University of East London

MSc. In Data Science

DATA ECOLOGY

DS7001

COURSEWORK &PORTFOLIO

BY

DAMILOLA OLUWASENI OMISORE

2272153

ON

DATA SCIENCE IN THE AGE OF COVID:

Artificial Intelligence in Public Perspective of COVID-19 Vaccine.

ABSTRACT

The primary method of preventing communicable infectious diseases is vaccination, but historically, the public has been reluctant to get vaccines, and the COVID-19 vaccine is no different. It has been established that social media use contributes to the low acceptability of vaccines. To better understand public perceptions, concerns, and emotions that may affect the achievement of mental strength, this study aims to review how artificial intelligence was able to identify the topics and sentiments in the public discussion on social media about the COVID-19 vaccine as well as any salient changes in those topics and sentiments over time.

Methods: Research works and analysis carried out during the pandemic to identify the public opinion on the COVID-19 vaccine will be critically reviewed. Methods used in the retrieval of data such as Twitter API will be discussed, applications used for sentiment analysis like VADER (Valence Aware Dictionary for Sentiment Reasoning) will be reviewed, and methods used for topic modelling such as LDA (Latent Dirichlet Allocation) will also be discussed. At the end of the review, significant drawbacks and future advancements will be highlighted.

Results: Sentiment analysis has been applied to various areas over time like social media monitoring, automobile/vehicle-based services, product/service review, travel/tourism-based services, healthcare-based services, and financial services. As for topic modelling, it is likely to remain a crucial method for examining and comprehending enormous volumes of text data in the future. The future of topic modelling analysis is expected to be shaped by several factors. One tendency is the expansion of massive datasets' accessibility and the improvement of computer resources, which will allow for the creation of more advanced and precise subject-modelling methods. Finally, the increased interest in multilingual and cross-lingual analysis will probably spur the creation of word clouds that facilitate cross-lingual comparison and analysis and can handle several languages.

Conclusions: This work has discussed the essential use of Artificial Intelligence and other applications during COVID-19 to understand the public perception of the COVID-19 vaccine. Overall, the development of more potent computational resources, the growing relevance of interpretability and explainability in machine learning, and the expanding availability of huge datasets are anticipated to be the driving forces behind the future of AI in sentiment analysis and topic modelling.

Table of Contents

Table of Figures

INTRODUCTION

A virus that causes the communicable illness (COVID-19) was first discovered in Wuhan, China, in December 2019. Then there was a spread of the illness throughout the world, causing the COVID-19 pandemic(*COVID-19 - Wikipedia*, n.d.). COVID-19 has significantly impacted people, communities, and nations all across the world. The virus has affected many parts of daily life, including work, education, travel, and social relationships. It has also caused widespread disease and death. Numerous firms and industries have suffered considerable losses as a result of COVID-19, and many individuals have lost their employment or are now struggling economically. The epidemic has also significantly impacted healthcare systems, as hospitals and medical staff have struggled to keep up with the demand for care and the sheer volume of cases(Kaye et al., 2021). The pandemic's social and psychological effects were extremely profound since people had to adjust to new lifestyles and deal with the stress and worry the crisis brought about. Due to certain groups being disproportionately afflicted by the virus, such as underprivileged populations and critical workers, the epidemic has also brought to light and worsened already existing inequities(Saladino et al., 2020).In general, COVID-19 has had a significant influence that will last for some time. In addition to governments and organizations working together to address the myriad issues posed by the pandemic, the vaccine was developed to protect against COVID-19. By injecting a tiny amount of the virus's genetic material into the body, these vaccines stimulate the immune system to create antibodies against the pathogen. If the person is exposed to the virus again in the future, these antibodies aid in preventing infection. Several COVID-19 vaccines are currently available and approved for usage in several international locations. These include the Johnson & Johnson vaccine, the Coronavac/Sinovac vaccine, the Pfizer/BioNTech vaccine, the Moderna vaccine, the AstraZeneca/Oxford vaccine, and others. Each of these vaccinations has completed clinical testing to prove their efficacy and safety, and it has been established that they are quite effective in preventing COVID-19(Dr. Mersad Alimoradi, 2022). But these vaccines are not perfect and have been reported to have side effects that only lasted for a few days which have been mild to moderate. Common side effects include injection site discomfort, fever, fatigue, headache, muscle soreness, chills, and diarrhoea. Various side effects have various post-vaccination risks depending on the specific vaccine(WHO, n.d.). And this has affected the perception of the public on getting the vaccine. The COVID-19 vaccinations reputation among the public might vary greatly. Some people could be eager to get vaccinated because they view it as a critical step in stopping the pandemic and resuming regular life. Others could be cautious or uncertain about receiving the vaccination due to worries about its safety or effectiveness, as well as due to misinformation or mistrust of the process used to create and distribute vaccines.

This work aims to review the technology used to analyze or identify the public opinion or response toward the COVID-19 vaccine. The focus of this review would be on social media, specifically Twitter. The reason for focusing on Twitter specifically is because it is a diverse platform that supports a variety of elements that fuel interactions like texts, photos, videos, hashtags, links, and numerals (for example sports scores or official statistics) and allows to easy pinpointing of the desired audience. Ultimately, Twitter is a network with distinct features that, in contrast to other social media platforms, allows both users and companies to let loose, develop connections, and maximize engagement(Gina Mueller, n.d.). A variety of research has been carried out on public perception which is what will be critically reviewed in this work.

METHODOLOGY

For this work, I will be reviewing the algorithms used to analyze public sentiment or public perception of the COVID-19 vaccine on Twitter. To carry out an analysis of public sentiment, data collection and pre-processing would have been put in place to retrieve tweets as datasets. First is a critical review of the applications that have been used to extract tweets regarding the COVID-19 vaccine such as the Twitter API, and Snscrape. After the data collection and pre-processing, the next phase would have been to carry out the sentiment analysis to identify and extract opinions, emotions, and evaluations expressed. Then, the critical review of the algorithms used for the sentiment analysis such as the VADER (Valence Aware Dictionary for Sentiment Reasoning) algorithm, and NRCLex (National Research Council Canada Lexicon). Then the topic modelling analysis would come after to identify the main topic or themes present in the group and the model used for this such as LDA (Latent Dirichlet Allocation) model will be critically reviewed. Finally, the size of a term reflects its frequency and popularity, and I will discuss the use of word cloud maps across predefined phases depending on the frequency of keywords appearing in Tweet content.

After the critical review and discussion, I am going to highlight the strength and weaknesses of the methods and the advantages and disadvantages of using the methods. Will also point out the future of these methods.

CRITICAL REVIEW

Data collection and pre-processing

The major source of data for public perception is Twitter as it is a well-known social media site with a sizable user base that produces a lot of material in the form of tweets. This content contains a variety of information that can be helpful for several purposes, including news, opinions, personal experiences, and more. Twitter's real-time information-sharing capabilities are one of the factors that make it such a significant data source for monitoring the most recent trends and developments across a wide range of issues. Twitter also offers APIs (Application Programming Interfaces), which make it simpler for developers to access and analyze the data created on the network and create applications that leverage it.

To retrieve tweets, Twitter provides the Streaming API and Search API for data collection. Real-time data is gathered using the Streaming API, while historical data is retrieved using the Search API. One of the three tiers (Standard, Premium, and Enterprise) can be used to access the Search API. A sample of recent Tweets that have been published in the last week is available using the Standard Search API. Due to their high costs, the other two options (Premium and Enterprise) are not frequently used by researchers. Real-time Tweet collection is made possible by streaming API. There are two choices when the goal is to gather all Tweets, regardless of their content which are the Free Streaming API and the Decahose Streaming API(Twitter Developer Platform, n.d.). Twitter API can be used in multiple programming languages like Python, R, and Java.

Another algorithm that was also adopted for data collection is Snscrape which is a scraper for social networking platforms (SNS). By scraping items like user profiles, hashtags, or searches, it is able to obtain the discovered objects, such as the relevant posts. Snscrape has varying services depending on the platform. For Twitter, it supports retrieving user profiles, hashtags, searches, tweets (single or surrounding thread), list posts, and trends. It also supports services for platforms such as Instagram, Facebook, Telegram, Reddit, Mastodon, Weibo, and VKontakte. Snscrape is a python package and

does not work for other programming languages. It requires having a Python 3.8 or higher version with installed python package dependencies such as libxml2 and libxslt(Snscrape, n.d.).

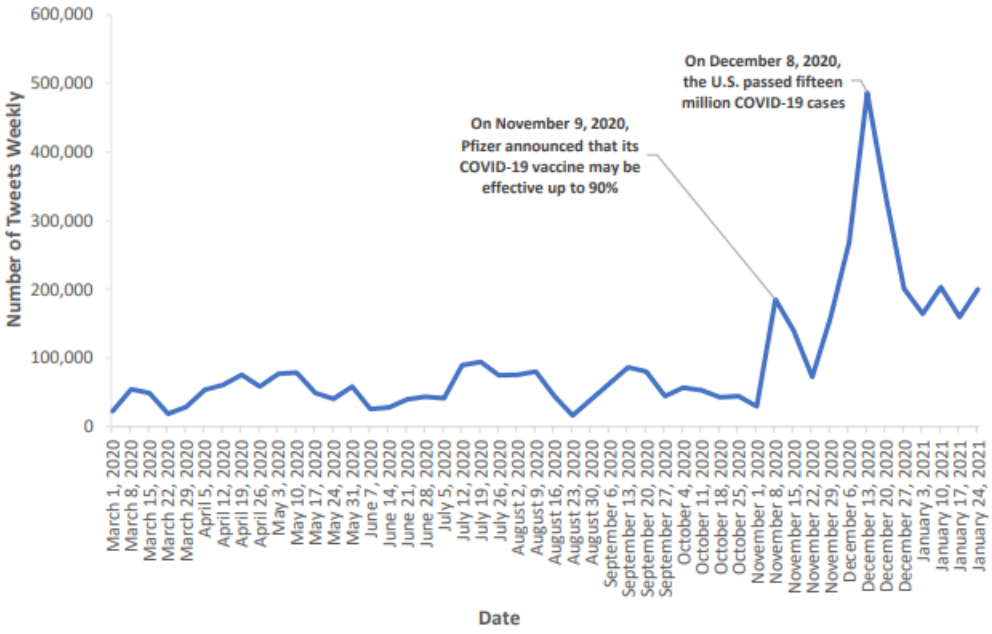

***Figure 1**: Total number of tweets mentioning COVID-19 vaccines each week in the US using Twitter API*(Xie et al., 2021).

Following the tweets' recovery, it was discovered that between March 5, 2020, and January 31, 2021, there were an estimated 4,438,203 tweets per week connected to the COVID-19 vaccine. Figure 1 shows how the conversation about COVID-19 vaccinations fluctuated and was lively in the beginning, but there were two significant peaks at the conclusion of 2020. The first peak was around November 9, 2020, when Pfizer announced that their coronavirus antibody may be effective up to 90% of the time. The second peak occurred on or around December 8, 2020, when there were around fifteen million COVID-19 cases in the United States.(Xie et al., 2021).

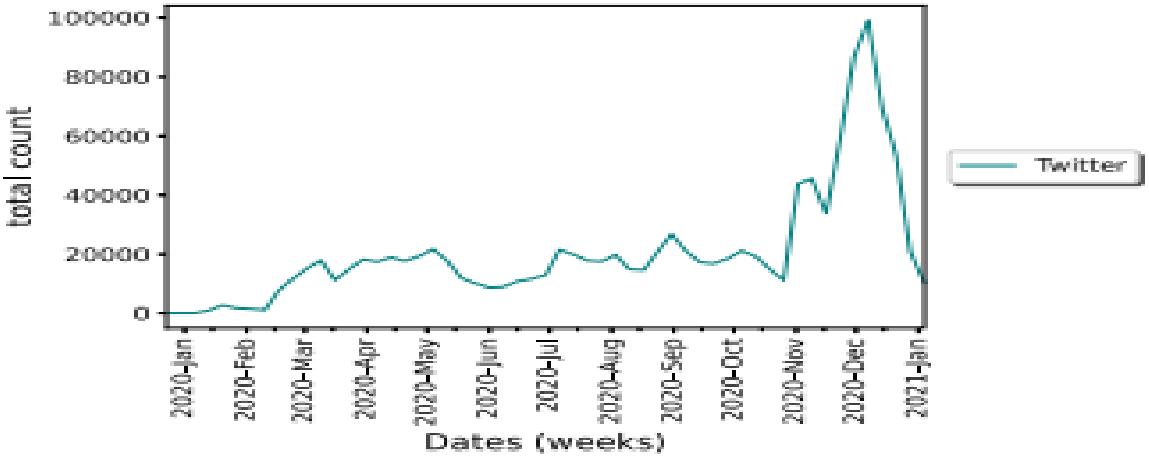

***Figure 2**: Distribution of COVID-19 vaccine-related tweets using Snscrape*(Yousefinaghani et al., 2021).

The above is a result of the visualization of the tweets retrieved from Twitter. After the data collection, it was observed that between January 7, 2020, and January 3, 2021, inclusive, 4,552,652 tweets were gathered. The second week of December saw the most tweets, with over 500,000 tweets being sent out. Due to news about the possibility that a tuberculosis vaccine could aid in the fight against SARS-CoV-2, there was a minor rise in the number of tweets on vaccines in June (weeks 13, 14, and 15). Most Twitter users wanted to know if COVID-19 could be stopped by using this vaccination, which is already available. The revelation of progress in the development of several vaccinations, however, was what caused the second surge at the beginning of November(Yousefinaghani et al., 2021).

Sentiment Analysis

Sentiment analysis, sometimes referred to as opinion mining or emotion artificial intelligence, is the methodical detection, extraction, measurement, and study of emotional states and subjective data using natural language processing, text analysis, computational linguistics, and biometrics. Sentiment analysis is frequently used in marketing, customer service, and clinical medical applications. It is used with consumer feedback products including testimonials and survey results, internet and social media material, and healthcare resources(*Sentiment Analysis - Wikipedia*, n.d.). The three primary approaches to sentiment analysis methods currently in use are knowledge-based techniques, statistical techniques, and hybrid methods(Cambria et al., 2013). Knowledge-based techniques group words based on the existence of clear-cut affect terms like joyful, sad, fearful, and bored, and categorize text by affect categories. Some knowledge bases feature words that have an influence, but they also give random words a potential "affinity" with certain emotions(Stevenson et al., 2007). Latent Semantic Analysis, Support Vector Machines, "bag of words" model, "Pointwise Mutual Information" for Semantic Orientation, semantic space models or word embedding models, and deep learning are examples of machine learning techniques that are used in statistical methodologies(*Sentiment Analysis - Wikipedia*, n.d.). Hybrid techniques combine both machine learning and knowledge representation components like ontologies and semantic networks to discover subtle semantics, such as through the examination of concepts that do not clearly transmit significant information but are implicitly related to other concepts that do(Cambria & Hussain, 2015). The following is a basic workflow for sentiment classification: dividing the data into test and training data, choosing a model architecture, training the model using "training data," evaluating the model's performance using "test data," and applying the trained model to new data to make predictions, which will be between -1.0 and 1.0 in this instance(VADER - Swayanshu, n.d.).

One of the major algorithms used for sentiment analysis is the Valence Aware Dictionary and Sentiment Reasoner (VADER) which will be reviewed along with a sample result of the analysis. VADER is a lexicon and rule-based tool for sentiment analysis that is tailored to social media-specific sentiments. A sentiment lexicon is a list of lexical features (such as words) that are typically labelled as either positive or negative based on their semantic orientation which is a combination VADER employs. Not only does VADER tell us about the score for positivity and negativity, but it also tells us how positive or negative a feeling is(VADER - GeeksforGeeks, n.d.). Note that the sentiment score ranges from -1.0 to 1.0. Very good sentiment is indicated by a high positive sentiment score, whilst more negative sentiment is indicated by a high negative sentiment score. Positive sentiment was defined as the sentiment score between 0.05 and 1.0, negative sentiment as between -1.0 and 0.05, and neutral sentiment as everything else(Xie et al., 2021).

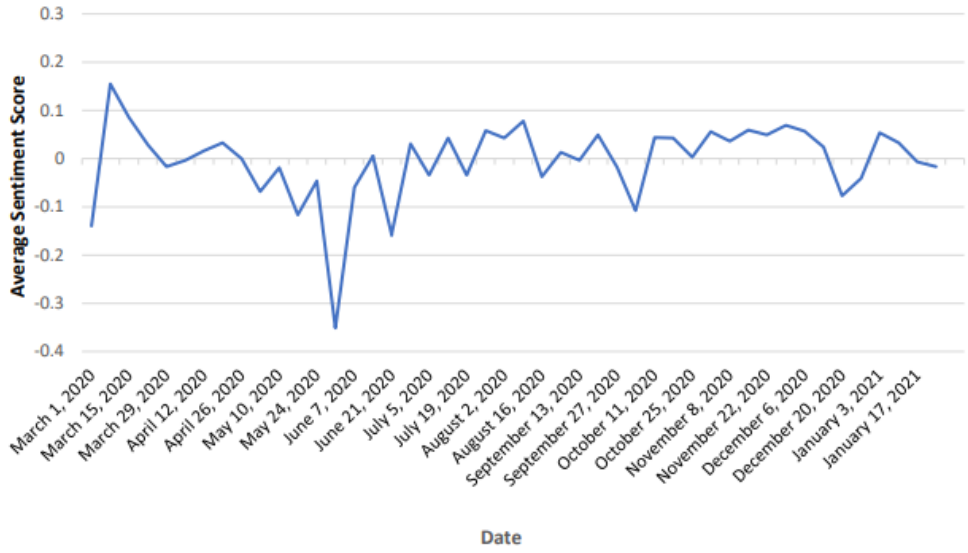

***Figure 3**: Average sentiment score of tweets mentioning COVID-19 vaccines each week using the VADER algorithm*(Xie et al., 2021).

The VADER algorithm calculated that the sentiment of the 4,438,203 tweets in the US that mentioned the COVID-19 vaccine was 34.78% (1,543,397) positive, 32.40% (1,438,107) negative, and 32.82% (1,456,699) neutral. There were two significant peaks. The COVID-19 vaccine-related tweets got the highest average sentiment score (0.155) from March 8, 2020, to March 14, 2020, and there were 54,364 tweets from the US during this time. With 58,903 tweets from the US over this period, the tweets from May 31, 2020, to June 6, 2020, received the lowest average sentiment score, which was -0.351(Xie et al., 2021).

Another algorithm that has been used over the years is the National Research Council Emotion Lexicon (NRCLex). The NRC Emotion Lexicon, otherwise called the "NRC Word-Emotion Association Lexicon," is a rundown of English words and their relationship with eight fundamental feelings (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two opinions (negative and positive). The National Research Council of Canada created the lexicon, which is available for research as well as commercial use. It was made by publicly supporting explanations from individuals all over the planet, who were approached to rate how they had an outlook on each word on a scale from 1 (not the slightest bit) to 9 (extremely)(*NRC Emotion Lexicon*, n.d.). For vaccine perception, four pairs of primary bipolar emotions are examined by NRCLex: happiness as opposed to sadness as a feeling; fear (feeling afraid) versus anger (feeling angry); trust (more grounded esteem and more fragile acknowledgement) versus disgust (feeling something is wrong or nasty); and the distinction between anticipation (looking forward to something positive) and surprise (being unprepared for something)(Hu et al., 2021).

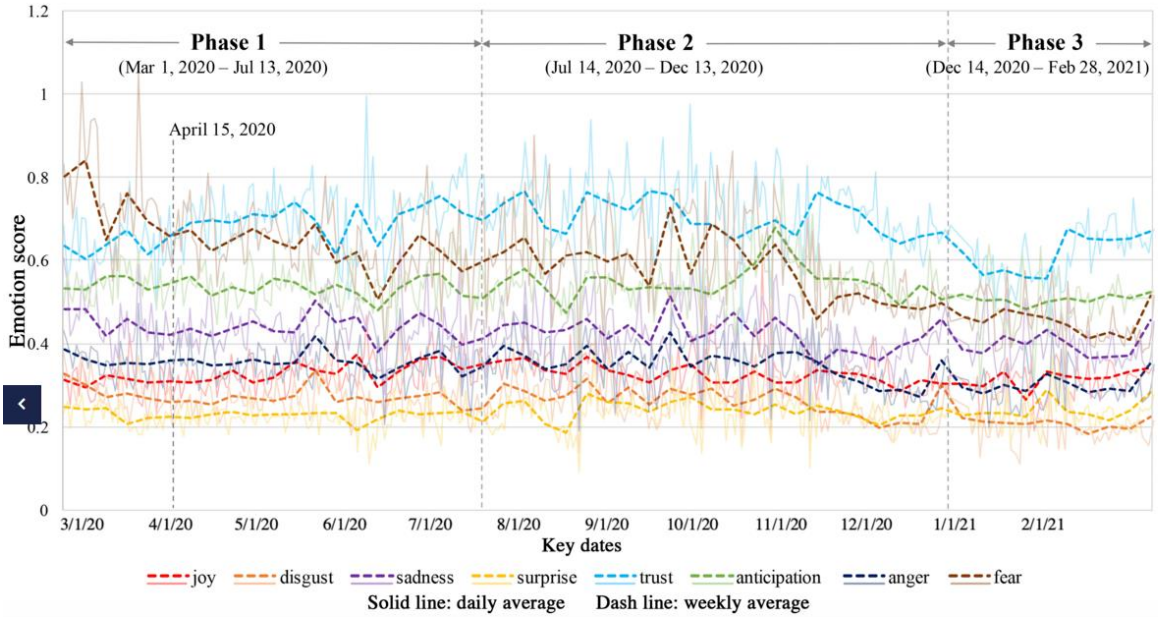

***Figure 4**: Average daily and weekly emotion scores over the entire study timeline at the national level using NRCLex*(Hu et al., 2021).

Figure 4 depicts the temporal patterns of the eight different types of emotions, including fear, joy, disgust, surprise, and anticipation. The vertical comparison of the weekly average trend lines revealed that trust (blue dashed line) had the highest weekly average ratings throughout the most of the timeline, followed by fear, anticipation, sadness, anger, joy, disgust, and surprise (dashed lines). It is important to note that before mid-April 2020, fear's weekly average emotion score was greater than trust's, presumably as a result of the early pandemic's rapid COVID-19 infection and inefficient viral spread management. These occurrences could have made people feel anxious, fearful, or even panicky. The public's trust in and expectation of vaccination were accompanied by a mixture of fear, grief, and anger, according to the general trend in emotion scores, even though variations in emotion scores (e.g., local peaks and troughs) could be seen within each of the 8 emotions(Hu et al., 2021).

Topic Modelling Analysis

Topic modelling is a text-mining technique that automatically derives the primary topics from a sizable corpus of documents. It can be used to reveal the underlying pattern in a group of texts and to pinpoint the primary themes that are covered. Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis are two techniques that can be used for topic modelling (LSA). LDA is a model that assumes that each document is made up of a given number of topics, and each word is chosen from one of those topics. In contrast, LSA is a method based on linear algebra that visualizes words and documents as vectors in a high-dimensional space and finds patterns and connections between them(Topic Modelling, n.d.). Preprocessing the text data is the initial stage in topic modelling. Stop words may need to be eliminated, words may need to be stemmed or lemmatized, and a document-term matrix may need to be built. The document-term matrix is a matrix where each row is a document and each column denotes a word. The entries in the matrix correspond to the frequency of each word in the document(Topic Modelling, n.d.).

The focus will be on the Latent Dirichlet Allocation (LDA) as it was majorly adopted for topic modelling. LDA is a probabilistic approach that can be used to identify the underlying subjects of a

group of texts. It is a generative model in which each word in the text is taken from one of a predefined number of themes, and each document is assumed to be a blend of those topics. LDA works on the fundamental premise that each document is composed of a variety of themes, each of which is defined by a distribution of words. The underlying subjects in a document are considered latent (unobserved) variables in Latent Dirichlet Allocation (LDA), which presupposes that the words in the document are produced by sampling from these latent topics(Blei et al., 2003).

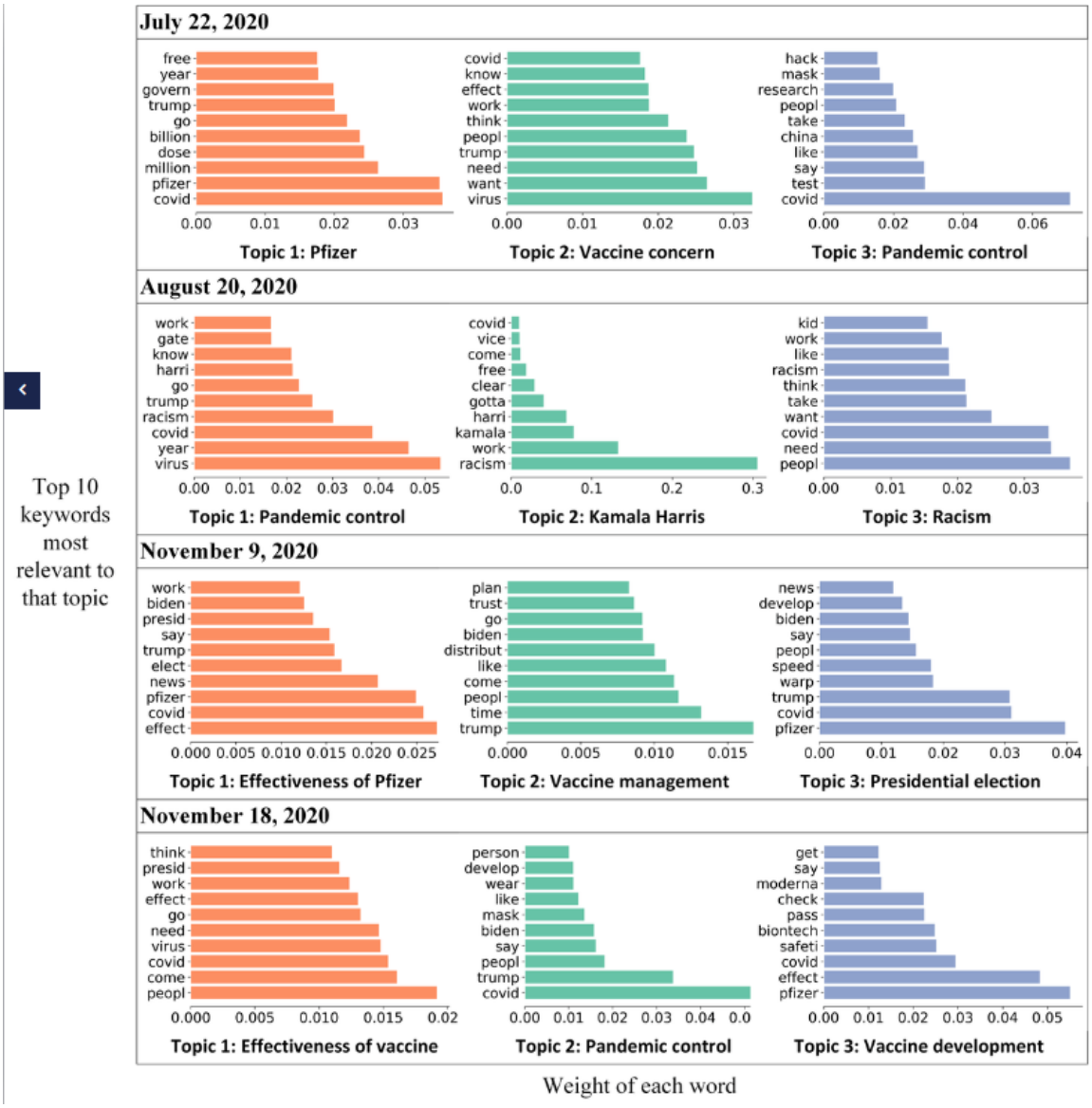

***Figure 5**: Three topics discussed on each of 4 key dates: July 22, 2020; August 20, 2020; November 9, 2020; and November 18, 2020, using LDA*(Hu et al., 2021).

On July 22, 2020, when Pfizer and the US government's cooperation hastened the manufacture and delivery of 100 million doses of COVID-19 vaccines, another significant increase in sentiment score was seen. Through topic modelling, the terms "pfizer," "govern," and "million" were commonly discussed as shown in Figure 5. The sentiment score significantly decreased on August 20, 2020, following Kamala Harris' official acceptance of the Democratic Party's vice presidential candidature at the 2020 Democratic National Convention. In reference to the backdrop of the anti-racism demonstrations for George Floyd and Breonna Taylor, Harris said that "there is no vaccine for racism". Through topic modelling, the terms "racist" and "kamala" were identified. When Pfizer

reported that its vaccine is 90% effective on November 9, 2020, the sentiment score also increased(Hu et al., 2021).

There is still the implementation of word cloud maps to visualize and summarize words. Based on frequency and significance, word cloud maps are used to emphasize well-known words and phrases. They give you immediate, straightforward visual insights that can inspire more thorough analysis(What Are Word Clouds? - Boost Labs, n.d.).

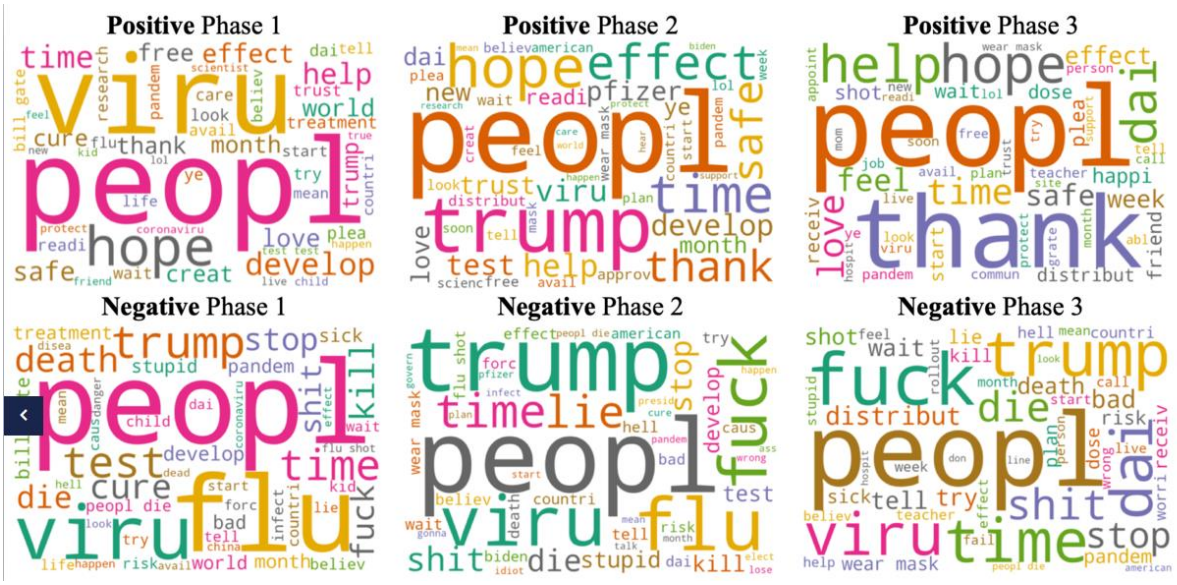

Figure 6*: Popular keywords associated with positive and negative sentiments over 3 phases using Word Cloud Map*(Hu et al., 2021).

50 popular words connected with both positive and negative attitudes were mapped into word clouds during the course of the three phases as shown in Figure 6. A word's magnitude indicates how widely used it is and how frequently it appears in tweets. Popular terms with positive connotations included "hope", "help", "thank", "love", "safe", "cure", and "free", though "peopl", which has a more neutral connotation, seemed to be the most well-liked. "Hope", "safe", and "thank" increased in frequency over the course of the three stages; "thank" in particular became the most widely used word in Phase 3. On the other hand, negative terms like "flu", "death", "trump", "fuck", "lie", "die", "kill", "shit", and "stupid" were frequently used. In contrast, "die," "fuck," "shit," and "trump" evolved to be larger from Phase 1 to Phase 3 during the course of the three phases; in particular, "trump" became predominate in Phase 2 perhaps as a result of Trump's rising popularity brought on by the 2020 Presidential Election(Hu et al., 2021).

DISCUSSION

On the above-reviewed algorithms or applications, I will be highlighting the strengths and weaknesses. Focusing on Sentiment analysis, it has been applied to various areas over time like social media monitoring, automobile/vehicle-based services, product/service review, travel/tourism-based services, healthcare-based services, and financial services. Apart from the already existing applications listed, there are areas associated with books, electronics, entertainment, fashion, online services, Malls and stores, etc, that sentiment analysis can still be applied(Nasreen Taj & Girisha, 2021). In many Natural Language Processing (NLP) tasks, deep learning models, such as neural networks, have produced state-of-the-art results, and, likely, they will also be used for sentiment

analysis in the future. Sentiment analysis models will probably advance and incorporate the context of a statement's utterance. They will be able to interpret statements more effectively and make more accurate predictions as a result. To provide a better understanding of text data, sentiment analysis will probably be combined with other NLP tasks, like entity recognition and summarization. It is expected that sentiment analysis will be utilized more frequently in a variety of industries as it gets more precise and accessible, enabling firms to better understand customer sentiment and make more informed decisions. The field of sentiment analysis has a bright future ahead of it as machine learning and natural language processing techniques continue to evolve, allowing for the development of more precise models that can better interpret the sentiment indicated in text data.

As for topic modelling analysis, one of the challenges is that there is usually a problem of data sparsity because it is difficult to carry out topic detection and sentiment extraction from short texts as it lacks contextual information. Other challenges in language interpretation include the use of sarcasm, imagery, metaphors, similes, humour, and figurative language, which depends on background knowledge and/or context and influences the accuracy of subject and sentiment classification(8 Limitations of TMA - LAZARINA STOY, n.d.). Topic modelling is likely to remain a crucial method for examining and comprehending enormous volumes of text data in the future. The future of topic modelling analysis is expected to be shaped by several factors. One tendency is the expansion of massive datasets' accessibility and the improvement of computer resources, which will allow for the creation of more advanced and precise subject-modelling methods. The importance of interpretability and explainability in machine learning is also rising, which will encourage the creation of subject-modelling algorithms that can produce outcomes that are clearer and easier to comprehend. The increasing use of deep learning techniques, which have the potential to dramatically increase the accuracy and efficiency of topic modelling algorithms, is another development that is likely to have an impact on topic modelling in the future. A topic modelling technique that can handle many languages and facilitate cross-lingual comparison and analysis will be developed in response to the growing popularity of multilingual and cross-lingual analysis. Last but not the least, the increased interest in multilingual and cross-lingual analysis will probably spur the creation of word clouds that facilitate cross-lingual comparison and analysis and can handle several languages.

CONCLUSION

This work has discussed the essential use of Artificial Intelligence and other applications during COVID-19 to understand the public perception of the COVID-19 vaccine. This research concept was dated about one to two years ago, yet there have been some recent advancements in this field that results in the evolution of new sentiment and topic modelling analysis methodologies. This work contributes to the review of the impact of Data science in the age of COVID-19 by discussing the applications in the collection of data, the applications and methods used over the years for sentiment and emotion analysis, the methods used over the years for topic modelling and identification of frequently used words. Also discussing some of the lags in these applications or methods along with future works. Overall, the development of more potent computational resources, the growing relevance of interpretability and explainability in machine learning, and the expanding availability of huge datasets are anticipated to be the driving forces behind the future of AI in sentiment analysis and topic modelling.

REFERENCES

8 Limitations of Topic Modelling Algorithms on Short Text - LAZARINA STOY. (n.d.). Retrieved December 21, 2022, from https://lazarinastoy.com/topic-modelling-limitations-short-text/

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. In *Journal of Machine Learning Research* (Vol. 3).

Cambria, E., & Hussain, A. (2015). *Sentic Computing*. Springer International Publishing. https://doi.org/10.1007/978-3-319-23654-4

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. https://doi.org/10.1109/MIS.2013.30

COVID-19 - Wikipedia. (n.d.). Retrieved December 20, 2022, from https://en.wikipedia.org/wiki/COVID-19

Dr. Mersad Alimoradi. (2022). *COMPARISON OF COVID 19 VACCINES*. https://myacare.com/blog/comparison-of-covid-19-vaccines

Gina Mueller. (n.d.). *Why Twitter is the Ideal Platform for Engagement*. Retrieved December 16, 2022, from https://www.convinceandconvert.com/social-media/twitter-engagement/#:~:text=Ultimately%2C%20Twitter%20is%20a%20network,BEST%20marketing%20education%2C%20totally%20free.

Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., She, B., & Li, Z. (2021). Revealing public opinion towards covid-19 vaccines with twitter data in the united states: Spatiotemporal perspective. In *Journal of Medical Internet Research* (Vol. 23, Issue 9). JMIR Publications Inc. https://doi.org/10.2196/30854

JustAnotherArchivist/snscrape: A social networking service scraper in Python. (n.d.). Retrieved December 20, 2022, from https://github.com/JustAnotherArchivist/snscrape

Kaye, A. D., Okeagu, C. N., Pham, A. D., Silva, R. A., Hurley, J. J., Arron, B. L., Sarfraz, N., Lee, H. N., Ghali, G. E., Gamble, J. W., Liu, H., Urman, R. D., & Cornett, E. M. (2021). Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives. In *Best Practice and Research: Clinical Anaesthesiology* (Vol. 35, Issue 3, pp. 293–306). Bailliere Tindall Ltd. https://doi.org/10.1016/j.bpa.2020.11.009

Nasreen Taj, M. B., & Girisha, G. S. (2021). Insights of strength and weakness of evolving methodologies of sentiment analysis. *Global Transitions Proceedings*, *2*(2), 157–162. https://doi.org/10.1016/j.gltp.2021.08.059

NRC Emotion Lexicon. (n.d.). Retrieved December 20, 2022, from https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Python | Sentiment Analysis using VADER - GeeksforGeeks. (n.d.). Retrieved December 20, 2022, from https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/

Saladino, V., Algeri, D., & Auriemma, V. (2020). The Psychological and Social Impact of Covid-19: New Perspectives of Well-Being. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.577684

Sentiment analysis - Wikipedia. (n.d.). Retrieved December 20, 2022, from https://en.wikipedia.org/wiki/Sentiment_analysis

Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the Affective Norms for English Words by Discrete Emotional Categories. *Behavior Research Methods*, *39*(4), 1020–1024. https://doi.org/10.3758/bf03192999

Topic Modeling: An Introduction. (n.d.). Retrieved December 21, 2022, from https://monkeylearn.com/blog/introduction-to-topic-modeling/

Twitter API Documentation | Docs | Twitter Developer Platform. (n.d.). Retrieved December 20, 2022, from https://developer.twitter.com/en/docs/twitter-api

VADER (Valence Aware Dictionary and sentiment Reasoner) Sentiment Analysis | by Swayanshu Shanti Pragnya | MLearning.ai | Medium. (n.d.). Retrieved December 20, 2022, from https://medium.com/mlearning-ai/vader-valence-aware-dictionary-and-sentiment-reasoner-sentiment-analysis-28251536698

What are Word Clouds? The Value of Simple Visualizations | Boost Labs. (n.d.). Retrieved December 21, 2022, from https://boostlabs.com/blog/what-are-word-clouds-value-simple-visualizations/

WHO. (n.d.). *Side Effects of COVID-19 Vaccines*. Retrieved December 16, 2022, from https://www.who.int/news-room/feature-stories/detail/side-effects-of-covid-19-vaccines

Xie, Z., Wang, X., Jiang, Y., Chen, Y., Huang, S., Ma, H., Anand, A., & Li, D. (2021). Public Perception of COVID-19 Vaccines on Twitter in the United States. *MedRxiv : The Preprint Server for Health Sciences*. https://doi.org/10.1101/2021.10.16.21265097

Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S. (2021). An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, *108*, 256–262. https://doi.org/10.1016/j.ijid.2021.05.059