaux sociaux. Avec la propagation rapide des réseaux sociaux, chaque déclaration, chaque information, chaque opinion peut devenir un sujet brûlant, qui peut affecter d'autres utilisateurs. Comment s'exprime le regard des utilisateurs de Twitter sur le vaccin anti-covid, ou quelles en sont les caractéristiques ?

A.2 Opinion des utilisateurs avec les vaccins contre Covid-19.

L'un des deux principaux objectifs de la mémoire est de classer les tweets en fonction des attitudes des utilisateurs. Ces avis sont divisés en deux catégories : positives ou négatives. Alors pourquoi diviser en deux, pas trois (commentaires neutres). Comme je l'explique plus haut, l'introduction par le gouvernement français de la réglementation sur la vaccination contre le Covid-19 peut également s'expliquer pour cette division d'opinion. Et comme tous les vaccins, vous avez deux choix : vous faire vacciner ou non. Si vous êtes vacciné contre le Covid, vous aurez un pass sanitaire et bénéficierez de quelques avantages pendant la période post-confinement comme manger au restaurant, pouvoir aller au cinéma, participer à certaines activités publiques. Mais la politique de la vaccination, nous fait, que cela nous plaise ou non, nous n'aurons que deux extrêmes dans nos opinions et attitudes envers les vaccins contre le Covid-19 dans un moment.

Negative	0.538962
Positive	0.461038

Avec 8611 tweets collectés avec les mots-clés vaccins Covid-19, nous avons un résultat que 46% des tweets ont une attitude positive envers les vaccins Covid-19, et 54% ont une attitude négative. Cela ne veut pas dire que 54% des utilisateurs ont une attitude négative (ou 46% des utilisateurs ont une attitude positive) car un utilisateur peut avoir des opinions à la fois positives et négatives sur les vaccins Covid-19. Alors est-ce illogique avec les arguments ci-dessus, quand on suppose que chaque personne n'a qu'une catégorie d'opinion positive ou négative envers le vaccin Covid-19 ? N'oubliez pas le mot un instant, car c'est le deuxième objectif de la mémoire, qui est d'observer l'évolution de l'opinion des utilisateurs. Selon Marine Cardot (écrit dans LesEschos), 16% des Français ne souhaitent pas se faire vacciner contre le Covid-19. Si le nombre de Français qui ne veulent pas se faire vacciner contre le Covid-19 apparaissent sur Twitter est aussi de 16%, le fait que jusqu'à 53% aient des Tweets exprimant une attitude négative envers les vaccins contre le Covid-19 est vraiment remarquable. Ceci dit, les militants du mouvement contre le vaccin Covid-19 sont actifs sur Twitter. Ils sont actifs et ont une réelle influence sur ce réseau de médias sociaux. Il est possible que sur ces 53% de tweets négatifs, tous ne proviennent pas de personnes qui sont réellement contre le vaccin Covid-19, il peut y avoir quelques personnes qui ont été influencées par les arguments anti-vaccination. Covid-19, ou ils ont également perdu confiance en ces vaccins après avoir été témoins de quelque chose.

Les attitudes et les sentiments envers les vaccins peuvent être observés chez les personnes les plus actives, en divisant le score total du label par le nombre total de tweets de cette personne. Plus la moyenne est basse, plus la personne est négative.

	name	label
135	Amor LOUHICHI	0.375740
247	Azeda	0.088235
281	BayesReality	0.305556
433	Carulu-andria PAZZONI euua	0.375000
500	Christophe R	0.500000
716	Dr Jean-Jacques Fraslin	0.466667
784	Ep	0.326797
885	Florquin Sylvie	0.181818
972	GeB. 🍌🍌🍌🍌🍌🍌	0.307692
1295	KINGDHARMA	0.675676
1379	Laurent BIBAS	0.558140
2211	Solange	0.741935
2519	backtobasics	0.208333
2636	heautontimoroumenos	0.344444
2665	jean pierre grosse	0.421429
2681	josiane MICHEL	0.696970
2746	mariesaw	0.257143
2841	sabine dumas	0.500000
3072	™ Colonel Kurtz	0.500000
3079	🍌 Les Mots de Cath 🍌	0.423077

13 personnes sur 20 dans le top 20 des personnes les plus actives ont plus d'opinions négatives que positives. Dans lequel, Azeda a le score le plus bas (0,088), ce qui signifie que

A.3 Le contenu des tweets.



45

Il y a un nom qui nous intéresse aussi beaucoup, quand elle se situe dans le coin supérieur gauche du nuage de mots : Emmanuel Macron, ou plus brièvement : Macron. Le nom du président français a été cité dans 650 tweets liés au vaccin Covid-19, si le champ de la recherche est élargi, ce nombre est probablement bien plus. Il est à noter que jusqu'à 60% des tweets liés au président français sont négatifs. Plus que la moyenne que nous avons calculée ci-dessus (53 % de négatif). Cela soulève pour nous une question : les utilisateurs de Twitter en France ne sont-ils pas d'accord avec les mesures prises par le gouvernement Macron simplement parce qu'ils sont préoccupés par leur propre santé, ou y a-t-il un autre facteur ? Un facteur politique les affecte.

A.4 Le point vue de politique des individus.

60% des tweets mentionnant Macron en lien avec le vaccin Covid-19 sont négatifs. Cela signifie-t-il que l'opposition aux mesures de vaccination du gouvernement français est là, et qu'il y a aussi une opposition à Macron. Mais que cette opposition soit due à des problèmes de santé, à des sentiments de méfiance, à la désapprobation des mesures du gouvernement simplement ou il y a des facteurs politiques aussi.

Parce que Twitter est un réseau social sécurisé sur les informations des utilisateurs et n'est pas capable de collecte automatique. Combiné avec le fait qu'il existe de nombreux comptes qui ne peuvent pas déterminer le sexe ou le nom de l'utilisateur. Dès lors, considérer un point de vue politique basé sur les 20 personnes les plus actives sur la question des vaccins contre le Covid-19 sera faisable et efficace.

La première personne à qui nous tenons est la personne la plus active, Monsieur Amor LOUHICHI. Il est l'ancien secrétaire national de l'Union démocratique et indépendante, un parti français. Sur Twitter, il a publiquement désapprouvé l'actuel président français, Emmanuel Macron. Amor LOUHICHI appelle Macron : La Honte Macron. Il n'est pas possible d'affirmer à 100% que son refus de soutenir Macron affecte son opinion sur le vaccin anti-covid. Mais c'est tout à fait possible. car comme nous l'avons identifié ci-dessus, chaque facteur social peut affecter nos émotions et nos actions. Amor LOUHICHI a déclaré qu'il ne soutenait pas Macron, qu'il ne soutenait pas les mesures proposées par le gouvernement français pour limiter la propagation du Covid-19 comme PassSanitaire, et qu'il doutait également de l'efficacité du vaccin pfizers. Dans l'un de ses tweets, Amor a soulevé des soupçons sur le vaccin pfizer et a qualifié Macron de menteur : *"le produit de pfizer qui demande de en d'injections chaque jr rapprochées n'a le droit de cité en tant que vaccin. c'est tout au un sparadrap qui évite certains effets du mais très temporairement macron a menti et continue à mentir ruiner a raison."*

La deuxième personne dont nous nous soucions est Azeda. Elle a le score le plus bas dans la moyenne des opinions (0,088), ce qui signifie que presque tous ses tweets soulèvent des objections, des doutes et une résistance aux vaccinations contre Covid-19. En faisant des recherches sur sa page Twitter, j'ai appris qu'elle est aussi une anti-Emmanuel Macron, alors qu'il y a des tweets négatifs sur l'actuel président français : désaccord sur la question du Pass Sanitaire, ou d'autres politiques. Il y a donc quelque chose entre M. LOUHICHI et Azeda, non seulement ils sont négatifs sur cette question de la vaccination, mais ils sont tous des

partisans non Macron. Alors est-il possible qu'on ait un lien entre l'anti-vaccination contre le Covid-19 et le non-soutien à Macron ? Ou le point de vue politique commun de ceux qui sont contre le vaccin Covid-19 est-il contre la politique du gouvernement Macron ?



Il y a une chose que j'ai trouvée sur la page personnelle d'Azed, qu'elle a soutenue et votée pour Mélenchon lors de l'élection présidentielle française de 2022. Mélenchon est un homme politique français, avec des enjeux liés à la politique de prévention de l'épidémie de Covid-19, il s'oppose à la politique du Pass Sanitaire et du Pass vaccinale, il pense que ce n'est pas vrai sur les questions de sécurité sanitaire. Non seulement LOUHICHI et Azeda, les militants qui ont des commentaires plus négatifs ont tous la caractéristique commune de ne pas soutenir Macron, s'opposant à des politiques comme le Pass Sanitaire. Lors de l'élection présidentielle française de 2022, ils choisissent souvent d'autres candidats comme Mélenchon et Zemmour, ou choisissent Macron en dernier recours.

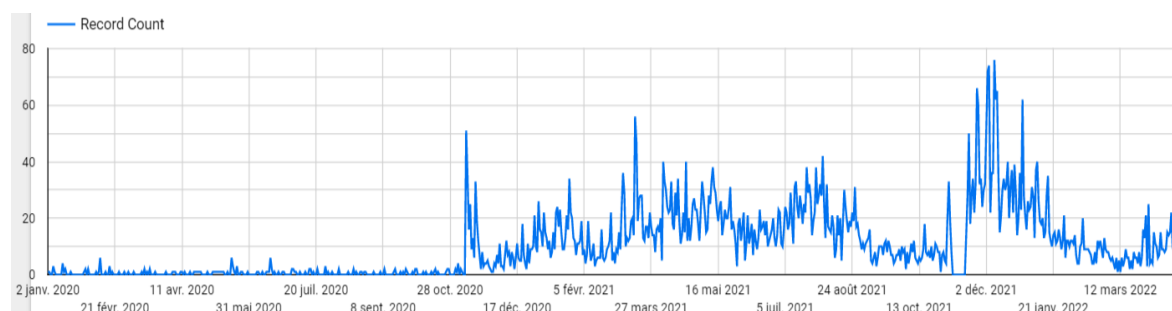
Alors qu'en est-il des personnes avec des commentaires plus positifs ? La personne avec le surnom de Solange a le pourcentage le plus élevé de tweets positifs. Et dans son profil sur Twitter, elle soutient et vote ouvertement Emmanuel Macron à l'élection présidentielle. Et aussi sur sa page personnelle, elle poste des tweets pleins de sens positifs envers les vaccins contre le Covid-19, annonçant et soutenant les mesures gouvernementales de lutte contre la maladie, de plus elle montre aussi un désaccord avec les anti-passants : « *les anti-pass sont une minorité, mais ils ont pas des marginales* ».

Suite à ces exemples des personnes les plus actives sur Twitter, on constate qu'il existe une interaction entre les opinions politiques et les opinions sur la question des vaccinations contre le Covid-19. Cela signifie que ce n'est pas seulement une question de santé, dans les messages des utilisateurs, on voit aussi leurs opinions politiques, par exemple : dans le même tweet sur le vaccin pfizers, LOUHICHI a également mentionné le total M. Macron est un

menteur. Ou plus profondément, c'est l'élection présidentielle (votez pour des candidats qui ont des vues similaires avec les politiques de prévention du Covid-19 du gouvernement). Nous avons considéré la politique de chaque individu, donc quels autres facteurs les affectent ou non. Venons-en maintenant aux analyses au fil du temps pour voir l'évolution des publications.

A.5 Evaluation des nombres de tweets.

L'analyse de la chronologie nous aide à observer l'évolution du nombre de tweets ou l'évolution émotionnelle de la communauté en ligne. Examinons d'abord l'évolution des tweets du 1er janvier 2020 à mars 2022.



On voit que de janvier 2020 à fin octobre 2020, le nombre de tweets est extrêmement faible, presque tous les jours il y a des tweets liés au vaccin Covid-19. C'est le moment où l'épidémie de Covid-19 fait rage avec la première vague de la maladie. Le 13 janvier 2020, le CDC a annoncé que le génome du virus SARS-CoV-2 avait été publié dans la base de données de séquences génétiques GenBank du NIH. C'est un pas en avant quand, grâce au décodage du génome du virus SARS-CoV-2, on pourra fabriquer un vaccin. Et dès la fin janvier 2020, le programme de recherche de vaccins Pfizer-Biotech a été lancé, suivi peu de temps après par le vaccin Moderna étudié par Le National Institute of Allergy and Infectious Diseases et Moderna Corporation des États-Unis. En février 2020, un groupe de chercheurs de l'Imperial College de Londres a publié sa volonté de réduire le temps de production et de développement "de trois à quatre ans à 14 jours". C'est pourquoi, même s'il n'y a pas d'informations officielles sur un vaccin contre le Covid, sur les sites de réseaux sociaux durant la période de début février 2020 à fin septembre 2020, des personnes affichent encore des lignes d'état liées au développement. Leurs lignes de statut sont majoritairement des pensées positives, souhaitant que la préparation et les tests du vaccin se fassent rapidement. Après ces jours, lorsque nous regardons le graphique ci-dessus, il y a trois ou quatre pics du graphique, représentant un pic de tweets sur le problème du vaccin Covid-19.

Le premier a eu lieu le 9 novembre 2020, lorsque le nombre de tweets est passé d'un ou deux tweets à 52 tweets. Ce nombre n'est pas le plus grand nombre de tweets en une journée sur les vaccins Covid-19, c'est la première fois qu'une nouvelle officielle sur l'essai du vaccin Covid-19 se concrétise. Selon la BBC, ou de nombreux autres sites, les vaccins Pfizer et BioNTech préviennent la maladie chez plus de 90 % des personnes vaccinées. Cet essai a été mené sur 43 500 personnes dans six pays et n'a soulevé aucun problème de

sécurité. Cette bonne nouvelle est annoncée dans une situation où la pandémie de Covid-19 montre à nouveau des signes de chaleur alors que précédemment le 7 novembre, le Royaume-Uni a interdit tous les visiteurs du Danemark, et les États-Unis ont enregistré plus de 132 700 cas d'infection à Covid-19. En France, début novembre 2020, le port du masque est également redevenu obligatoire dans tous les lieux publics. L'annonce que le processus de développement du vaccin a eu un signal positif lors des tests chez l'homme a provoqué des réactions très positives de la communauté Twitter en France.

Julien Stephan: *“notre maître à tous, FN Lutte, yosoleno media venir même efficace à c'est bon. les français ont enfin moins peur avec un vaccin même partiellement efficace”*

Riad Lemhachheche: *“Très bonne analyse de l'annonce pfizer france sur le potentiel vaccin covid.”*

Au milieu de ces nouvelles positives, on voit encore des tweets moquant le gouvernement français. Cependant, ce n'est qu'une petite partie, rien qu'en ce 9 novembre 2020, 67% des tweets sont positifs.

Mais cette attitude positive a été modifiée par le deuxième pic du graphique, le 15 mars 2021. Le président français a annoncé la décision d'arrêter la vaccination par AstraZeneca en France ainsi que dans 9 autres pays européens, en raison d'effets secondaires chez certains patients, dont la thrombose. La suspension de la vaccination contre AstraZeneca a déjà été effectuée par d'autres pays. L'Autriche a cessé d'utiliser cette série de vaccins le 8 mars suite au décès d'une infirmière. Le Danemark, la Norvège, l'Islande et les Pays-Bas ont interrompu indéfiniment tous les lots du vaccin AstraZeneca, suivis de l'Allemagne, de l'Italie et de l'Espagne. La communauté Twitter a été choquée par cette nouvelle alors qu'un jour plus tôt, le Premier ministre Jean Castex avait annoncé que le gouvernement n'arrêtera pas la vaccination contre AstraZeneca. Cela a conduit beaucoup à se moquer du gouvernement et de son incohérence dans ses actions et ses décisions. De plus, presque tout le monde a annoncé que le président Macron arrêterait la vaccination de ce vaccin, et a également fait quelques commentaires selon lesquels il se fait vacciner par Pfizer ou un autre vaccin. Le Dr Jean-Jaques Fraslín conseille à tous de se faire injecter Janssen car il n'a pas la même catégorie que Pfizer et AstraZeneca. Il est à noter qu'une source du surnom Solange a déclaré qu'en fait, les médecins au Royaume-Uni ont reçu de nombreux rapports de caillots sanguins après l'injection de Pfizer et d'Astrazeneca depuis février 2021, mais aucun d'eux n'a dépassé les attentes des fourmis. Mais il semble que la communauté en ligne ne soit pas très intéressée par les informations ci-dessus, beaucoup de gens pensent que seul le vaccin AstraZeneca provoque la coagulation du sang, pas Pfizer. Mais est-ce que ce sont les actions et les décisions du gouvernement français qui intéressent la communauté en ligne ?

“après avoir entendu parlé castex hier sur twitch de son comité de vigilance pour évaluer astrazeneca puis macron aujourd' hui de la science et de la haute autorité européenne de la santé sur le même sujet je me reprends à penser que je suis dans brazil”

“et à part ça hier on nous disait qu' il fallait avoir confiance. et après ils ne comprennent pas pourquoi on ne les croient plus emmanuel macron annonce la suspension de la vaccination avec astrazeneca en france bfmtv.”

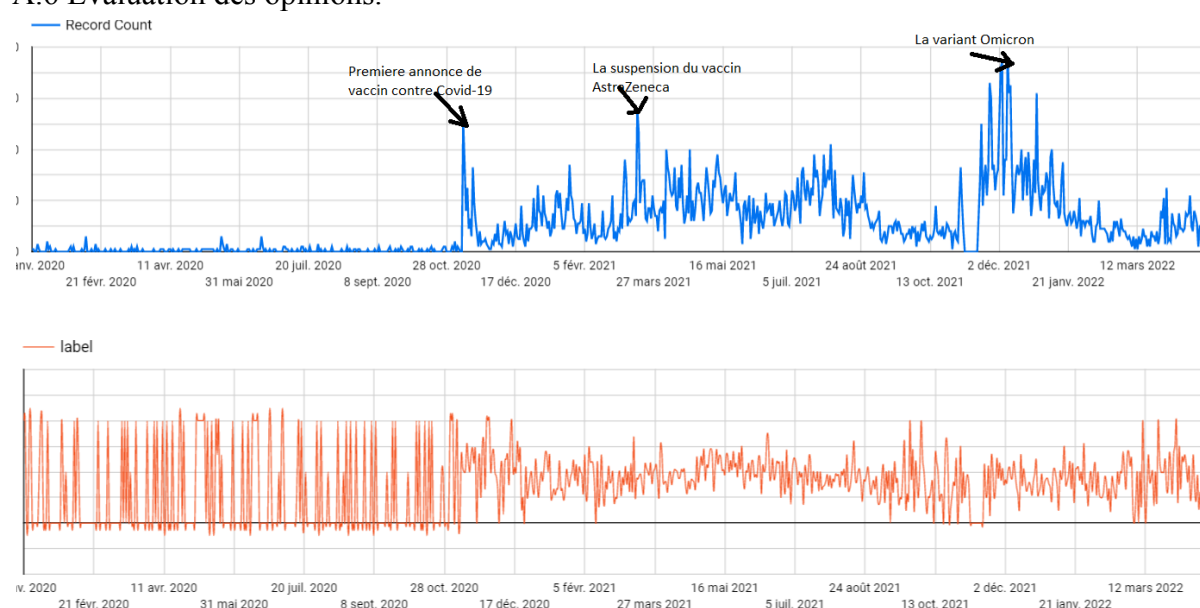
Pas tous, mais certains ont également remis en question la manière dont le gouvernement gère les problèmes liés aux vaccinations contre le covid-19.

Les 18 et 19 mars 2021, le Premier ministre Jean Castex a annoncé la mise en place d'un confinement dans les seize régions les plus touchées par la pandémie. Dans le même temps, il a annoncé le redémarrage du programme de vaccination contre le Covid-19 avec le vaccin AstraZeneca. Dans le même temps, a annoncé le redémarrage du programme de vaccination contre le Covid-19 avec le vaccin AstraZeneca. Durant ces deux jours, les gens ont évoqué la question de la ré-autorisation du vaccin AstraZeneca, certains n'ont pas compris pourquoi il était interdit puis ré-approuvé après seulement 3 jours, et d'autres aussi envisagent cette ré-autorisation pour que tous les Français puissent être vacciné.

Et le troisième temps auquel il faut faire attention, c'est aussi le moment où le plus de tweets sont postés liés à la question du vaccin Covid-19, soit de fin novembre 2021 à début janvier 2022. En fait, depuis août, la communauté en ligne a été agitée lorsque Pass sanitaire a été officiellement approuvé et exploité. Dans la rue, de nombreuses manifestations ont eu lieu avec des dizaines de milliers de personnes présentes dans toute la France (selon les calculs du ministère de l'intérieur). Le 25 novembre 2021, le ministre de la Santé a annoncé de nouvelles mesures pour faire face à la recrudescence des cas de Covid-19. Début décembre 2021, une nouvelle souche de Covid-19, Omicron fait son apparition en France. Et selon les rapports de l'époque, avec cette nouvelle souche, l'efficacité du vaccin Pfizer pourrait être réduite. Pour ceux qui reçoivent les trois doses du vaccin Pfizer, l'efficacité est de 70 %, mais à partir de la 10e semaine après la dose de rappel, l'efficacité n'est plus que de 45 %. Ce sont probablement les raisons pour lesquelles le nombre de tweets a augmenté au cours de cette période. Quand les mesures sociales se durcissent, il est temps pour ceux qui ne soutiennent pas de prendre la parole. De nombreuses personnes ont exprimé l'efficacité du Pass sanitaire, ou alors les doutes entourant l'efficacité des vaccins. De nombreux politiciens se sont également prononcés contre le pass sanitaire, ou contre la politique du gouvernement français de l'époque, comme Amor LOUHICHI, et Mélenchon. Début janvier 2022, Mélenchon affirmait que le laissez-passer vaccinal était 100 % inefficace. Peu de temps après, le ministre de la Santé a répondu en qualifiant cela de "populisme" de l'opposition et en mettant en garde contre la variante Omicron. Le 4 janvier 2021, le président Macron a annoncé que : *“ Moi, je ne suis pas pour emmerder les Français. Je peste toute la journée contre l'administration quand elle les bloque. Eh bien là, les non-vaccinés, j'ai très envie de les emmerder. Et donc on va continuer de le faire, jusqu'au bout ”*. D'autres candidats à la présidentielle ont critiqué sa déclaration.

Et peut-être pas seulement les candidats à la présidentielle mais aussi la communauté en ligne. La partie suivante permet de mieux observer l'évolution des émotions et des opinions des utilisateurs de Twitter en France sur la question des vaccins et vaccinations contre le Covid-19.

A.6 Evaluation des opinions.



Selon le graphique ci-dessus, au 09 novembre 2020, la moyenne d'opinion est de 0,7. Fidèle à notre jugement, la nouvelle que le vaccin Pfizer a été testé sur des humains avec un niveau d'efficacité allant jusqu'à 90% a attiré l'attention de tous et a réagi avec une attitude positive. Les publications négatives sont principalement dues à des préoccupations concernant la sécurité alors qu'il a fallu moins de 10 mois aux chercheurs pour amener le vaccin Pfizer à des essais sur l'homme. Il s'agit d'un doute purement logique car le temps moyen qu'il faut pour qu'un vaccin soit disponible pour utilisation se situe entre 10 et 15 ans. Mais quelle est la raison pour laquelle tout le monde est si excité par cette nouvelle ? Considérons la situation à cette époque en France. Après la fête de Toussaint, les signes de la deuxième vague de Covid-19 apparaissent peu à peu, et tout le monde vient de traverser près de cinq mois de blocage social. L'ampleur de l'impact de Covid-19 sur la vie est incontestée. Et tout le monde semble souhaiter que la vie revienne rapidement à la normale. Et puis, la nouvelle du vaccin Pfizer a été comme une lueur d'espoir de retrouver une vie normale.

Les attitudes et les opinions de la communauté Twitter sont maintenues à 0,55-0,7 jusqu'à fin janvier 2021, lorsque le changement se produit. Le score d'opinion moyen diminue

progressivement, atteignant parfois seulement 0,2, ce qui signifie que sur 10 tweets, seuls 2 tweets sont positifs. A cette heure, la négativité s'explique par l'impatience de la communauté twitter face aux progrès de la vaccination contre le Covid-19. Lorsque les nouveaux vaccins n'étaient utilisés que pour les personnes de 50 ans et plus. Notamment, le 12 janvier 2021, une étudiante à Lyon a tenté de sortir par la fenêtre, un acte qui démontre la gravité de l'instabilité mentale des jeunes dans le contexte des restrictions sociétales liées au Covid-19. Cela semble n'avoir rien à voir avec les vaccins ou les vaccinations, mais cela affecte les émotions des gens, en particulier les jeunes. Ils seront plus impatients et impatients avec la question de la vaccination contre le Covid-19.

Nous voyons que dans les périodes ultérieures, jusqu'à fin novembre 2021, les tweets quotidiens sont de l'ordre de 20 à 40 tweets par jour, le score d'opinion moyen varie de 0.25 à 0.6, ce qui prouve que le niveau négatif de la communauté en ligne est apparue, et n'est pas un phénomène temporaire, ce qui montre que l'opposition et le désaccord avec la question de la vaccination anti-Covid relève d'un processus, ou est causé par une décision ou une mesure de longue haleine. Ceci, en plus des problèmes liés au vaccin, a conduit à un effet négatif à un moment donné au cours de cette période.

Pendant la période du 19 novembre 2021 à début 2022, où les conflits liés au pass sanitaire, et la mutation Omicron réduisaient l'efficacité du vaccin Covid, le score d'opinion moyen n'était atteint qu'au seuil de 0,4, il y a des jours où il n'atteint que 0,25 (24 novembre) ou 0,13 (12 décembre). Ces indicateurs sont tout à fait cohérents avec nos déclarations ci-dessus.

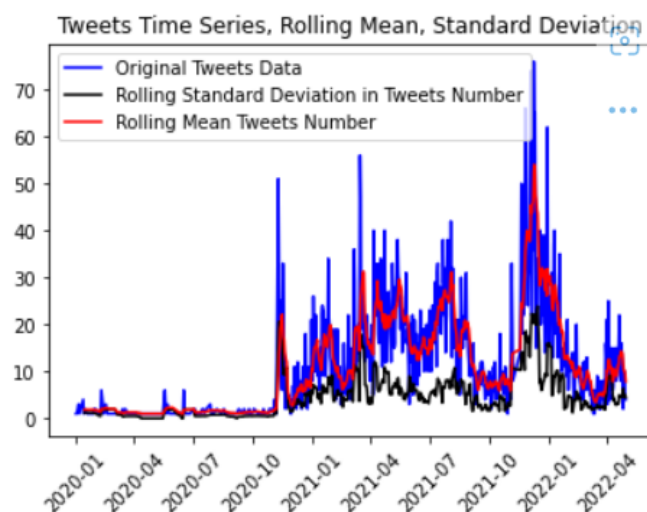
Il y a encore une chose quand on regarde ces deux graphiques côte à côte, on voit qu'à certaines périodes, avec moins de tweets, la communauté en ligne semble être plus active, alors qu'à certains moments, avec beaucoup de tweets, elle semble être plus négative. Cela reflète-t-il le comportement et les habitudes d'une partie des utilisateurs de Twitter ? Considèrent-ils Twitter comme un simple lieu d'expression de leurs frustrations, de leurs ennuis et de leur mécontentement, ou est-ce également une arme de communication efficace à des fins politiques ? Ils utilisent Twitter comme arène, débat entre partis, opinions.

Pour passer aux étapes de vérification du comportement et des habitudes des utilisateurs de Twitter, vérifions d'abord la stabilité des données.

A.7 Stationnarité

La stationnarité est un élément clé de l'analyse des séries chronologiques. En termes simples, la stationnarité signifie que la manière dont les données des séries chronologiques changent est constante. Une série chronologique stationnaire n'aura pas de tendances ou de modèles saisonniers. Vous devez vérifier la stationnarité, car non seulement cela facilite la modélisation des séries chronologiques, mais il s'agit d'une hypothèse sous-jacente dans de nombreuses méthodes de séries chronologiques. Plus précisément, la stationnarité est supposée pour une grande variété de méthodes de prévision de séries chronologiques, y compris la moyenne mobile autorégressive (ARMA), ARIMA et ARIMA saisonnière (SARIMA).

Nous utiliserons le test de Dickey Fuller pour vérifier la stationnarité de nos données. Ce test générera des valeurs critiques et une p-value, ce qui nous permettra d'accepter ou de rejeter l'hypothèse nulle qu'il n'y a pas de stationnarité. Si nous rejetons l'hypothèse nulle, cela signifie que nous acceptons l'alternative, qui stipule qu'il y a stationnarité.



Nous voyons que nos données ne sont pas stationnaires, car la valeur de p est supérieure à 5 % (8 %) et la statistique expérimentale est supérieure à la valeur critique. Le fait que les données des Tweets ne soient pas stables est quelque chose que nous pouvons complètement voir à l'œil nu, car cela change, parfois un peu, du nombre de tweets. Cela prouve que le fait qu'il y ait plus ou moins de tweets dépend d'un certain changement social, affectant les utilisateurs de la plateforme Twitter.

A.8 Autocorrelation.

Le test du degré d'autocorrélation est également une partie importante de l'analyse des données de séries chronologiques. Il s'agit d'une mesure de la façon dont les données de séries chronologiques sont corrélées à un moment donné avec des valeurs passées. Dans notre cas, nous verrons si le comportement de l'utilisateur de Twitter durera longtemps ou s'il sera simplement diffusé pendant une courte période.

One Month Lag: 0.7379733957704617

Avec un décalage d'un mois, nous constatons une corrélation de 0,74, ce qui signifie que si le nombre de tweets est important en ce moment, il y a 73 % de chances qu'il en soit de même après un mois.

Three Month Lag: 0.5966259755431302

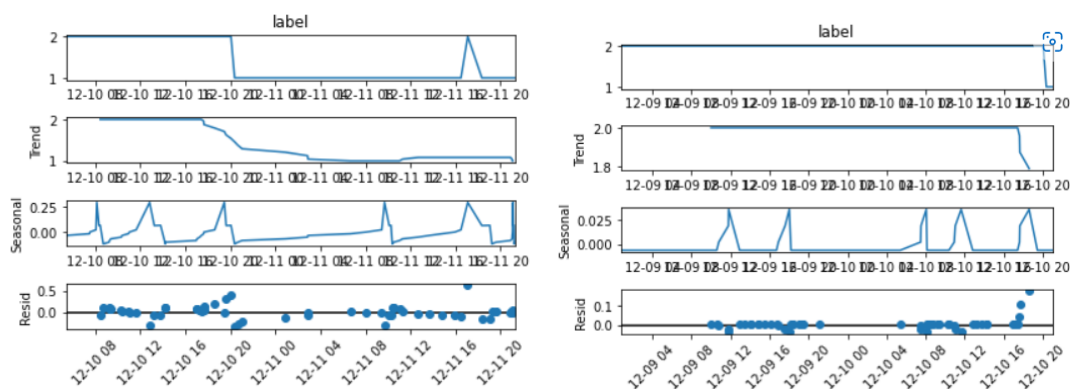
Six Month Lag: 0.6382702125947161

Nine Month Lag: 0.558386798007838

Et après 3 mois, 6 mois, 9 mois, le niveau de corrélation n'est pas aussi fort qu'après 1 mois, mais toujours significatif. De plus, nous constatons une reprise de la corrélation au bout de 6

mois. Cela montre en fait que tous les 6 mois, le nombre de personnes qui tweetent montre des signes d'augmentation. Cela est évident en novembre 2020 et mars 2021 lorsqu'il y a une reprise des Tweets, ce qui signifie que quelque chose lié au vaccin Covid-19 est à nouveau sur la canopée.

A.9 Saisonnalité.



Tenez compte de la saisonnalité dans la publication de tweets liés au vaccin Covid-19. On constate que le nombre de tweets est majoritairement posté le jour, de 8h à 16h et diminue la nuit. A partir de l'habitude de poster ces lignes de statut, peut-on déceler des caractéristiques des utilisateurs de Twitter ? Bien sûr, nous ne pouvons pas identifier exactement tous les utilisateurs, mais nous pouvons identifier les caractéristiques de base de la majorité des utilisateurs de Twitter. Pour ceux qui exercent des métiers qui ne sont pas liés aux réseaux sociaux, ou qui n'utilisent pas forcément les réseaux sociaux, bien sûr, il n'est pas possible d'utiliser twitter pour poster des émotions entre 8h et 16h. Et à partir des données des 9 et 10 2021, qui sont les jeudis et vendredis, on peut aussi quelque peu deviner les groupes de personnes les plus susceptibles de poster ces tweets.

L'un est un groupe de personnes qui utilisent les réseaux sociaux, Twitter comme outil pour leur travail : journalistes, dirigeants politiques, militants sociaux. Tous, utilisant Twitter comme canal de communication, expriment leurs points de vue pour influencer les autres utilisateurs de Twitter.

Le deuxième groupe : peut-être des personnes qui travaillent dans des domaines spécialisés, mais qui sont maintenant à la retraite, elles vont en ligne pour partager leurs points de vue, leurs opinions, ainsi que pour partager des informations utiles.

Le troisième groupe : Ceux qui ont beaucoup de temps, ou passent le plus clair de leur temps sur les réseaux sociaux, surfant sur le web pour trouver des informations, exprimer des opinions, ou soulager une frustration.

Comme pour les autres groupes sociaux comme les étudiants, les personnes qui vont travailler, bien sûr, peuvent aussi afficher des statuts à ces moments-là, mais seulement une minorité, insignifiante.

B : Analyse du Baromètre de la Santé.

Dans ce chapitre, nous examinerons les problèmes liés aux vaccins à dose unique ou à dose unique favorables/défavorables. Cela ne signifie pas que les personnes qui sont défavorables à tous les vaccins sont une personne anti-vaccination, elles ne font tout simplement pas confiance au vaccin. Mais on comprend aussi que les gens qui sont défavorables avec tous les vaccins auront tendance à rejeter les vaccins, et ils expriment leur point de vue sur les nouveaux vaccins, donc ils auront un potentiel plus que d'autres, devenu anti-vaccination.

Comme nous le disions aussi dans la théorie du Cadre sur l'anti-vaccination, que la définition de l'anti-vaccination, de nombreux sociologues la trouve aussi trop stricte, on peut utiliser le mot méfiance, doute ou défavorable à la vaccination. Ce sont des gens qui utilisent leur énergie pour répandre leurs pensées contre les vaccinations, ou certaines substances dans certains vaccins, à d'autres.

Ainsi, être anti-vaccination, ou défavorable à la vaccination est à la fois une référence à une perte de confiance dans la question de la vaccination.

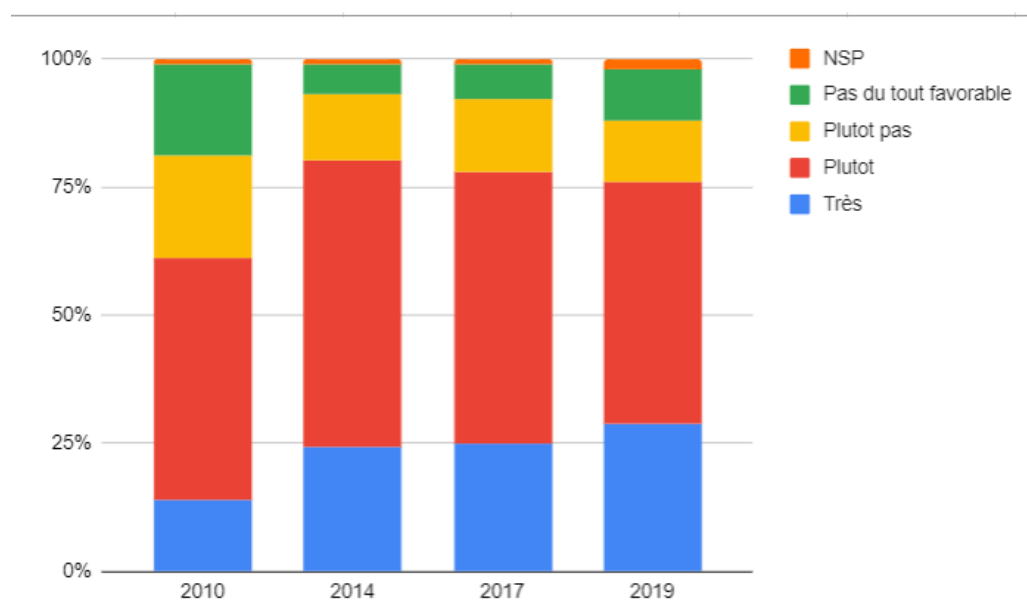


Figure 5 Source: Barometres de Sante (2010-2019)

Nous voyons un changement dans la pensée favorable à la vaccination. Plus de 75 % des personnes interrogées étaient favorables à la vaccination, au lieu de 60 % comme en 2010. La confiance dans la vaccination a été rétablie après des scandales autour du vaccin contre l'hépatite B, ou les vaccins contre la rougeole, les oreillons et la rubéole. Cependant, 10% des répondants qui ne sont pas du tout favorables à la vaccination disent que c'est une

augmentation par rapport à 2014 et 2017. De plus, en 2019, environ 2% des répondants étaient défavorables à tous les vaccins, ce chiffre en 2017 était de 1,5%. Si la confiance dans la vaccination a été restaurée, la résurgence de la proportion de personnes totalement défavorables à la vaccination est également inquiétante. Cela représente une polarisation dans la question de la vaccination.

B.2 Différenciation dans les caractéristiques sociales à l'opinion de la vaccination.

Sexe.

Il y a égalité des sexes dans l'opinion concernant la vaccination. Les hommes et les femmes ont des ratios de semence assez similaires, hormis l'avis très favorable sur la vaccination, les hommes font légèrement mieux que les femmes (16,9 et 12,7 en 2010 ; 25,4 et 23,5 en 2014) ; 26,9 et 24,2 en 2017 ; 31,6 et 27,2 en 2019), mais cela se compense en faveur du taux d'opinion plutôt favorable pour trouver un équilibre entre les deux sexes en opinion favorable versus la vaccination en général.

Age.

Selon le Baromètre de la Santé 2019, 33,9% des personnes de plus de 65 ans sont très favorables à la vaccination en général, ce qui est supérieur à la moyenne des quatre tranches d'âge (18-25 ans, 25-45 ans, 45- 65 ans et plus de 65 ans) est de 29,2%, mais aussi 13,1% des personnes de plus de 65 ans ne sont pas complètement favorables, dans la tranche d'âge des 18-25 ans ce chiffre est de 5,1%, et chez les 25-45 ans c'est 8,7%. Cela peut s'expliquer par le fait que les personnes âgées sont souvent plus fragiles que les âges précédents, leurs attitudes et opinions peuvent donc être plus extrêmes.

Diplôme.

Outre l'égalité des deux caractéristiques sociales ci-dessus, il existe une différence marquée d'opinions sur la vaccination lorsque l'on compare les diplômés. Jusqu'à 86% des personnes titulaires d'un master ou plus sont favorables à la vaccination en général, ce chiffre n'est que de 66% chez les personnes sans diplôme. Cela signifie que près de 9 personnes sur 10 titulaires d'une maîtrise ou plus ont confiance dans les vaccinations en général. En termes de vaccin défavorable, seulement 28% des personnes titulaires d'une maîtrise ou plus étaient défavorables à certains vaccins, contre 36% des personnes sans diplôme. Les taux favorables à la vaccination augmentent également avec le degré.

VI. Conclusion.

Notre article de recherche vise à mieux comprendre les opinions, les attitudes et les changements d'attitude des utilisateurs de Twitter envers les vaccins contre le Covid-19. L'épidémie de Covid-19 qui sévit depuis le début de l'année 2020 a eu un grand impact sur notre vie sociale. Par conséquent, il est également devenu le sujet de nombreux articles et rapports scientifiques. Analyser l'attitude et le comportement d'un individu face aux impacts de la société, surtout lorsqu'ils le montrent sur une plateforme numérique, est également un sujet nouveau et intéressant. Notre mémoire de recherche part du contexte de recherche, du cadre théorique de sujets de recherche connexes tels que la sociologie des émotions, des sentiments, ou encore le rappel des facteurs sociaux affectant le comportement et les décisions individuelles. Vient ensuite la section méthodologie où nous apprenons à savoir comment nous l'utilisons, en incorporant l'apprentissage automatique, l'analyse Times Seris à la sociologie. C'est aussi l'un des objectifs, et notre souhait, lorsqu'en incorporant des méthodes d'apprentissage automatique, ou d'apprentissage profond et éventuellement d'intelligence artificielle, à l'étude des questions sociologiques, aidera les chercheurs à gagner du temps et à gagner en efficacité. Et nous avons des résultats intéressants.

Premièrement, on observe une évolution des opinions et des sentiments des utilisateurs de Twitter envers les vaccins Covid-19, mais pas seulement, ils intègrent également des émotions avec les vaccinations Covid-19-19, voire avec le gouvernement dans ses lignes de statut. Selon les résultats de l'étude, la transition émotionnelle du positif au négatif reflète en partie ce qui se passe dans la réalité. Quand il y a eu les premières informations sur le vaccin contre le covid-19, la joie et l'espoir que la maladie allait bientôt disparaître sont apparus, puis sont progressivement devenus négatifs, quelque peu indignés quand le bruit, les vaccins ont des effets secondaires. Mais les gens répondront-ils simplement aux vaccins Covid-19 ?

La communauté en ligne a également réagi aux décisions gouvernementales, à l'ambiguïté, aux retards ou aux désaccords (renversement des opinions des utilisateurs de Twitter depuis l'arrêt de la vaccination d'AstraZeneca). C'est dans le rapport du gouvernement français qu'il reconnaît le retard ou l'erreur de calcul de la gravité de l'épidémie.

Non seulement cela reflète les frustrations de la société, mais les tweets en ligne contiennent également des messages politiques. C'est le deuxième résultat.

Non seulement les tweets énonçant des attitudes et des opinions, de nombreuses personnes profitent de cette situation pour publier des lignes de statut émotionnel politique et influencer la communauté en ligne. Quant à savoir si la politique est la principale influence sur le comportement et les attitudes des utilisateurs, nous ne sommes pas sûrs. Mais l'existence de la politique est indiscutable (soutenir d'autres candidats, s'opposer au gouvernement actuel). Ce deuxième résultat confirme également que Twitter est une arme de communication efficace pour les militants.

Le troisième résultat que nous obtenons est l'analyse du comportement des utilisateurs. Selon les statistiques du gouvernement français, 16% des personnes ne souhaitent pas se faire vacciner contre le Covid-19. et selon le Baromètre de la Santé, en 2019, la

proportion de personnes totalement défavorables à la vaccination représentait 10 %, et 2 % de personnes étaient défavorables à tous les vaccins. Mais sur Twitter, nous avons jusqu'à 54% de tweets négatifs. Cela montre que, bien qu'ils soient minoritaires, ceux qui sont favorables, défavorables ou anti-vaccin au Covid-19 sont la partie sociale la plus active sur Twitter. Et saura-t-on localiser, ou identifier, les caractéristiques de ce groupe social ?

En se basant sur la saisonnalité des tweets, plus l'analyse de la partie B, on peut en déduire que le groupe social le plus actif se caractérise par beaucoup de temps libre dans la journée, ce qui signifie que l'on est les groupes sociaux les plus actifs qui sont ceux qui sont au chômage au moment où ils rejoignent l'affichage des lignes d'état. Le prochain groupe actif est celui des militants politiques et sociaux, ils utilisent Twitter comme un lieu pour diffuser leurs pensées, ils investissent du temps et des efforts dans cette plateforme.

Voici les résultats obtenus à partir de l'étude. Au cours du processus de rédaction du rapport, j'ai rencontré quelques difficultés car l'utilisation des données d'un site de réseautage social entraînerait un manque de variables sociales. Ensuite, mais mentionné, l'analyse des émotions et des opinions est un aspect difficile à la fois en sociologie et en apprentissage automatique, donc le temps n'est pas un petit problème, si ce sujet continue à étudier, nous pouvons construire un modèle d'apprentissage automatique avec une plus grande précision qui est capable de mieux analyser la situation que la méthode actuelle. Et enfin, le manque de données pertinentes, si elles sont disponibles, j'aimerais observer certaines données politiques telles que l'élection présidentielle française de 2022, si ces données sont disponibles sur Twitter, l'argument politique est incontestable.

Bibliographie :

- Morelle A. (1996), *La défaite de la santé publique*, p.24
- Lawrence K. Altman (consulté le 25 mai 2021), "RARE CANCER SEEN IN 141 HOMOSEXUALS", *The New York Times*, 3 juillet 1981; <https://www.nytimes.com/1981/07/03/us/rare-cancer-seen-in-41-homosexuals.html>
- Mesko, B. Conclusion. (2013) In *Social Media in Clinical Practice*, p.139-141
- Pang, B. and Lee, L. (2008) *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval
- Enrico Bergamini (2020), *How COVID-19 is laying bare inequality*
- Blumenshine et al. (2008), *Sydenstricker, 1931*
- Jacques Boucher de Perthes (1863). *Sous dix rois: souvenirs de 1791 à 1860*, vol 5.
- Massimo Borlandi (2017), *Sociologie et histoire : la contribution de Durkheim*, p.83-104 Jacob
- Carmichael, Jason (2012), *Social Control*.
- Cayton, H. The alienating language of health care (2006), *J. R Soc. Med*
- Conley, Dalton (2017), *You may ask yourself: An Introduction to thinking like a sociologist*, p.197
- Boullier, D., & Lohard, A. (2012). *Opinion mining et Sentiment analysis*. S.1. : OpenEdition Press
- Devlin et al, BERT: *Pre-training of Deep Bidirectional Transformers for Understanding*
- Abderrahman Essebbar et al. (2021), *Aspect Based Sentiment Analysis using French Pre-Trained Models*
- Cambria, E., and White, B. (2014). *Jumping NLP curves: A review of natural language processing research*. IEEE Computational Intelligence Magazine, p48-57
- Lindzey, Gardner (Ed) (1954) *Handbook of Social psychology. I. Theory and method. II. Special fields and applications*
- Goncalves et al., (2013). *Comparing and combining sentiment analysis methods*. In Proceedings of the first ACM conference on Online social networks (pp.27-38)
- Hollingshead et al. (1941), *The Concept of Social Control*. *American Sociological Review*, p217-224.
- Macionis John, J. (1987), *Sociologie*, p.154
- Michelle Zancarini-Fournel (2016). *Les luttes et les rêves: Une histoire populaire de la France de 1685 à nos jours*. p.251, chap 7.
- Loi du 15 février 1902 relative à la protection de la santé publique, JO du 19 février 1902 ; Bourdelais P., *Les épidémies terrassées, une histoire de pays riches*, Éd. de la Martinière, 2003.
- Dominique Mesda (2020), *France Culture*.
- Mauss M (1950). *Rapport réels et pratiques de la psychologie et de la sociologie*
- Thierry Mainaud (2020), *Les inégalités sociales à l'épreuve de la crise sanitaire*, p.18
- Christopher Manning and Hinrich Schütze (1999). MIT Press. *Foundations of Statistical Natural Language Processing*.
- MacDonald NE (2015). *Vaccine hesitancy: definition, scope and determinants*

Nathalie Vallerand (2014), *Magazine Naitre et grandir*

Weissmann (2012), *G. Epigenetics in the Age of Twitter: Pop Culture and Modern Science*

Hobson-West (2007). *Trusting blindly can be the biggest risk of all: organized resistant to childhood vaccination in the UK.*

Données référencées d'Attention Is All You Need (2018)

<http://nlp.seas.harvard.edu/2018/04/03/attention.html>

Documents parlementaires, Chambre, annexe n° 5131, séance du 31 octobre 1918, p.1793.

https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/dares_acemo_covid19_synthese_17-04-2020.pdf

Annexes:

Code de Analyse des sentiments

```
In [1]: # Importing standard libraries for every machine/deep Learning pipeline
import pandas as pd
import torch
from tqdm import tqdm, trange
import numpy as np

# Importing specific libraries for data preprocessing, model architecture choice, training and evaluation
from sklearn.model_selection import train_test_split
from keras.preprocessing.sequence import pad_sequences
from torch.utils.data import TensorDataset, DataLoader, RandomSampler, SequentialSampler
from transformers import CamembertTokenizer, CamembertForSequenceClassification
from transformers import AdamW
import torch.optim as optim
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import seaborn as sns
```

```
In [3]: df=pd.read_csv('C:\\Users\\User\\datafinal.csv')[['author_id','name','username',
                                                         'text','created_at','Sexe']]
```

```
In [4]: df.head()
```

```
Out[4]:
```

	author_id	name	username	text	created_at	Sexe
0	1.330000e+18	Alexandre M A Caron	Alexand14248460	Tellement fier d'être vacciné trois fois contr...	2022-04-29T15:47:14.000Z	M
1	1.330000e+18	Alexandre M A Caron	Alexand14248460	@JeMouth Tellement fier d'être vacciné trois f...	2022-04-29T15:33:49.000Z	M
2	1.469749e+09	GeB. 🍌🍌🍌😘😘😘	GerardBondeau	@DrElDavid @idrissaberkane 📌 Si... #Moderna b...	2022-04-29T08:46:59.000Z	M
3	1.469749e+09	GeB. 🍌🍌🍌😘😘😘	GerardBondeau	@JeanYvesCAPO @maillardjeanch3 @alainhouper P...	2022-04-26T19:57:10.000Z	M
4	1.469749e+09	GeB. 🍌🍌🍌😘😘😘	GerardBondeau	"L'ancien responsable de la rech respiratoire ...	2022-04-21T17:42:44.000Z	M

Activat
Go to Set

```
In [6]: # Defining constants
epochs = 5
MAX_LEN = 128
batch_size = 16
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
```

```
In [7]: # Load the dataset, I selected only 5000 sample because of memory limitation
df2 = pd.read_csv('C:\\Users\\User\\Desktop\\Base des donnees\\french_tweets.csv').sample(5000).reset_index(drop=True)
df2.head()
```

```
Out[7]:
```

	label	text
0	0	Arrête de pleurer ton bonheur
1	0	J'aimerais être en stockholm maintenant ...
2	0	Fait une mission
3	0	En regardant le spectacle de jay leno dernier ...
4	0	J'ai enseigné jordan à prendre la tête ce soir...

```
In [8]: # Initialize CamemBERT tokenizer
tokenizer = CamembertTokenizer.from_pretrained(
    'camembert-base', do_lower_case=True)
```

```
In [9]: # Creates List of texts and Labels
text = df2['text'].to_list()
labels = df2['label'].to_list()

#user tokenizer to convert sentences into tokenizer
input_ids = [tokenizer.encode(sent, add_special_tokens=True, max_length=MAX_LEN) for sent in text]

# Pad our input tokens
input_ids = pad_sequences(input_ids, maxlen=MAX_LEN, dtype="long", truncating="post", padding="post")

# Create attention masks
attention_masks = []
# Create a mask of 1s for each token followed by 0s for padding
for seq in input_ids:
    seq_mask = [float(i>0) for i in seq]
    attention_masks.append(seq_mask)
```

Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to `truncation`.

```
In [10]: # Use train_test_split to split our data into train and validation sets for training
train_inputs, validation_inputs, train_labels, validation_labels, train_masks, validation_masks = train_test_split(input_ids, labels,
                                                                                                                random_state=42, test_size=0.2)

# Convert all of our data into torch tensors, the required datatype for our model
train_inputs = torch.tensor(train_inputs)
validation_inputs = torch.tensor(validation_inputs)
train_labels = torch.tensor(train_labels)
validation_labels = torch.tensor(validation_labels)
train_masks = torch.tensor(train_masks)
validation_masks = torch.tensor(validation_masks)

# Create an iterator of our data with torch DataLoader. This helps save on memory during training because, unlike a for loop,
# with an iterator the entire dataset does not need to be loaded into memory

train_data = TensorDataset(train_inputs, train_masks, train_labels)
train_sampler = RandomSampler(train_data)
train_dataloader = DataLoader(train_data, sampler=train_sampler, batch_size=batch_size)

validation_data = TensorDataset(validation_inputs, validation_masks, validation_labels)
validation_sampler = SequentialSampler(validation_data)
validation_dataloader = DataLoader(validation_data, sampler=validation_sampler, batch_size=batch_size)
```

Activat

```
In [11]: # Load BertForSequenceClassification, the pretrained BERT model with a single linear classification layer on top.
model = CamembertForSequenceClassification.from_pretrained("camembert-base", num_labels=2)
model.to(device)
```

Some weights of the model checkpoint at camembert-base were not used when initializing CamembertForSequenceClassification: ['roberta.pooler.dense.weight', 'lm_head.layer_norm.weight', 'lm_head.dense.bias', 'lm_head.decoder.weight', 'roberta.pooler.dense.bias', 'lm_head.layer_norm.bias', 'lm_head.bias', 'lm_head.dense.weight']
- This IS expected if you are initializing CamembertForSequenceClassification from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing CamembertForSequenceClassification from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).
Some weights of CamembertForSequenceClassification were not initialized from the model checkpoint at camembert-base and are newly initialized: ['classifier.dense.bias', 'classifier.out_proj.weight', 'classifier.dense.weight', 'classifier.out_proj.bias']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

```
Out[11]: CamembertForSequenceClassification(
  (roberta): RobertaModel(
    (embeddings): RobertaEmbeddings(
      (word_embeddings): Embedding(32005, 768, padding_idx=1)
      (position_embeddings): Embedding(514, 768, padding_idx=1)
      (token_type_embeddings): Embedding(1, 768)
      (layer_norm): LayerNorm((768), eps=1e-05, elementwise_affine=True)
```

```
In [12]: param_optimizer = list(model.named_parameters())
no_decay = ['bias', 'gamma', 'beta']
optimizer_grouped_parameters = [
    {'params': [p for n, p in param_optimizer if not any(nd in n for nd in no_decay)],
     'weight_decay_rate': 0.01},
    {'params': [p for n, p in param_optimizer if any(nd in n for nd in no_decay)],
     'weight_decay_rate': 0.0}
]

optimizer = AdamW(optimizer_grouped_parameters, lr=2e-5, eps=10e-8)

# Function to calculate the accuracy of our predictions vs labels
def flat_accuracy(preds, labels):
    pred_flat = np.argmax(preds, axis=1).flatten()
    labels_flat = labels.flatten()
    return np.sum(pred_flat == labels_flat) / len(labels_flat)
```

C:\Users\User\anaconda3\lib\site-packages\transformers\optimization.py:306: FutureWarning: This implementation of AdamW is deprecated and will be removed in a future version. Use the PyTorch implementation torch.optim.AdamW instead, or set `no_deprecation_warning=True` to disable this warning
warnings.warn(

Activate

```
In [13]: # Store our loss and accuracy for plotting if we want to visualize training evolution per epochs after the training process
train_loss_set = []

# trange is a tqdm wrapper around the normal python range
for _ in trange(epochs, desc="Epoch"):
    # Tracking variables for training
    tr_loss = 0
    nb_tr_examples, nb_tr_steps = 0, 0

    # Train the model
    model.train()
    for step, batch in enumerate(train_dataloader):
        # Add batch to device CPU or GPU
        batch = tuple(t.to(device) for t in batch)
        # Unpack the inputs from our dataloader
        b_input_ids, b_input_mask, b_labels = batch
        # Clear out the gradients (by default they accumulate)
        optimizer.zero_grad()
        # Forward pass
        outputs = model(b_input_ids, token_type_ids=None, attention_mask=b_input_mask, labels=b_labels)
        # Get Loss value
        loss = outputs[0]
        # Add it to train Loss list
        train_loss_set.append(loss.item())
        # Backward pass
        loss.backward()
        # Update parameters and take a step using the computed gradient
        optimizer.step()

        # Update tracking variables
        tr_loss += loss.item()
        nb_tr_examples += b_input_ids.size(0)
        nb_tr_steps += 1

    print("Train loss: {}".format(tr_loss/nb_tr_steps))
```

Act
Go t

```
# Tracking variables for validation
eval_loss, eval_accuracy = 0, 0
nb_eval_steps, nb_eval_examples = 0, 0
# Validation of the model
model.eval()
# Evaluate data for one epoch
for batch in validation_dataloader:
    # Add batch to device CPU or GPU
    batch = tuple(t.to(device) for t in batch)
    # Unpack the inputs from our dataloader
    b_input_ids, b_input_mask, b_labels = batch
    # Telling the model not to compute or store gradients, saving memory and speeding up validation
    with torch.no_grad():
        # Forward pass, calculate logit predictions
        outputs = model(b_input_ids, token_type_ids=None, attention_mask=b_input_mask, labels=b_labels)
        loss, logits = outputs[:2]

    # Move logits and labels to CPU if GPU is used
    logits = logits.detach().cpu().numpy()
    label_ids = b_labels.to('cpu').numpy()

    tmp_eval_accuracy = flat_accuracy(logits, label_ids)

    eval_accuracy += tmp_eval_accuracy
    nb_eval_steps += 1

print("Validation Accuracy: {}".format(eval_accuracy/nb_eval_steps))
```

Epoch: 20% ██████	1/5 [58:16<3:53:05, 3496.35s/it]
Validation Accuracy: 0.7698412698412699 Train loss: 0.4124925308227539	
Epoch: 40% ██████	2/5 [1:58:34<2:58:24, 3568.15s/it]
Validation Accuracy: 0.7926587301587301 Train loss: 0.29708193969726565	
Epoch: 60% ██████	3/5 [2:57:55<1:58:49, 3564.86s/it]
Validation Accuracy: 0.7926587301587301 Train loss: 0.1955599757358432	
Epoch: 80% ██████	4/5 [3:57:02<59:17, 3557.87s/it]
Validation Accuracy: 0.7807539682539683 Train loss: 0.13299895744770765	
Epoch: 100% ██████	5/5 [4:56:55<00:00, 3563.17s/it]

Tableau de tris croisés des Baromètres de Santé

```
> cprop(table(baro2010$RS_SEXE,baro2010$VA1,useNA="no"))
```

	Tres	Plutot	Plutot pas	Pas du tout favorable
Homme	52.0	42.6	44.1	45.7
Femme	48.0	57.4	55.9	54.3
Total	100.0	100.0	100.0	100.0

	5	All
Homme	59.6	44.9
Femme	40.4	55.1
Total	100.0	100.0

```
> rprop(table(baro2010$RS_SEXE,baro2010$VA8A,useNA="no"))
```

	Oui	Non	NSP
Homme	45.8	52.8	0.8
Femme	58.9	39.6	0.8
All	53.0	45.5	0.8

	[Oui, à toutes les vaccinations en général]
Homme	0.6
Femme	0.7
All	0.7

```
> rprop(table(baro2019$qage, baro2019$val, useNA="no"))
```

	Tres	Plutot	Plutot pas	
De 18 à 25 ans	29.2	52.2	10.7	
De 25 à 45 ans	26.7	49.7	13.6	
De 45 à 65	27.9	48.0	12.6	
Plus de 65 ans	33.9	40.8	11.2	
All	29.2	46.9	12.4	

	Pas du tout favorable	5	Total
De 18 à 25 ans	5.1		2.8 100.0
De 25 à 45 ans	8.7		1.3 100.0
De 45 à 65	10.0		1.5 100.0
Plus de 65 ans	13.1		1.0 100.0
All	10.1		1.4 100.0

	Tres	Plutot	Plutot pas	
Aucun Diplome	25.5	41.0	9.3	
Diplome	24.2	45.7	14.4	
Bac	28.7	49.4	11.4	
Bac +2,+3	28.2	51.0	13.8	
Master, Doc et plus	41.0	44.8	9.1	
Autre	33.3	50.0	0.0	
15	16.7	33.3	16.7	
16	71.4	0.0	14.3	
All	29.2	46.9	12.4	

	Pas du tout favorable	5	Total
Aucun Diplome	21.7		2.5 100.0
Diplome	14.4		1.2 100.0
Bac	8.6		1.9 100.0
Bac +2,+3	5.7		1.3 100.0
Master, Doc et plus	4.2		1.0 100.0
Autre	16.7		0.0 100.0
15	33.3		0.0 100.0
16	14.3		0.0 100.0
All	10.1		1.4 100.0