EPiC Series in Education Science

Volume 5, 2023, Pages 54–68

Proceedings of NEMISA Digital Skills Conference 2023: Scaling Data Skills For Multidisciplinary Impact

Sentiment Analysis of South African tweets about COVID-19 vaccines

Thato Ditsele[1], Wallace Chigona[1] and Malebo Sephodi[1]

[1] University of Cape Town, Cape Town, South Africa

dtstha004@myuct.ac.za,

wallace.chigona@uct.ac.za,gllmal003@myuct.co.za

Abstract

The COVID-19 pandemic and lockdowns have led to a surge in the use of social media for information sharing and learning about the virus and the vaccine. Our study aimed to understand the sentiments of South African Twitter users towards COVID-19 vaccines. Using the Twitter API version 2, we collected a total of 21,084 tweets from 1 January 2021 to 31 December 2021, during the government's rollout of the vaccine. A sentiment analysis was performed using the VADER lexicon-based classifier to categorise the tweets into positive, neutral, and negative sentiments. The results showed that 40% of the tweets were positive towards the vaccine, 32% were neutral, and 28% were negative. The analysis also revealed that people expressed their opinions on vaccinations more frequently during the early months of the year (January-March 2021) in response to the government's announcement for the vaccine rollout. However, the attitudes towards the vaccine changed throughout the year, indicating that people were sceptical of the government's vaccine rollout strategy, which could have affected the overall vaccine adoption. The findings of this study can provide valuable insights for policymakers and healthcare organisations in shaping effective strategies for promoting vaccine adoption.

Keywords— Sentiment analysis, social media, vaccine hesitancy South Africa 2021, Twitter-Sentiment analysis, public opinions, Covid-19 vaccines

1 Introduction

The global impact of the Corona virus disease (COVID-19) since its emergence in China in December 2019 has been significant, leading to the widespread spread of the virus. South Africa reported its first COVID-19 case in February 2020 and has since reported over 3.7 million confirmed cases, with approximately 103,000 deaths (Department of Health, 2022). In response, the South African government employed a 5-level alert system to handle the gradual relaxation of the lockdown, aimed at controlling the spread of the virus (Department of Health, 2022).

Vaccination has been widely recognised as a critical tool in preventing the spread of COVID-19, with human studies indicating that COVID-19 vaccines are 89% safe, with a low to moderate risk of serious disease among older individuals (Bam, 2021). Despite this, there has been growing vaccine scepticism in South Africa, with only 67% of those polled expressing certainty in getting vaccinated and the remaining 33% uncertain or against getting vaccinated (Bam, 2021).

The South African government's failure to prioritise public data in the early days of the COVID-19 outbreak was identified as a mistake that impacted policymakers' ability to deal with the pandemic effectively, leading to a loss of public trust in the government's ability to manage the vaccine rollout (Marivate & Combrink, 2020; Puri et al., 2020). The increasing concern about the impact of social media in increasing anti-vaccination sentiment and low vaccine uptake is a major concern among public health professionals (Rotolo et al., 2022).

Twitter data provides real-time insights into public opinion on a variety of pandemic-related issues, including vaccinations, and has been used to determine South African sentiments about COVID-19 vaccines throughout 2021. Our study aimed to answer the following research questions:

1. What are the sentiments expressed in South African tweets about COVID-19 vaccines?
2. What are the central themes discussed on Twitter about COVID-19 vaccines in South Africa?
3. How did South Africa's Twitter sentiments on COVID-19 vaccines change throughout 2021?

Our study used a methodology that involves opinion mining and sentiment analysis, and its results can be applied in multidisciplinary investigations involving social media data analysis.

2 Literature Review

2.1 Twitter social media platform

Twitter is a real-time microblogging platform that allows individuals or groups to share their opinions and perspectives on various topics through multimedia, text, and other forms of media. The messages shared on Twitter are known as "tweets" and are displayed on a timeline, a collection of tweets in chronological order (Pawar et al., 2015; Shoaei & Dastani, 2020). The availability of data through the Twitter API v2 enables researchers to examine Twitter data in depth, offering valuable insights into public opinions and sentiments (Mohamed Ridhwan & Hargreaves, 2021).

Twitter is a popular source of health care information in South Africa, with around 8 million users among social media networks (van Heerden & Young, 2020; Hoffman et al., 2021). During the COVID-19 pandemic, a tweet about the virus was reported every 45 milliseconds on Twitter, and the hashtag #coronavirus was one of the most frequently used globally in 2020 (Puri et al., 2020). The legitimacy of a Twitter account is determined by whether it is verified or unverified. Verified accounts are those of genuine public interest, while the legitimacy of unverified accounts is unknown (Mir et al., 2022).

2.2 Use of twitter to determine public sentiments on Covid-19 vaccines

Twitter has become an important source of information for individuals seeking information about health care, including information about vaccines (Hoffman et al., 2021). In the context of the COVID-19 pandemic, Twitter has been used to track public sentiment and reactions to vaccine rollout (Hung et al., 2020). Twitter users engage with one another through direct messaging, updates, replies, likes, and retweets, which can indicate user engagement and reveal their attitudes towards (Kang et al., 2017).

Studies have used Twitter data to gain a better understanding of how misinformation about vaccines, including vaccine-induced acquired immunodeficiency syndrome (VAID), has spread on the platform (Shahi et al., 2020a). This misinformation has led to confusion and misperceptions about vaccines, and researchers have used sentiment analysis to track these attitudes and reactions (Maciuszek et al., 2021; Mønsted & Lehmann, 2022).

Twitter data can be collected and analysed through open academic access via API v2, making it a valuable tool for researchers, government organizations, and health care professionals (Luo et al., 2021). A number of studies have analysed tweets using techniques such as logistic regression, deep learning algorithms, topic modelling, Latent Dirichlet Allocation (LDA), and machine learning to better understand the conversations, concerns, sentiments, and reactions elicited by tweets (Li Han Wong et al.,2022; Amjad et al. 2021; Deverna et al. 2021; Anuratha et al., 2021; Alfatease et al., 2021; Piltch-Loeb et al., 2021).

In conclusion, Twitter plays a crucial role in shaping public perception and understanding of health care, including vaccines. By analysing Twitter data, researchers and health care professionals can gain a deeper understanding of public sentiment and reactions, and use this information to inform vaccine policy and communication efforts. Vaccine manufacturers can also use sentiment analysis to quickly detect negative feelings towards their vaccine and respond accordingly (Anuratha et al., 2021; Hou et al., 2021).

2.3 Challenges to vaccine acceptance

Misinformation or disinformation, refers to misleading information shared intentionally rather than when erroneous information is transmitted accidentally. Misinformation on the internet has also been linked to reduced vaccination uptake. Other studies noted that misinformation was spread, stating that people are testing HIV-positive after vaccination just like COVID-19. Vaccination sceptics continued to dominate the Twitter dialogue, with information centred on government scepticism of boosters and requirements, intellectual property around the vaccine, and who benefits from the vaccine (Department of Health, 2022). Furthermore, misinformation can erode trust in science and public health authorities, leading to a drop-in vaccine uptake (Steffens et al., 2019; Bonnevie et al., 2020b).

Whereas unpleasant effects (31%) and concern about their effectiveness (21%) were common, loss of trust (18%) and anti-vaccination feelings were uncommon (14%) (Burger et al.,2021). Fears about mandated vaccinations appearing on social media can cause fear and scepticism about vaccines, according to the Africa CDC (2021). Conspiracy theories and religious ideas published on social media are two common types of misinformation that influence user perception (Bam, 2021; Chou et al., 2021; Shahi et al., 2020a).

Vaccine hesitation / hesitancy describes the refusal or postponing of vaccination despite its availability (Chadwick et al., 2021; Hoffman et al., 2021; Wilson & Wiysonge, 2020). South Africa's COVID-19 vaccine hesitancy can be addressed through collective efforts to ensure that there are no constraints related to the supply and system that administers vaccines. The data creates a picture of the vaccine debate on Twitter as highly polarised, with individuals who express similar views on vaccinations more likely to communicate with one another and post content from comparable sources. In other studies, developed economies such as Singapore rely heavily on their government to provide accurate information and demonstrate trust in science and their vaccine program (Mohamed Ridhwan & Hargreaves, 2021).

Pro- and anti-vaccine information may spontaneously separate into distinct groups, possibly as a result of social media self-selection that brings together communities with similar viewpoints (Puri et al., 2020). Vaccination opponents and proponents can be classified as social groups rather than by more specific criteria such as gender, nationality, or affiliation. People in a variety of countries voiced scepticism COVID-19 vaccines, as well as serious worries about vaccine safety and distrust in governments and doctors (Hou et al., 2021; Puri et al., 2020). Some researchers use sentiment analysis scores to infer that positive sentiment is pro-vaccine and negative sentiment is anti-vaccination (Lyu et al., 2022; Maciuszek et al., 2021b).

3 Methodology

Twitter sentiment analysis is the process of analysing and categorising opinions expressed in Twitter posts (tweets) into positive, negative, or neutral classes. The goal is to determine the overall public opinion or emotions towards a specific topic or brand (Alsayed, Alharthi, & Adeyemo, 2019). As illustrated in Figure 1, the study went through five phases (Mohamed Ridhwan & Hargreaves, 2021).

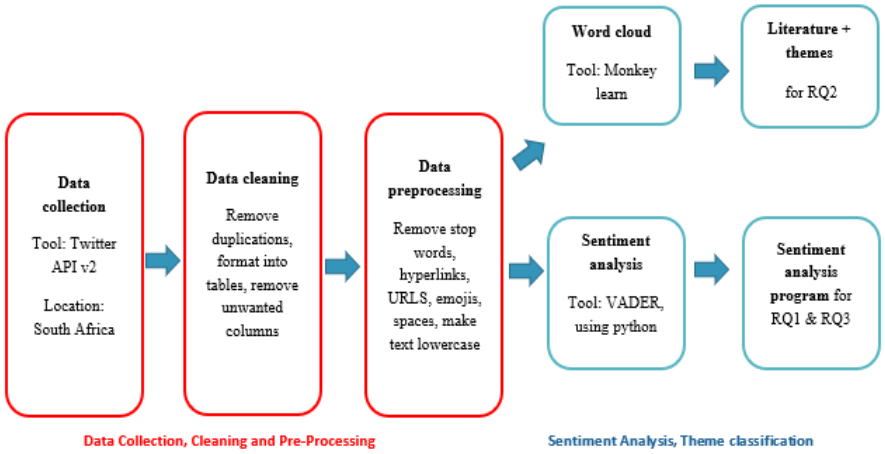


Figure 1: Five phases used in the study

3.1 Phase 1: Data collection

We utilised the Twitter historical tweets downloader, after obtaining developer access to Twitter version 2. Our focus was to search and download historical tweets from South Africa. The keywords selected from trending topics were related to vaccines, Covid-19, and anti-vaccination and were selected based on their relevance and popularity (Li Han Wong et al., 2022). The keywords included:

Vaccineforall, Vaccine, Antivaccine, Vaccinationcovid, Covid19, Vaccine, Shot, covid-19, Coronavirus, AstraZeneca, modernavax, anti-vaccination, anti-vax, anti-vaxxers, pro-vax, covid19jab, vaccinesideeffects, Antivax, Antivaxxer, Antivaxxers, vaccine, Moderna, pfizersideffects, j&jsideaffects, vaccinesideeffects, novaxx

The date range for this study was from January 1st, 2021 to December 31st, 2021.The metadata extracted from the tweets included: tweet content, location, date and time of post, author, language, type of account, account verification status, author friends, retweet count, and likes count. This information was crucial in analysing the tweets and understanding the perspectives of the authors on the topics related to vaccines and Covid-19.

3.2 Phase 2: Data cleaning

The Twitter dataset includes South Africans' location fields, as well as race, gender, age, and languages. We used the keyword 'vacc,' which is a prefix for words like vaccine, vaccinated, vaccinate, and vaccination, to narrow down the results. The tweets were filtered to exclude non-English tweets. Some tweets contain English and other languages but categorised as English by Twitter, which we could not remove entirely. We eliminated 43 656 tweets, leaving 21 084 tweets suitable for data analysis.

3.3 Phase 3: Data selection

The final dataset consisted of 21 084 tweets. We used Python's Panda package to transform the data into a data frame for the study. We used the Natural Language Toolkit (NLTK) to perform tokenization, sentiment analysis (using the NLTK Sentiment Intensity Analyzer), and emotive analysis on the text.

Tokenisation is the process of dividing large text blocks into smaller chunks. In this case, tokens could be letters, numbers, or sub words (Ioana-Andreea et al., 2021). To tokenize, the Word Tokenize package was imported into Python Pandas. The analysis of word sequences aids in understanding the meaning of the text (Mir et al., 2022). All punctuation was removed during standardisation (Kumar & Sebastian, 2012). The tweet id and user id were anonymised.

3.4 Phase 4: Sentiment analysis of tweets

VADER is a pre-trained model that employs rule-based values tailored to social media sentiments. It employs dictionary phrases with evaluation capabilities to classify tweets based on the compound score. The compound score was used to decide whether a tweet may be labelled (good, negative, or neutral). A compound score of ≥ 0.05 satisfies the labels positive, and between -0.05 and 0.05 was neutral and a score of ≤ -0.05 was negative (Mir & Sevukan, 2022).We used the Natural Language Toolkit (NTLK) package in the first step of sentiment analysis to tokenize the content data and remove commonly used words. Based on the tokens, a frequency dictionary was created. The Python script then generated a frequency list in descending order. Stop words are frequently used words such as "drink," "please," "guys," or "covid," "corona" are among the top keywords. These words, along with the rest of the NTLK stop words list, were removed to gain insight from the analysis (see Appendix A).

3.5 Phase 5: Word frequency and themes

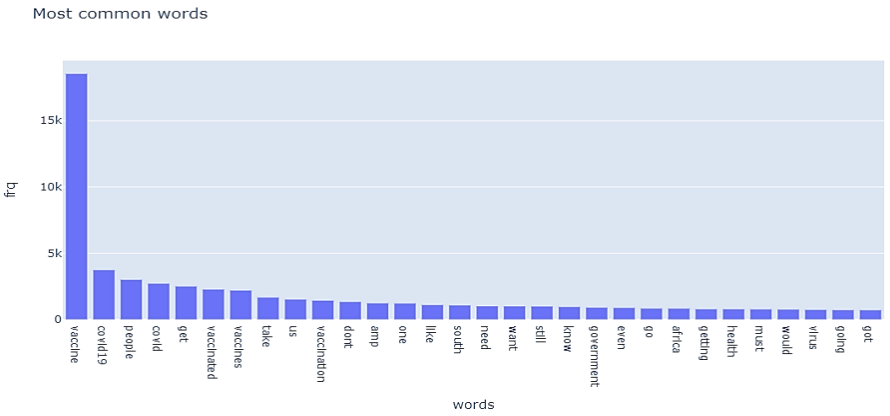
The data cleansing resulted in 21 084 tweets in English. Using the Python df.sample random sample formula (n=200, random state=1), we obtained a random sample of 200 tweets. Previous research was divided into two categories: pro-vaccination and anti-vaccination. We added a third type of vaccination hesitancy (Maciuszek et al., 2021b; Monsted & Lehmann, 2022b). The central themes were determined through literature and word cloud topics. The word cloud was created using the Monkeylearn web API.

The choice of a sample size of 200 tweets was because of time and resource constraints, and the need for a representative sample that accurately captures the overall themes of the population.

4 Findings and discussion

4.1 Sentiments about COVID-19 vaccines

An overall sentiment analysis using the VADER lexicon-based method and was performed on 21 084 tweets. Figure 2 shows the most frequent words generated from the dataset without stop words to find the 30 most frequently used words in the tweets. The top three keywords used by Twitter users were “vaccine,” “covid19”, and “people”. This was expected given that the filter list used to filter the tweets was specifically looking for vaccine discussion; this validates our Twitter data. Figure 3 shows the distribution of sentiments in percentages. The graph depicts the overall sentiment distribution of tweets from 1st January 2021 to 31st December 2021. Most tweets were positive 41% (8543), seconded by neutral tweets 32% (5894). Finally, negative tweets account for only 28% of all tweets (6647).



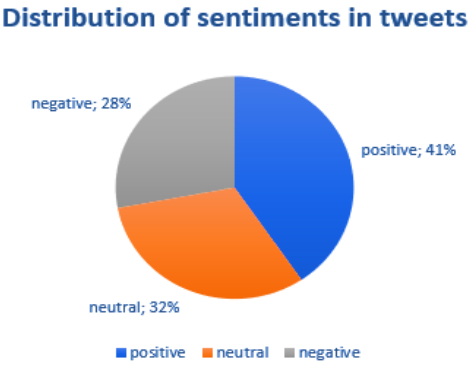


Figure 3: The distribution of sentiments in tweets

Table 1 shows a selection of tweets and their sentiment classifications. We categorised the hesitant tweets with neutral tweets.

Positive tweets (n = 8543)	**Neutral tweets (n = 5894)**	**Negative tweets (n = 6647)**
"Please get sputnik V vaccine is the only I will take"	*"Why are we not educating people on the vaccine the same way people who advocate for Prep etc"*	*"IM NOT TAKING THAT VACCINE!!!!"*
"I wanna be getting my vaccine too"	*"Why is the vaccine roll out so low in Northern Cape?"*	*"Say No to the vaccine"*
"Finally took my vaccine today. No pain, no effect whatsoever"	*"Why should people be forced to take the Covid-19 vaccine"*	*"I have no interest in all your vaccine lies! God sees you! Petition Against Mandatory Vaccinations in South Africa"*

Table 1: Examples of tweets for the different sentiments

4.2 Central themes discussed on Twitter about COVID-19 vaccines in South Africa

Figure 4 depicts a word cloud of the diversity of multilingual hashtags used to discuss COVID-19 vaccinations in a Twitter dataset. The larger the size of the word in the cloud, the more frequently it occurs in the corpus (Aguilar-Gallegos et al., 2020; Kang et al., 2017; Mir & Sevukan, 2022). The word cloud shows that the most common words were 'vaccine', 'government', 'covid19', 'country', 'vaccine information'. Additional topics that were in the corpus were: "country has no confidence", "puppet master bill gates", "south african vaccine rollout strategy", "severe symptom" amongst topics relating to our central themes.

A total of 183 vaccine-related tweets were analysed and 9 themes were manually coded based on previous literature and the word cloud, as shown in Figure 5. These themes include misinformation, vaccine hesitancy, government, conspiracy, religion, pro-vaccination (pro-vaxx) and anti-vaxx vaccination (anti-vaccination) and humour. We identified the themes by comparing the words from the word cloud to what previous studies had found.

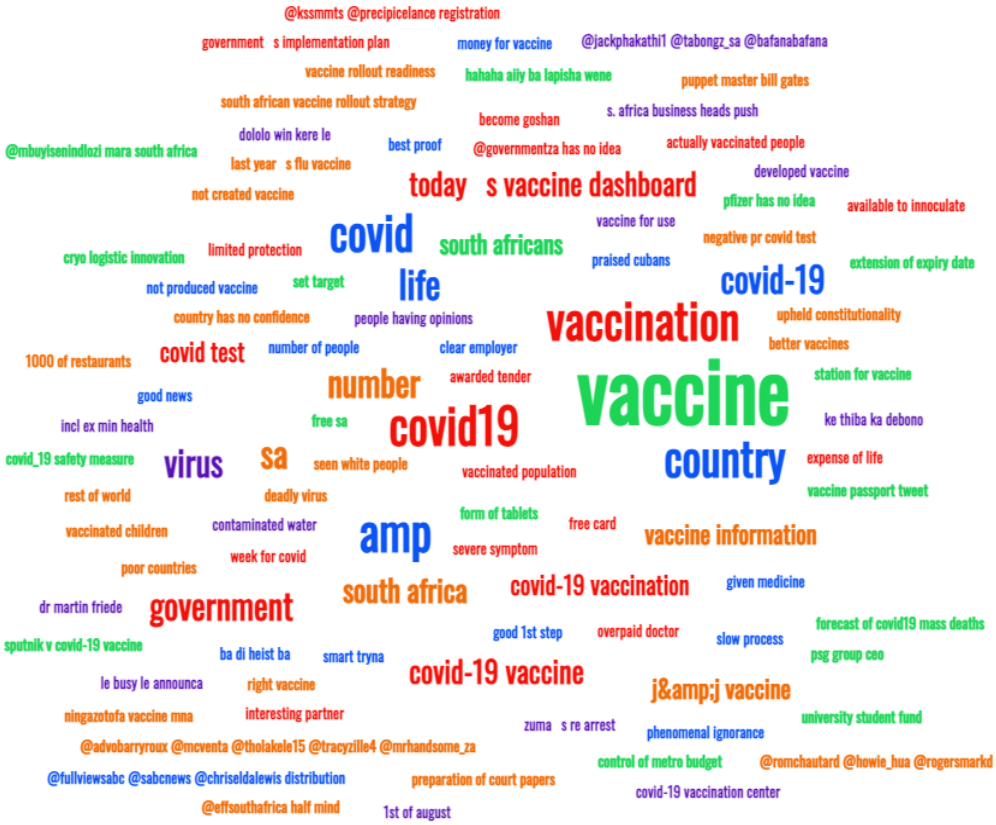


Figure 4: Word cloud of Top 100 words amongst tweets about COVID-19 vaccines

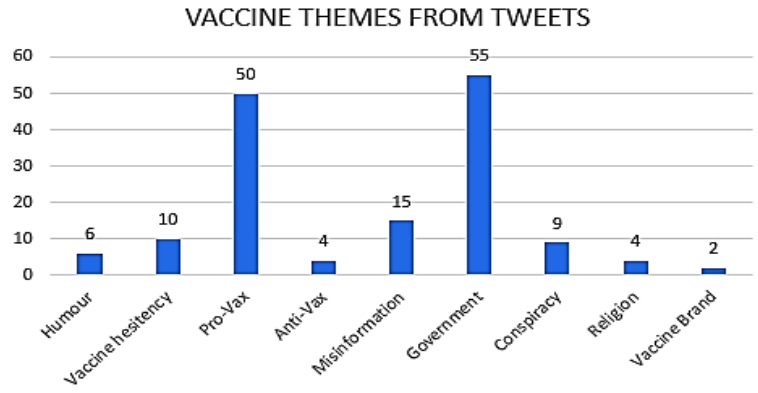


Figure 5: Vaccine themes that emerged from tweets

4.2.1 Government

"Government" was the most common theme (28%, n=55) in vaccine-related discussions. This contrasts with a previous study in Singapore where the main focus was Singapore's government initiatives (Mohamed Ridhwan & Hargreaves, 2021). One of the factors leading to vaccine hesitancy is the possible a link with political issues and disillusionment.

4.2.2 South African governments vaccine rollout plan

Our findings indicate that people had little trust in the South African government's handling of the vaccine rollout. Many people believed that the country can and should produce its own vaccines, but was unwilling to do so (Aguilar-Gallegos et al., 2020; Kang et al., 2017; Mir & Sevukan, 2022).

"@FloydShivambu Cde Floyd, Africa has resources to produce own vaccines but we don't have leaders. Let's stop blaming the West for our own ineptitude. The least we could have done was to fund research into vaccine discovery. We deserve to be sidelined, West taxpayers can't fund our us forever".

Many tweets in the corpus came from verified accounts. This resulted in higher tweet engagement than non-verified tweets. The hashtag #VaccineRolloutSA was more widely used to inform the public about the plan to administer vaccines in order to achieve the anticipated herd immunity.

"The first batch of the J&J #COVID19 vaccine arrived at @ortambo_int last night. Frontline healthcare workers will be vaccinated during first phase of the #VaccineRolloutSA"

Government corruption leads to mistrust of authorities: Some people lost confidence in the government's involvement in the rollout of vaccines. This is supported by evidence of tender fraud involving COVID-19 personal protective equipment (PPE) by senior government officials (Mpofu, 2021, p.52). According to reports, corruption did not anticipate the use and allocation of COVID-19 funds (Mpofu, 2021, p.52).

"ANC has taught us that corruption is the way, we gonna buy those fake vaccine certificates or whatever y'all call it, anything for majwals ??"

"Medical aids don't pay a cent more than they have too why would they allow corruption in the vaccine roll out. @DrZweliMkhize @SundayTimesZA allow for market forces to deliver the drug just regulate the cost per jab."

The word corrupt appears in 401 tweets in the corpus. People were concerned about political corruption and theft, and they questioned the legitimacy of the vaccine rollout strategies. Since most previous studies came from the developing countries, Tweeter users did not complain as much about corruption.

4.2.3 Vaccine polarity

The second most common theme was "Pro-Vax" (25%, n=50), followed by "vaccine hesitancy" (5%, n=10) and "Anti-Vax" (2%, n=4). These themes were combined because they are the most frequently occurring themes in literature and our tweets, and they frequently overlap with positive, negative, and neutral sentiments. When the vaccine for COVID-19 becomes available, the tweets express a willingness to vaccinate against it. These tweets expressed gratitude for vaccines and how they are made available to save people's lives.

Pro-Vaccination: The Pfizer vaccine was given in two doses. Twitter users used hashtags like #IChooseVaccination to express their desire to get vaccinated and to encourage other Twitter users to do the same. These tweets demonstrate a willingness to receive the vaccine.

"Get your second dose to be fully vaccinated 😉 #IChooseVaccination #staysafe"

A global study comparing people's willingness to vaccinate noted that lower and middle-income countries like South Africa had a higher acceptance rate (Hou et al., 2021). This is similar to our findings; Figure 3 shows only 41% positive sentiments. The majority of South Africans, however, were eager to receive their first or second dose of the vaccine.

Anti-Vaccination: Anti-vaxxers are people who do not want to take or are opposed to vaccines (Hou et al., 2021). Some authors suggest that one of the things that distinguishes vaccine supporters from vaccine opponents is their attitude toward research and scientific knowledge (Maciuszek et al., 2021a).

The user stated that they were against mandatory vaccinations and cited religious texts to back up their position. There are numerous reasons why people are opposed to the vaccine, but two stand out: "freedom" and "choice."

"@RevMeshoe @CyrilRamaphosa He must not force us to be vaccinated, some of us our religion and beliefs doesn't allow us to take any vaccine"

"Would love to join a space discussing the science behind Vaxx and Anti-vaxx controversy. My submission would be around why it is scientifically absurd for people who have contacted covid, mild/severe to vaccinate"

The literature frequently mentions how anti-vaccine persons tend to deny scientific data demonstrating the safety and effectiveness of vaccinations and cast doubt on the authenticity of pertinent research, knowledge, and medical authority (Maciuszek et al., 2021a).

4.2.4 Vaccine Hesitancy

There were some comments about the uncertainty surrounding vaccine acceptance due to safety concerns. Some of these vaccine-related concerns include the claims that Johnson and Johnson vaccines were not thoroughly trailed/tested. Though people were concerned about receiving the vaccines, their concerns were not necessarily that the vaccines were unsafe, but that the vaccine development themselves was rushed.

"Do you honestly trust that the vaccine is safe?"

4.2.5 Vaccine brands safety and effectiveness

This theme is more neutral. It contains tweets that are for or against a specific vaccine brand. The vaccine brands were debated as people were choosing their preferred vaccine.

"Personally I want the pfizer. Vaccine Cocktail immunity"

"Me and my younger sister were vaccinated in May with Johnson & Johnson vaccine and didn't experience any side effects. My mom had a headache and was swollen where she was vaccinated"

"Is not because of Bill Gates' depopulation project... No one is infect advocating for Russian or China vaccine they all pushing for Bill Gates vaccine covering their faces with Russia and China!!"

4.2.6 Misinformation

The third most common theme was "Misinformation" (8%, n=15), "Conspiracy" (4.5%, n=200). This theme looked at tweets that were making misleading statements that had no evidence. Some of the contents of the tweets, included comments such as:

"I heard the Vaccine kills your entire family in one swoop, but I ain't no doctor".

"All the best with the J&J vaccine. I think it's a matter of probability & you might be the one to get the blood clot. As they say, "you might be history".

To manage false and dangerous media content and to convey reliable information intended to encourage the uptake of a potential COVID-19 vaccine to end the pandemic, researchers suggest media partnerships should be expanded (Niemiec, 2020).

4.2.7 Conspiracy theories

Accounted for 4.5% of the sample tweets. Bam (2021) associates these individuals with anti-scientific beliefs (believing in knowledge based on evidence). The corpus contained retweets with phrases like "they are killing people because they want to reduce population". There were a few tweets in the corpus that mentioned the depopulation agenda. These conspiracy theories were linked to one of the most frequently occurring words in Figure 4, "Bill Gates".

"Some countries are protesting against the vaccine, they claim Bill Gates is pushing population control publicly. Australia, Brazil and some EU countries are protesting against the vaccine."

There are several conspiracy theory subcategories in our dataset and in other literature, including "microchips in vaccines," "vaccines cause infertility," and "vaccines are for profit" (Nuzhath et al., 2021).

4.2.8 Religious beliefs

Religious opinions made up 2% of the tweets. This number may appear to be insignificant at first glance. However, famous religious leaders have a large Twitter following with high tweet engagement, which may sway their followers' opinions. Some users were leaning toward wanting to practice their own religious beliefs:

"If im entitled to right to a right to practice my religion, and for whatever i believe that the vaccine "generates from some form of Satanism" i see no reason why not to enforce my right to religion."

According to tweets like the one above, some people are unwilling to vaccinate for religious reasons. It also corresponds with literature indicating that social media users from all over the world hold similar religious beliefs about vaccines (Nuzhath et al., 2021; Wilson & Wiysonge, 2020).

4.2.9 Humour

Less than a percent of tweets were jokes. The use of humour in dealing with a serious matter could have been a form of coping mechanism (Mpofu, 2021). Example of jokes that were found were:

"Lol I saw one yesterday. These people had a sign saying "God is my vaccine" lol I laughed. Crazy yt people"

4.3 Changes of South Africa's Twitter sentiments on COVID-19 vaccines

Figure 6 depicts the average sentiments throughout the year. The frequency of each sentiment from 1st January 2021 to 31st December 2021 shows that positive tweets dominate. Positive, negative, and neutral tweets are closer together and overlap on specific days. The high spikes indicated that engagement was higher in the first three months of 2021, from January 2021 to March 2021. Figures 6 and 7 show an increase in Twitter engagement as the year began on the 3rd January 2021 (A), with 116 positive tweets, 88 negative tweets, and 73 neutral tweets. Minister Zweli Mkhize's public briefing statement revealed South Africa's COVID-19 vaccine strategy. On the first day, there was a high volume of tweet engagement because of an important event announced by an official with a verified account. Positive sentiments came from people who were optimistic about vaccination because of their personal experiences.

"*The lackluster manner in which ANC govt is rolling out vaccination processes is as if Covid 19 is yet to happen when in actual fact the pandemic has been with us for a year. It's like embarking on a long trip riding a donkey which has no sense of urgency inspite of eminent perils"*

This is understandable given that prominent individuals and media organisations frequently comment on and cover breaking news in real time using Twitter as a platform to amplify their messaging (Chen et al., 2020). The first substantial rise occurred on 11th January 2021 (B). On this day, South African President Cyril Ramaphosa declared during a national address that the goal of the parliament in designing the path to recovery was to develop a vaccination drive (Business Solutions, 2021). With 198 neutral tweets, 209 positive tweets and 154 negative tweets. The 1st February 2021 (C) had 361 counts of neutral tweets, 282 positive and 197 negative tweets. 8th February (D) had 229 tweets that were neutral, 228 positive, 208 negatives. After that period, there has been a significant drop in engagement then a slight rise in tweets on the 17th February there was a media statement that stated that the total number of COVID-19 cases in South Africa was 1 496 439, and since the last report, 2 320 more cases had been found ("Update on COVID-19 (17th February 2021)," 2021).

From April 2021 to August 2021, there was no significant event that caused a high spike in tweets in this period. The counts were averaging below 50 tweets per day. No major announcements were made by prominent public figures except on the 27th June when a government gazette was released on the adjusted level of lockdowns to level 3 (Gazette et al., 2021). The last significant spike for the year occurred on 20th August, 2021 (F). The National Institute for Communicable Diseases (NICD), a component of the National Health Laboratory Service, continues to monitor and supervise COVID-19 in order to inform the public health response ("Latest confirmed cases of COVID-19 in South Africa (20 August 2021)," 2021). They issued a report that detailed the COVID-19 cases' statistics, such as hospital admissions and vaccination rollout.

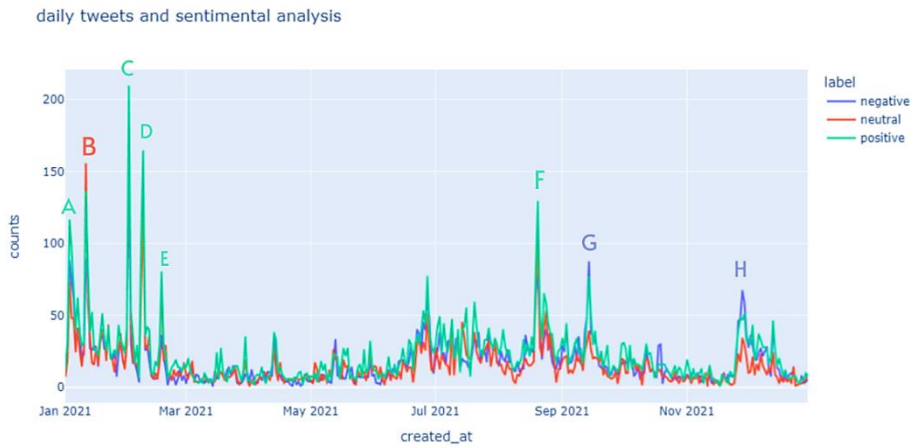


Figure 6: Line graph showing change in tweet sentiment from January 2021 to December 2021

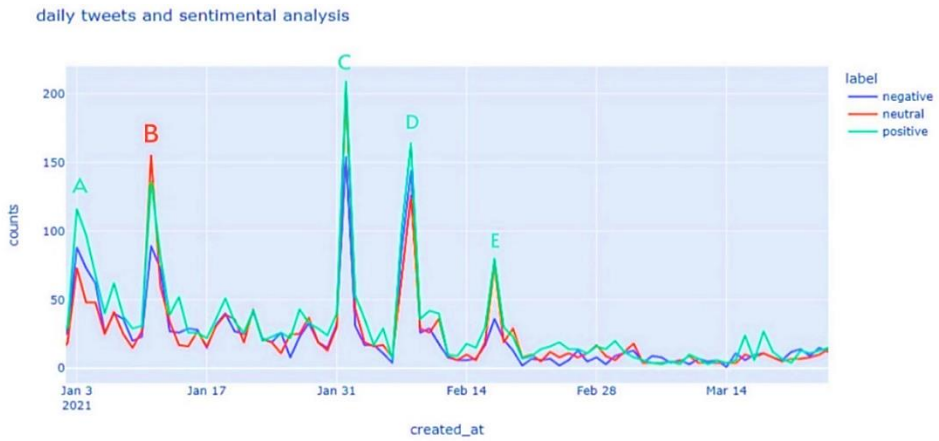


Figure 7: Line graph showing change in tweet sentiment from January 2021 to March 2021.

The 20th August 2021 was the day when people 18 years and older could get the vaccine. There was a drop in engagement from that day on until the 14th September 2021 (G). This was the first spike that contained high negative sentiments with 87 tweets, 77 positive tweets and a low 39 neutral tweets.

"Dear President Before you force us to be vaccinated, please force us to get the RDP's, Food parcels and force your people to bring back R500 billions. Then the vaccine will be the last."

On this day, topics associated with conspiracies, misinformation, and government distrust were more prevalent. The year's final spike may indicate that 2021 ended on a negative note. The spike on the 28th November 2021 (H), was lower than the previous spikes, with 67 negative tweets, 47 positive tweets, and 34 neutral tweets. President Cyril Ramaphosa issued a national declaration announcement (Gov,2021). According to the president's public announcement, the vaccine rollout had reached approximately 41% of the total South African population, with 25 million vaccine doses administered.

"Then why are infections on the rise for vaccinated people? Answer - compromised immune systems "from vax"

"That man his telling the truth about vaccines go n make full research you will see vaccine is from luciferase enzyme wich mean lucifer devil"

In this period there was higher misinformation and hesitancy with tweets such as the one above questioning the need for vaccines. The rest of the year remains with tweets averaging less than 50 tweets a day. This section saw the change in tweet sentiment over time.

5 Conclusion

More than half of the tweets examined contained positive sentiments, with the remaining two-thirds neutral or negative. The findings do not represent a complete picture of the South African demographic, but rather a snapshot of sentiments. Because social media is more popular among younger people, the results may be skewed toward them. We recommend the inclusion of African languages in NTLK dictionaries because it will allow more people and cultures to be accommodated in analysing data.

The use of Twitter data to understand public sentiment can help advance Information Systems as a discipline by providing new data collection techniques and insights into the impact of big data on organisations and society as a whole. Such research can provide valuable information to governments, organisations, and businesses to make informed decisions and respond to the needs and opinions of their stakeholders.

The study was completed prior to Elon Musk's acquisition of Twitter; changes in the way Twitter runs may occur and provide a distinct dimension while performing research on Twitter. For a government to understand public sentiment on Twitter, they can collect relevant tweets, perform sentiment analysis to classify them as positive, negative or neutral, and analyse the results to identify common themes and opinions. This information can help the government make informed decisions and understand public sentiment on important issues. In this study people were sceptical of the South African government's vaccine rollout strategy, which could have hampered overall vaccine adoption. The study concludes that vaccination attitudes changed throughout 2021, particularly in response to government announcements addressing the public on the status of vaccination rollout.

References

Aguilar-Gallegos, N., Romero-García, L. E., Martínez-González, E. G., García-Sánchez, E. I., & Aguilar-Ávila, J. (2020). Dataset on dynamics of Coronavirus on Twitter. Data in Brief, 30. https://doi.org/10.1016/j.dib.2020.105684

Alfatease, A., Alqahtani, A. M., Orayj, K., & Alshahrani, S. M. (2021). The impact of social media on the acceptance of the covid-19 vaccine: A cross-sectional study from saudi arabia. Patient Preference and Adherence, 15, 2673–2681. https://doi.org/10.2147/PPA.S342535

Alsayed, A., Alharthi, S., & Adeyemo, D. A. (2019). Twitter Sentiment Analysis Techniques: A Survey. Journal of King Saud University - Computer and Information Sciences, 31(3), 189-201.

Amjad, A., Qaiser, S., Anwar, A., Ijaz-Ul-Haq, & Ali, R. (2021, September 7). Analysing Public Sentiments Regarding COVID-19 Vaccines: A Sentiment Analysis Approach. 2021 IEEE International Smart Cities Conference, ISC2 2021. https://doi.org/10.1109/ISC253183.2021.9562904

Anuratha, K., Sujeetha, S., Nandhini, J. M., Priya, B., & Paravthy, M. (2021). #Vaccine: Using hashtags from Indian Tweets to Capture and Analyse the Sentiments of People on Vaccination for Covid'19 Pandemic. Advances in Parallel Computing, 39, 88–92. https://doi.org/10.3233/APC210183

Bam, N. E. (2021). Understanding the motivations for covid-19 vaccine hesitancy in south africa: narrative literature review.

Burger, R., Buttenheim, A., English, R., Maughan-Brown, B., Köhler, T., & Tameris, M. (2021.). WAVE 4 3 12 May 2021 COVID-19 vaccine hesitancy in South Africa: Results from NIDS-CRAM Wave 4.

Chadwick, A., Kaiser, J., Vaccari, C., Freeman, D., Lambe, S., Loe, B. S., Vanderslott, S., Lewandowsky, S., Conroy, M., Ross, A. R. N., Innocenti, S., Pollard, A. J., Waite, F., Larkin, M., Rosebrock, L., Jenner, L., McShane, H., Giubilini, A., Petit, A., & Yu, L. M. (2021). Online Social Endorsement and Covid-19 Vaccine Hesitancy in the United Kingdom. Social Media and Society, 7(2). https://doi.org/10.1177/20563051211008817

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR Public Health and Surveillance, 6(2). https://doi.org/10.2196/19273

Chen, X., Association for Computing Machinery. Special Interest Group on Information Retrieval, Association for Computing Machinery. Special Interest Group on Hypertext, H. and W., & ACM Digital Library. (2012.). CIKM'12 : the proceedings of the 21st ACM International Conference on Information and Knowledge Management : October 29 - November 2, 2012, Maui, Hawaii, USA.

Chou, W. Y. S., Gaysynsky, A., & Vanderpool, R. C. (2021). The COVID-19 Misinfodemic: Moving Beyond Fact-Checking. Health Education and Behavior, 48(1), 9–13. https://doi.org/10.1177/1090198120980675

Deverna, M. R., Pierri, F., Truong, B. T., Bollenbacher, J., Axelrod, D., Loynes, N., Torres-Lugo, C., Yang, K.-C., Menczer, F., & Bryden, J. (2021). CoVaxxy: A Collection of English-Language Twitter Posts about COVID-19 Vaccines. https://doi.org/10.5281/zenodo.4526494

Gazette, G., Gazette, R., & Notice, G. (2021). Disaster Management Act: Regulations: Alert level 3 during Coronavirus COVID-19 lockdown: Amendment. www.gpwonline.co.za

Gul, S., Mahajan, I., Nisa, N. T., Shah, T. A., Jan, A., & Ahmad, S. (2016). Tweets speak louder than leaders and masses: An analysis of tweets about the Jammu and Kashmir elections 2014. Online Information Review, 40(7), 900–912. https://doi.org/10.1108/OIR-10-2015-0330

Ioana-Andreea, G., Popa, S., & Ioana, M. (2021). *COVID-19 Vaccine infodemic: Sentiment analysis of the twitter content*. https://global.oup.com/academic/content/word-of-the-year/?cc=ro&lang=en&

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. Vaccine, 35(29), 3621–3638. https://doi.org/10.1016/j.vaccine.2017.05.052

Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis on Twitter. www.youtube.com

Li Han Wong, B., Vybornova, O., Grima, S., & Nia, M. Z. (2022.). Public sentiments toward COVID-19 vaccines in South African cities: An analysis of Twitter posts.

Luo, C., Chen, A., Cui, B., & Liao, W. (2021). Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: Semantic network analysis of two social media platforms in the United States and China. Telematics and Informatics, 65. https://doi.org/10.1016/j.tele.2021.101712

Lyu, H., Wang, J., Wu, W., Duong, V., Zhang, X., Dye, T. D., & Luo, J. (2022). Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. Intelligent Medicine, 2(1), 1–12. https://doi.org/10.1016/j.imed.2021.08.001

Marivate, V., & Combrink, H. M. (2020). Use of available data to inform the COVID-19 outbreak in South Africa: A case study. Data Science Journal, 19(1), 1–7. https://doi.org/10.5334/dsj-2020-019

Mir, A. A., Rathinam, S., & Gul, S. (2022). Public perception of COVID-19 vaccines from the digital footprints left on Twitter: analyzing positive, neutral and negative sentiments of Twitterati. Library Hi Tech, 40(2), 340–356. https://doi.org/10.1108/LHT-08-2021-0261

Mir, A. A., & Sevukan, R. (2022). Sentiment analysis of Indian Tweets about Covid-19 vaccines. Journal of Information Science. https://doi.org/10.1177/01655515221118049

Mohamed Ridhwan, K., & Hargreaves, C. A. (2021). Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. International Journal of Information Management Data Insights, 1(2). https://doi.org/10.1016/j.jjimei.2021.100021

Mønsted, B., & Lehmann, S. (2022). Characterizing polarization in online vaccine discourse—A large-scale study. PLoS ONE, 17(2 February). https://doi.org/10.1371/journal.pone.0263746

Niemiec, E. (2020). COVID-19 and misinformation . EMBO Reports, 21(11). https://doi.org/10.15252/embr.202051420

Pawar, K. K., Shrishrimal, P., & Deshmukh, R. R. (2015). Twitter Sentiment Analysis: A Review. International Journal of Scientific & Engineering Research, 6(4). http://www.ijser.org

Puri, N., Coomes, E. A., Haghbayan, H., & Gunaratne, K. (2020). Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Human Vaccines and Immunotherapeutics, 2586–2593. https://doi.org/10.1080/21645515.2020.1780846

Rotolo, B., Dubé, E., Vivion, M., MacDonald, S. E., & Meyer, S. B. (2022). Hesitancy towards COVID-19 vaccines on social media in Canada. Vaccine, 40(19), 2790–2796. https://doi.org/10.1016/j.vaccine.2022.03.024

Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2020a). An Exploratory Study of COVID-19 Misinformation on Twitter. http://arxiv.org/abs/2005.05710

Mpofu, S. (2021). Digital Humour in the Covid-19 Pandemic. Palgrave Macmillan.

Shoaei, M. D., & Dastani, M. (2020). The role of twitter during the COVID-19 crisis: A systematic literature review. In Acta Informatica Pragensia (Vol. 9, Issue 2, pp. 154–169). University of Economics - Prague. https://doi.org/10.18267/J.AIP.138

van Heerden, A., & Young, S. (2020). Use of social media big data as a novel HIV surveillance tool in South Africa. PLoS ONE, 15(10 October). https://doi.org/10.1371/journal.pone.0239304

Appendix A: Stop words for text processing (100 words)

['setsidiki1','meant','covid19','vaccine','booster','irvinjimsa','sihlewasembo','arw','forcing','thing','kuthi','respect','covid','care','putting','pressure','us','vaccinate','pls','siyekeni','sizaziyela','ngelethu','xesha','vaccinated','curfew','lifted','thank','jesus','vaccinefree','curfew','cabinet"hhaybo','guys','drink','won't','take','vaccine','doctoriamthe','happened','natural','immunity','infection','successful','recovery','deltaomricon','still','discriminated','medical','institutions','advocating','vaccine','employer','theres"robot','vaccine','pfizerwhistleblower','yerlydave','petersweden7','moosefucker','logic','go','get','vaccinated','stop','blabbing','tweeter','vaccinate','dammit','tell','partner','protect','times','goes','cause','shes','double','jabbed','dont','make','supergirl','must','still','follow','covid"guidelines','period','pfizerwhistleblower','decide','vaccinate','vaccine','get','jampj','pfizer','ceo','pfizer','taken','vaccine','pfizerwhistleblower','guys','believe']