Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Technologies
FEM CZU Prague

Master's Thesis

Analysis of Twitter content related to Covid-19 vaccination using NLP

Tegegn Addiskidan Samuel

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

BSc. Addiskidan Tegegn

Informatics

Thesis title

Analysis of Twitter content related to Covid-19 vaccination

Objectives of thesis

The main objective of the thesis is to analyze the relationship between the Covid-19 vaccination rates and Twitter content.

Partial objectives

- To identify relevant hashtags, download the Twitter dataset and conduct sentiment analysis.
- To build a model and run statistical analysis.
- To evaluate the results and interpret findings.

Methodology

·The methodology of solving the theoretical part of the diploma thesis will be based on the study and analysis of professional information sources. Based on the theoretical knowledge gained in work, the author will formulate specific research questions, collect relevant datasets, and build and run statistical analysis. The results will be interpreted and contrasted with other similar studies. Based on the synthesis of theoretical knowledge and the results of the practical part, the conclusions of the work will be formulated.

The proposed extent of the thesis

80 pages

Keywords

Covid-19, vaccination, Twitter, relationship, moderation

Recommended information sources

COPPING, L. T. Anxiety and covid-19 compliance behaviors in the UK: The moderating role of conspiratorial thinking. Personality and Individual Differences, 2022, 192: 111604.

DI DOMENICO, Giandomenico, et al. Fake news, social media and marketing: A systematic review. Journal of Business Research, 2021, 124: 329-341.

KÜÇÜKALI, Hüseyin, et al. Vaccine hesitancy and anti-vaccination attitudes during the start of COVID-19 vaccination program: A content analysis on Twitter data. Vaccines, 2022, 10.2: 161.

SHAREVSKI, Filipo, et al. Misinformation warnings: Twitter's soft moderation effects on covid-19 vaccine belief echoes. Computers & security, 2022, 114: 102577.

SHAREVSKI, Filipo, et al. (Mis) perceptions and Engagement on Twitter: COVID-19 Vaccine Rumors on Efficacy and Mass Immunization Effort. International Journal of Information Management Data Insights, 2022, 100059

Expected date of thesis defence

2022/23 SS – FEM

The Diploma Thesis Supervisor

Ing. Miloš Ulman, Ph.D.

Supervising department

Department of Information Technologies

Electronic approval: 14. 7. 2022

doc. Ing. Jiří Vaněk, Ph.D.

Head of department

Electronic approval: 28. 11. 2022

doc. Ing. Tomáš Šubrt, Ph.D.

Dean

Prague on 22. 03. 2023

Declaration

I declare that I have worked on my master's thesis titled "Analysis of Twitter content related to Covid-19 vaccination using NLP" by myself, and I have used only the sources mentioned at the end of the thesis. As the author of the master's thesis, I declare that the thesis does not break any copyrights.

In Prague on 31/03/2023 ____________________________

Acknowledgement

First and foremost, I would like to thank almighty God who has helped me throughout my journey. I would like to express my deepest gratitude to my thesis advisor Ulman Miloš whose guidance and support have been valuable throughout this process.
Their expertise, patience, and encouragement have been instrumental in shaping this work.

I would also like to thank the faculty and staff of Economics and Management for their support and encouragement throughout my academic journey. Their dedication to teaching and research has been a source of inspiration and motivation for me.

I would like to extend my heartfelt thanks to my family and friends for their unwavering support and encouragement. Their love and encouragement have been my constant motivation throughout my studies.

Finally, I would like to thank all the participants who generously shared their time and insights for this study. Their contributions have been instrumental in making this research possible.

Abstract

This thesis examines the use of natural language processing (NLP) techniques to analyze tweets related to COVID-19 vaccination on Twitter. The study aims to identify the sentiments, topics, and opinions expressed by Twitter users regarding the COVID-19 vaccine, and to assess the impact of these expressions on public health communication efforts. Using a dataset of over one million tweets related to COVID-19 vaccination, the study employs various NLP techniques, including sentiment analysis, topic modelling, and opinion mining. The results show that there is a wide range of sentiments and opinions expressed on Twitter regarding the COVID-19 vaccine, with both positive and negative sentiments present. The study also identifies several key topics that are frequently discussed, including vaccine safety, vaccine efficacy, vaccine hesitancy, and vaccine mandates.

The findings of the study highlight the potential of NLP techniques in analysing social media data to gain insights into public opinions and attitudes towards the COVID-19 vaccine. The results can help inform public health communication strategies aimed at addressing vaccine hesitancy and increasing vaccine uptake. Additionally, the study provides valuable insights for policymakers, healthcare professionals, and researchers who seek to understand the impact of social media on public health communication efforts.

Keywords:

COVID 19, Vaccination, sentiment analysis, Twitter.

Abstraktní

Tato práce zkoumá využití technik zpracování přirozeného jazyka (NLP) k analýze tweetů souvisejících s očkováním proti COVID-19 na Twitteru. Cílem studie je identifikovat pocity, témata a názory vyjádřené uživateli Twitteru ohledně vakcíny COVID-19 a posoudit dopad těchto projevů na úsilí o komunikaci s veřejným zdravím. S využitím datové sady více než jednoho milionu tweetů souvisejících s očkováním proti COVID-19 studie využívá různé techniky NLP, včetně analýzy sentimentu, modelování témat a získávání názorů. Výsledky ukazují, že existuje široká škála názorů a názorů vyjádřených na Twitteru ohledně vakcíny COVID-19, s pozitivními i negativními náladami. Studie také identifikuje několik klíčových témat, o kterých se často diskutuje, včetně bezpečnosti vakcín, účinnosti vakcín, váhání s vakcínami a mandátů k očkování.

Závěry studie zdůrazňují potenciál technik NLP při analýze dat sociálních médií, aby bylo možné získat náhled na veřejné mínění a postoje k vakcíně COVID-19. Výsledky mohou pomoci informovat o komunikačních strategiích veřejného zdraví zaměřených na řešení váhání s očkováním a zvýšení absorpce vakcín. Studie navíc poskytuje cenné poznatky pro tvůrce politik, zdravotníky a výzkumné pracovníky, kteří se snaží porozumět dopadu sociálních médií na úsilí o komunikaci v oblasti veřejného zdraví.

Klíčová slova:

COVID 19, očkování, analýza sentimentu, Twitter.

Table of Content

List of Tables

List of Figures

LIST OF ABBREVIATION

NLP	Natural language processing
CNN	Convolutional neural network
GRU	Gated recurrent unit
RNN	Recurrent neural network
LSTM	Long short-term memory

Chapter One

1.1 Introduction

Social media has become an essential forum for exchanging information and where individuals share their opinions about the most diverse topics, making it an excellent data source for researchers on public opinion. Twitter is at the forefront of social media platforms where individuals share their views with other users worldwide. This global reach can create public discussions and influence individuals' preferences, beliefs s and choices. In late December 2019, an outbreak of mysterious pneumonia characterised by fever, dry cough, fatigue, and occasional gastrointestinal symptoms happened in a seafood wholesale wet market, the Huanan Seafood Wholesale Market, in Wuhan, Hubei, China (Wu et al. 2020). The coronavirus outbreak has seriously threatened global public health (Covid 19). In 2020, COVID-19 became the centre of all public attention and the same translated into trending topics on Twitter. International health authorities were committed to containing COVID-19 outbreaks in light of this circumstance, requiring residents to utilise personal protective equipment and limiting their freedom of movement, among other measures. Therefore, the swift creation of a COVID-19 vaccination that is both safe and effective immediately became a top priority for global institutions (Pirrotta et al. 2022). The evolution of the pandemic and the appearance of vaccination options are widely discussed on the Twitter feed. With many different opinions on how to perceive it, a constant flow of misinformation was also noticed. Social media is an excellent means of communicating with others, especially during a pandemic, to spread information on current restrictions, precautions, and news about the virus. However, as anyone can write whatever they want on those platforms, many said facts have not necessarily been verified by professionals. More often, people have taken the uneducated conspiratorial theories about COVID-19 seriously, leading to increased anxiety from reading frightening information and anti-vaccination propaganda on the internet faster than the official truthful news. Therefore, social media is only a means of getting more information and communicating with others for them. However, when looking at the total population, many cannot think critically and believe everything they read. In that case, the many misleading posts on social media can lead to them not trusting the vaccine and other scientific information about the virus SARS-CoV-2. To this end, the literature devotes understanding how Twitter influences the COVID -19 vaccination and hesitancy. The subject of COVID-19 vaccine hesitancy has different outcomes.

The results can be divided into two groups - those that say that social media add to vaccine hesitancy and those that dispute that.

Twitter sentiment analysis towards the COVID- 19 vaccination is imperative, engaging more users in the discussion and contributing to the global public health vaccination campaign. The present study has the goal of analysing the overall sentiment of public opinion towards vaccines and how that translates into the vaccination rate of individuals using the NLP.

Chapter Two

Objectives and Methodology

2.0 Main Objectives

The main aim of the thesis is to analyse the relationship between the COVID-19 vaccination rates and Twitter content.

2.1 Specific Objectives

1. To evaluate the perception of COVID-19 vaccination through Twitter content
2. To monitor the attitudes of Twitter users towards COVID-19 vaccination
3. To determine the positive, negative, and neutral sentiment behind tweets

2.2 Methdology

To achieve the objectives of the study, the methodology to be used in this research follows three main steps, first collecting Twitter data then data preprocessing and then analaysing using sentimental analaysis, Decision trees and By using Deep Machine learning techniches finally providing outputs of findings. The method used in this research is python Programming.

Chapter Three

Literature Review

3.0 Brief Description of Twitter

Twitter is a social medium widely utilised by ordinary people and influential characters. On Twitter, ideas are shared within networks of like-minded individuals, not isolated from the rest of the platform. This likely leads to an echo-chamber effect that can disseminate informative, fallacious, or biased information (Thelwall et al. 2021). Understanding the demographics of Twitter users can help picture a more precise panorama of Twitter's influence on the world's population. According to the Digital Report (2022), the total potential reach of advertisements on Twitter was 486 million users in July 2022, namely 6.1 percent of the world's total population and 9.7 percent worldwide internet users. Of these, 27.3 percent were female users, while 72.7 percent were male users. The five countries with the highest share of Twitter users were the United States of America, with 83.4 million active users; Japan (60.7 million users); India (24.8 million users); Indonesia (21.2 million users); and the United Kingdom (19.8 million users). In the US, the most comprehensive group of Twitter users included people who were aged 30-49 years in 2019, which accounted for 44 percent of the total number of Twitter users on the national level; besides, their income and education level were higher than the ones of the average US citizen (Wojcik et al. 2019).
Although Twitter is not the most popular and used existent social medium – Instagram, for comparison, had 1.44 billion users in July 2022, 18.1 percent of the world's total population (Digital Report 2022) –Twitter still represented one of the leading thought and informational hubs for regular users, healthcare professionals and political leaders during the years of the COVID-19 pandemic (Shoaei and Dastani 2020).

3.1 Elon Musk’s purchase of Twitter

In October 2022, Elon Musk purchased Twitter for forty-four billion dollars, accenting helping humanity and promising a platform essential to lead civilisation by promoting positive debate concerning briefs and a violent-free environment (Cassidy 2022). However, Cassidy (2022) argues that Musk’s insistence on free speech can create a bias and compromise in Twitter content which has in the past years come under pressure to restrict

Twitter content for users do not glorify violence, violent extremism, self-induced harm, or discriminate against religion, sexuality, disability, or ethnicity.

The supporters of Musk show optimism in the financial incentives and investment to make the platform successful. Thus far, proposals from Musk have committed to increasing the subscription base, eliminating fake accounts, longer twits and the possibility to edit twits (Cassidy 2022). Additionally, Barrie (2022) pinpoints that Musk's Twitter acquisition shows a positive and significant increase in user post engagement on Twitter. The study indicated an increase in Twitter traffic during the purchase month's end of October, with an increase of 28 per cent in retweets and 4 per cent likes.

On the other hand, the pessimistic of Musk's Twitter acquisition contends that it is naive and tones down to a laissez-faire approach to content. Despite Musk's claims of being politically centric, pessimistic argue that Musk's last Tweets have targeted individual abuse (Cassidy, 2022). Further, Chiu (2022) contends that the acquisition of Twitter by Musk and the agenda of transforming twitter into a common digital platform is a threat to Twitter users. This notion is highlighted by Oxford Analytica (2022), which argues that promoting free speech on Twitter will promote disinformation. The further contention is that the addition of Twitter from Tesla's Chief executive officer would raise an exciting rest in non-democratic countries such as China, which are leaders in Tesla with a perception of content moderation on Twitter (Oxford Analytica, 2022).

3.2 Background of Corona Virus Disease (COVID-19)

The root of the COVID-19 pandemic goes back to Wuhan, Hubei Province, China (Ciotti et al., 2019). The Coronavirus disease (COVID-19) is a highly contagious viral infection caused by the SARS-CoV-2 virus and transmitted through droplets expelled by infected people coughing, sneezing, or speaking. COVID-19 is a respiratory illness that, despite causing mild symptoms in most individuals, can arise in acute ones in the elderly and people affected by other medical conditions such as diabetes, cancer, and chronic respiratory or cardiovascular diseases.

In such cases, COVID-19 can be hazardous to life and needs to be treated medically (United Nations Development Programme 2020). Though transmission might be feasible before symptoms appear in patients, it is considered most contagious when people are ill. The specific days between exposure and symptoms start between two and fourteen days.

Shortness of breath, coughing, fever, and sneezing are typical symptoms. Potential complications include acute respiratory distress syndrome, pneumonia, and sore throat (Ahmad 2020).

Firstly, identified in Wuhan, China, in December 2019, COVID-19 spread worldwide, leading most countries to implement restrictions on assemblages and international travel. Such decisions heavily damaged the global economy and resulted in recession, while healthcare systems came under pressure (Qorib et al. 2022). Until November 2020, the main anti-infection precautions included wearing a face mask, sanitising one's hands, maintaining a safe distance of at least one meter from other people, sneezing and coughing in one's elbow crease, quarantining in case of infection or symptoms, and avoiding unnecessary movements (United Nations Development Programme 2020).

Figure 1 shows the highest number of confirmed COVID-19 cases from January 2020 until December 15, 2022. The highest number of confirmed COVID-19 cases were reported in Europe (257 081 702 cases in total), and the lowest number of COVID-19 cases stated by WHO were in Africa, with only 12 974 380 cases.

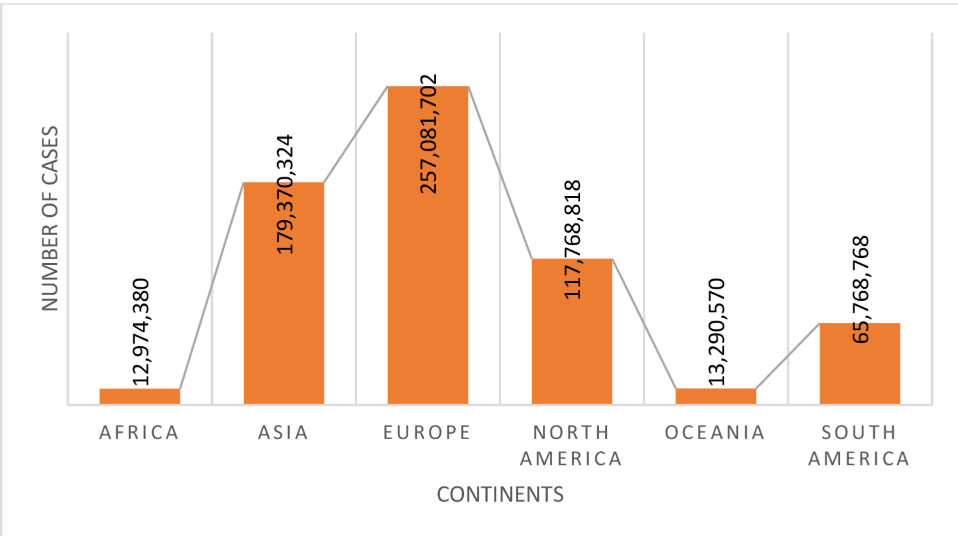

Figure 1 Global COVID-19 rates by continents, January 2020 until 15 December 2022. https://www.who.int/data)

3.2 Covid-19 and the Anxiety Effect

During the initial stage of the COVID-19 epidemic in China, local researchers conducted research on psychological responses in the general population and found that 28.8% of respondents reported moderate to severe anxiety symptoms, 16.5% said mild to severe depressive symptoms, and 53.8% rated outbreak's psychological impacts as moderate or severe. (Sher 2020).

An increase in anger, sleep disturbances, distress, depression, and anxiety were found among inhabitants of all affected places, and the decrease was found in life satisfaction and most positive emotions. This problem is not found only among non-healthcare people but also among healthcare professionals who suffer from depression, insomnia, anxiety, and distress (Sher 2020). As this pandemic made significant economic and social threats to society, it was vital to preserve healthy minds. The experts recommend connecting with others, as loneliness appears a lot during this time, finding sources of inspiration and joy, daily rhythms to keep track and staying mindful (Peteet 2020).

In November 2020, the first vaccines were officially released, setting the conditions for systematic vaccination campaigns worldwide to combat the COVID-19 pandemic (Qorib et al. 2022). About 200 different vaccines were created against SARS-CoV-2. Some were more efficient than others, with Pfizer and Modena being presented as one of the best. However, developing the vaccines was just the beginning of the process. Since vaccination works best when most of the population is vaccinated, the aim was to convince as many people as possible to get the shot, and social media played a massive role in that matter (Parker et al. 2020)

3.4 covid vaccination and the factors affecting covid-19 vaccination

To significantly lower morbidity and death from COVID-19, an effective and safe vaccination had to be quickly and widely distributed to the population as soon as it was developed. To ensure widespread immunological protection, a vaccine must be available and well-received by the medical profession and the general population (Schaffer Deroo et al. 2020). To ensure that promising vaccines can be produced in sufficient numbers and fairly distributed to all impacted areas, especially low-resource countries, solid international coordination, and cooperation between research groups engaged in vaccine research and development is required.

As countries strive to stop the spread of the virus and its mutations, the Coronavirus Disease 2019 (COVID-19) pandemic has had adverse health and economic effects worldwide. To facilitate the creation, distribution, and use of COVID-19 diagnostics, treatments, and vaccinations, WHO and its partners introduced ACT-A in April 2020.

The World Health Organization (WHO), the European Commission, and France organised access to COVID-19 through this ACT to hasten the development, manufacture, and equitable and efficient access to COVID-19 diagnostics, vaccinations, and medicines. The COVID-19 Vaccines Global Access Facility (COVAX) was subsequently established by the World Health Organization (WHO) and Epidemic Preparedness Innovations (CEPI) to assist all nations or areas in obtaining enough COVID-19 doses through the distribution of vaccine contracts (Li et al. 2021). Technically, COVAX's system for distributing vaccines worldwide is efficient, equitable, and doable. The distribution, however, cannot be carried out in a way that fully ensures effectiveness and justice – in reality. The existing vaccine distribution system is inefficient and unfair across nations, regions, and social levels (Li et al. 2021).
Many analysts argued that at the current rate of vaccine distribution, COVID-19 vaccinations would not be widely accessible in Low Medium Income Countries until at least the year 2022.
The great majority of manufactured doses of the COVID-19 vaccination have been given out in high-income countries as follows 1 in 4 of their populations, whereas, in low-income countries, this ratio is 1 in 500 (Binagwaho et al. 2022). The right to life and health is never based on socioeconomic status. The COVID-19 vaccination should be affordable to most individuals, particularly those from disadvantaged and low-income groups (Binagwaho et al. 2022). On the other hand, studies on the factors of COVID-19 vaccination attitudes find them to be more unfavourable among low-income, less educated, ethnic minorities, and the young. International comparisons show that Asian nations with high confidence in the central authority have the most favourable sentiments. In contrast, Central and Eastern Europe have more negative attitudes, presumably as a legacy of Soviet communism (Kachurka et al. 2021).

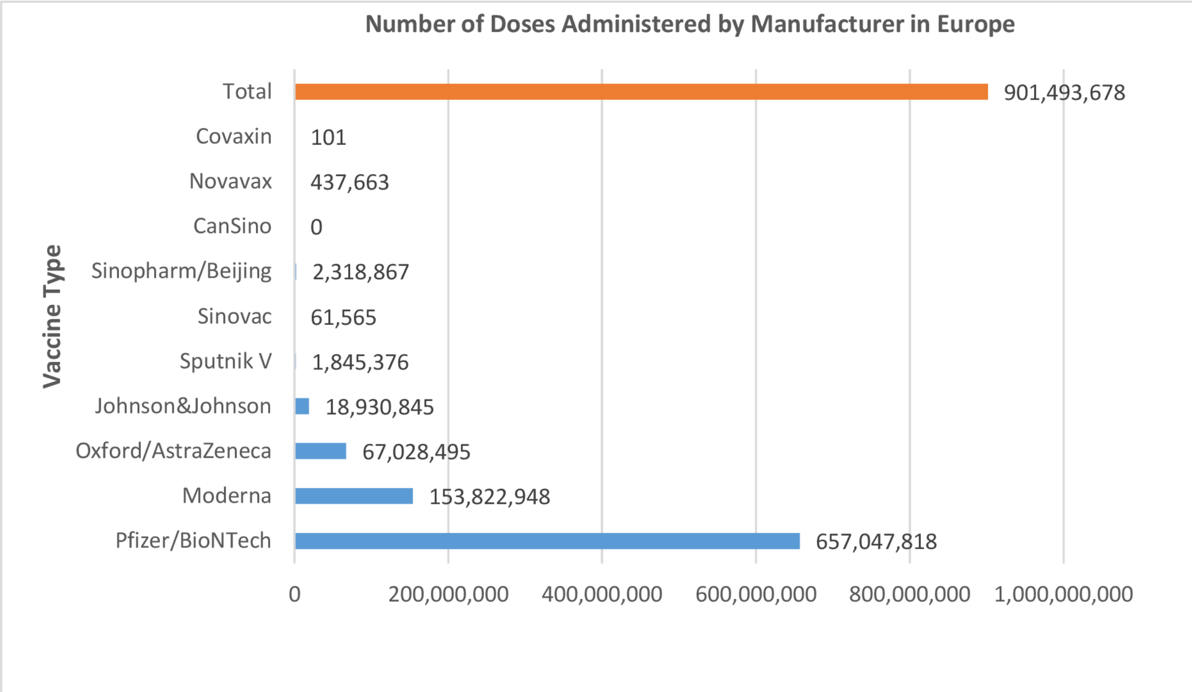

Figure 2 COVID-19 vaccine doses administered by manufacturer, European (https://www.who.int/data)

A study by Chen et al. (2022) concentrated on exploring factors that influence COVID-19 vaccination intention with a focus on the selection of different media. The results indicate that higher knowledge and risk perception contribute positively to vaccination intention. Another finding reveals a correlation between a selected media source and knowledge. Respondents who preferred text media over videos showed higher levels of expertise than those who favoured video usage. That suggests a link between text media usage and a rise in vaccination intention.According to Figure 2, there has been a total of 901 493 678 administered doses of different manufacturers and types in the EU. For 657 047 818 of those cases, Pfizer (or BioNTech) is responsible, making up for 73 % of all administered vaccines in the EU. The second most used vaccine type was Moderna with 153 822 948 vaccines, and the third was Oxford (or AstraZeneca) vaccine with 67 028 495 vaccines administered.

Figure 3 shows the number of countries administering a particular vaccine type. Out of the 27 types, ten vaccines are distributed in more than ten countries. Namely Oxford/AstraZeneca, Pfizer/BioNTech, Moderna, Johnson & Johnson, Sinopharm/Beijing, Sinovac, Gamaleya, Novarax, Covaxin (Bharat Biotech), and CanSino. The remaining 17 vaccine types are primarily regional or only administered in one country.

The most widespread vaccine type is Oxford/AstraZeneca, distributed in 185. The second most used vaccine type is Pfizer/BioNTech, with 165 countries; the third is the vaccine type named Moderna, which is being administered in 113 countries.

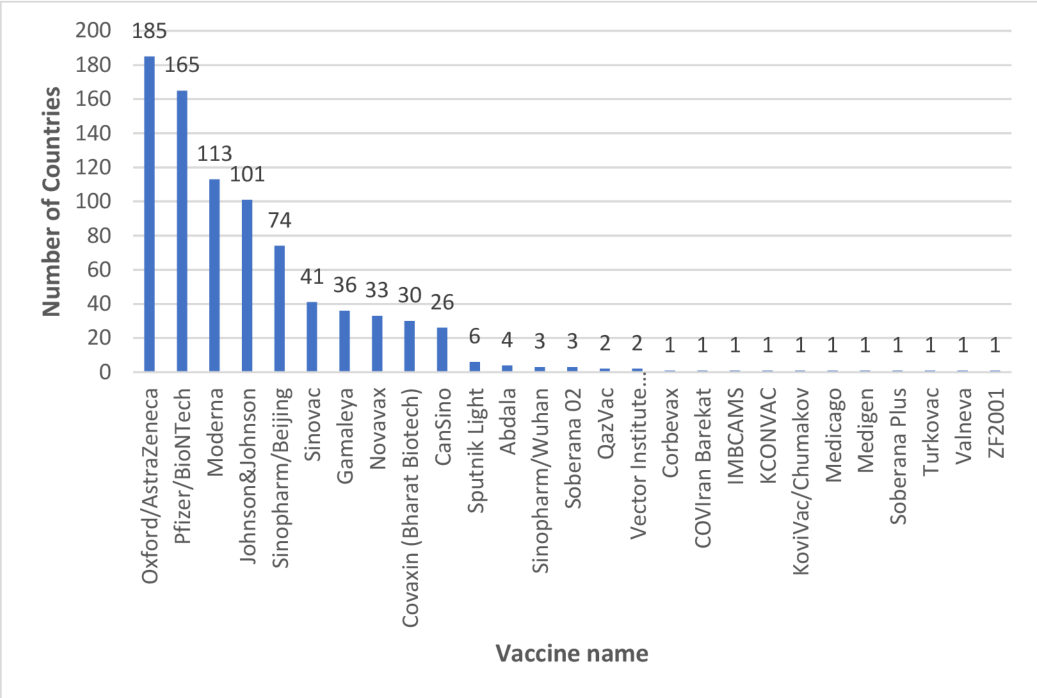

Figure 3. The number of countries that administer a particular vaccine type (https://www.who.int/data)

3.5 The influence of twitter on the uptake of covid -19 vaccines

Public health experts face difficulty when a pandemic strikes because they must plan a quick and efficient communication strategy to tell the public about the hazards of the pandemic and the best way to respond behaviourally to those risks (McNeill et al. 2016). During this epidemic, people were urged or required to stay home, engage in social seclusion, and carry out most of their work and daily routines remotely. Due to this isolation, more people are using social media (Twitter) to read news and share their thoughts (Rosenberg et al. 2020). The internet – and social media for the most part – have been the primary means of disseminating official and unofficial information regarding the pandemic and the vaccines (Igoe 2019). Freely expressing people's opinions and perspectives, social media have often embodied and reflected people's uncertainty about COVID-19 vaccines, disseminating both misinformation and disinformation (Igoe 2019).

Since the start of the SARS-CoV-2 pandemic, social media has emerged as one of the primary platforms for debating the different topics connected to the COVID-19 epidemic's global spread. However, the material shared on these platforms frequently lacks scientific backing, which fuels public uncertainty, mistrust, and terror (Muñoz-Sastre et al. 2021).

Misinformation is fallacious that depicts an incomplete panorama of a given topic. Disinformation intentionally disseminates false and biased news to promote an ideological or economic agenda. Social media's misinformation and disinformation have instigated COVID-19 vaccine hesitancy, which – according to the World Health Organisation (2019) – has been one of the significant threats to public and global health during the recent pandemic. This affects the mental health of patients, endangering both their physical and psychological health. There may be a rise in suicidal thoughts or attempts, which the world must increasingly prepare for during this epidemic. The two most prominent examples are the increase in vaccination scepticism and social media's role in propagating false and unfavourable information (Rosenberg et al. 2020). No matter how well-intentioned or malicious, the propagation of false information causes mistrust in medical advice due to the abundance of false information and causes panic, incorrect prescribing, and reduced reaction to warnings on concerns like social distancing (McNeill et al. 2016; Rosenberg et al. 2020). However, users can draw inspiration from tales of heroic deeds, admirable role models, and international initiatives to fight the pandemic.

These stories can help people work through this issue together, from frontline healthcare personnel's accounts of immeasurable sacrifices to the humorous accounts of non-medical users merely making it through a quarantine period (Rosenberg et al. 2020; Pirrotta et al. 2022).

3.6 Sentimental Analaysis of public opinion towards covid-19 vaccine

Sentiment can be described as one's opinions and feelings. Sentiment analysis examines people's interests, assessments and emotions based on the content they choose to write and share (Chinnasamy 2022). People frequently seek out other people's opinions when making decisions; hence what someone shares on social media might considerably impact the conclusions one might take in the future. Therefore, social media has emerged as a crucial forum for the public to gather information and voice ideas, and people may share various viewpoints, beliefs, and feelings (Xu et al. 2022).

This has been extremely visible regarding the COVID-19 pandemic and the vaccination campaigns. Public confidence in the vaccine is an increasingly important public health issue globally, so addressing negative vaccine sentiments is essential for COVID-19 prevention. Assessing the overall sentiment is quite challenging; analysing social media data, which contains people's unfiltered thoughts, can offer valuable information that could guide the institution in promoting vaccines more efficiently (Sun and Budhwani, 2022). Chinnasamy et al. (2022) observed from their research that vaccines had a beneficial impact on the population, with most individuals assuming either a neutral or positive attitude towards the different types/brands of vaccines available. Overall they also observed that people have a high level of trust in the vaccines. Still, fearful feelings are also quite present in the online community, mainly due to the side effects that many immunisations cause. Other studies have shown the same trend, with the prevalence of positive as the dominant message and higher engagement from Twitter users (Yousefinaghani et al. 2021).

During the COVID-19 pandemic, traditional and social media were filled with various levels of information. Some sources provided good-quality, fact-checked analysis, while others concentrated on affecting society by distributing false information. Jemielniak and Krempovych (2021) found that the tweeting was partly well-organised and had at least some motivation for political influence.

These authors also draw attention to the case of AstraZeneca vaccinations that were demonised by the Russian news network aiming to increase the desirability of the Russian-manufactured Sputnik vaccine. Furthermore, Yousefinaghani et al. (2021) confirm that political activists, bots and state media were crucial players in producing meaningful content during the pandemic. What is interesting in their findings is that overall, the anti-vaccination tweets were more widespread than the pro-vaccination ones.

Also, different actors were in favour of different types of opinions. While official governmental agencies were pro-vaccination, the political activists and bots circulated negative views via automated processes, resulting in many tweets in a short time. Despite the negative posts outnumbering the positive feed, negative posts generally got lower engagement and retweet rates compared to the positive tweets.

This could be due to people seeking trusted information during a crisis and refraining from engaging with suspicious material. On the other hand, people might have wanted to spread and engage with positive emotions related to the positive tweets.

As the COVID pandemic was passing through the world, new vaccinations with ground-breaking technology raised many opinions among scholars, scientists and ordinary individuals. Consequently, understanding public opinion has become more critical than ever before. Sharevski et al. (2022) discovered that false information affected the willingness to take COVID-19 vaccination. If a person was hesitant or against the vaccination administration, they were more likely to support the conspiracy theories and false information. On the contrary, people who did not engage with anti-vaccination content on Twitter had a more vaccination-positive attitude. As Liu and Liu (2021) mentioned in their study, Twitter has become essential for understanding public opinion regarding wide societal phenomena. Acknowledging current public opinion could help governments understand when and how to implement public health initiatives. Hence, efficient timing could be one of the critical factors for successful public health initiatives. Chen et al. (2022) agree that analysing social media outputs could help understand the wider population's opinion. However, sometimes developing methodology is challenging. Artificial intelligence and models prepared before any specific case could help understand the content efficiently.

In summary, the overall Twitter sentiment towards the COVID vaccines was positive, engaging more users in the discussion and contributing to the global public health vaccination campaign. On the other, much misinformation and negative sentiment were mainly created by bots or fomented by institutions interested in promoting their vaccines, like the case of Russia Today. In this sense, Twitter is a good mirror of our society, and the positive sentiment also represents individuals' trust in their governments, health institutions and professionals. Since the campaign started, there have been several variations in the sentiment. Still, as more people became vaccinated and the mortality rates began to drop, the sentiment became more positive, influencing more and more Twitter users to get inoculated. Lastly, from the existing literature, it is impossible to translate the sentiment into the actual vaccination rate of the users since we don't have this data. However, it is visible that the global trend was for more and more individuals to decide on taking the vaccine, following the overall positive sentiment described in the literature.

3.7 The Effect of Twitter on Public Attitudes on Covid 19 Vaccinations

The different posts about vaccination on Twitter were assessed in a research study, 'Public attitudes on social media toward vaccination before and during the COVID-19 pandemic, by Shah et al. (2022). The studied tweets were posted in five different languages - French, English, Swedish, Dutch and Italian. The study's results made it clear that there were positive and negative opinion shifts, but the overall opinion changed in favour of the vaccination. The COVID-19 vaccine acceptance rate has yet to attain the necessary level for herd immunity. To create effective implementation strategies to promote COVID-19 vaccines, it is essential to comprehend why some people are eager to receive vaccinations while others are not. Social media has acted as a vehicle and a source for the spread of misinformation about vaccines. The algorithms of many sites, including Twitter, facilitate the grouping of like-minded groups that spread identical anti-vaccination information in a feedback loop that has persuasive power with the hesitant (Carrieri et al. 2019; Pirrotta et al. 2022).

.

Research Aim	Sample Size	Study Period/Data Collection	Methods of Analysis	Main findings	Reference
To examine the intentions on the uptake of COVID-19 vaccination.	1205 respondents.	16 March to 29 April 2020. Cross-sectional online survey	Multivariate regression and logistic regression methods.	The covid vaccination rate was 49 per cent, and those who intended to be vaccinated were 63 per cent.	Kwok et al. (2021).
To investigate the opinions and interests of the Public on COVID-19 vaccines.	undefined	2021. Panda's library data set was converted to a data frame.	Sentiment analysis-Tweeter API and Natural Language Toolkit.	The sentiment analysis shows that more responses were neutral than positive, and negatives were the least.	Chinnasamy et al (2022).
To investigate vaccination attitudes from Twitter posts.	2,198,090 tweets from Western Europe	January 20, 2020, and March 15, 2021.	Natural Language Processing (NLP)-Natural Language Toolkit (NLTK)	The findings indicate the validation of the text-based NLP, which promotes vaccination uptake. Sixty per cent show a positive attitude, while 20 per cent have a negative attitude.	Chen et al. (2022).

The study outlines social media's position on the vaccine and its relationship with compliance with vaccination campaigns.	A total of 213 respondents, while only 207 were analysed.	In April 2020. A web link to an online survey.	Multivariate analysis and construction of a decision tree.	Sixty-eight perfect social media users did not get vaccinated against influenza in 2019. Hence the researchers perceived a similar pattern for covid vaccines.	Benis et al. (2021).
To investigate trends in the sentiment of COVID-19 vaccine-related tweets.	2,678,372 COVID-19 vaccine-related tweets.	November 1, 2020, to January 31, 2021. English-language tweets.	Valence Aware Dictionary and Sentiment Reasoner Tool and Latent Dirichlet Allocation Analysis.	The findings show positive sentiments in tweets with 42.8 per cent, 26.9 per cent neutral and 30.3 per cent negative.	Liu and Lui (2021).
The role of social media and the determinants of public vaccine and hesitancy.	Sample size 29,755	February to March 2021. Cross-national survey (27 EU Member States).	Multivariate regression models.	Social media is a primary source of COVID-19 vaccinations. Statistical significance was found in the hesitancy of vaccines among social media users.	Mascherini and Nivakoski (2022).

To investigate the opinion of the public on the Covid-19 vaccine and hesitancy.	Undefined	2021. Live streaming public tweets search-Application Programming Interface (API)	Sentiment analysis methods (VADER, Azure Machine Learning, and TextBlob). Learning algorithms such as Naïve Bayes, Logistics Regression, Decision Tree, Random Forest, and LinearSVC). Vectorisation methods- (TF-IDF, Doc2Vec, and CountVectorizer)	The results showed that the public had confidence in the COVID-19 vaccination, and this reduced vaccine hesitancy.	Qorib et al. (2023).
To examine the perception and engagement on Twitter regarding the COVID-19 vaccine.	606 participants	January and February 2021 using the questionnaire survey of the tweets.	The attitudes were analyzed on a yes / No response on i) expect the effective vaccine to be developed, ii) will take a COVID-19 vac- cine (Yes/No/I neutral); and iii) children to get a COVID-19 vaccine (Yes/No). Twitter users' engagement on a 7-point Likert scale (1-extremely likely; 7-extremelyunlikely).	The findings indicate that the vaccine rates were influenced by twitter users' alteration of content affected the COVID-19 vaccine and engagement.	Sharevski et al. (2022).

Table 1 The Effect of Twitter on Public Attitudes on Covid 19 Vaccinations

3.8 COVID-19 Vaccination Hesitancy

Vaccination hesitancy is a significant obstacle to increasing the acceptance of the COVID-19 vaccine (Willis et al. 2023). Vaccine hesitancy is the diffidence or reluctance of a particular portion of society toward vaccines' safety, efficacy, and usefulness. This uncertainty and doubt lead people to refuse vaccination and is likely to slow down global resilience against the spread and impact of dangerous diseases (World Health Organisation 2019).

Twitter greatly fostered COVID-19 vaccine hesitancy: giving voice to all people, ordinary users, influential political leaders, and other informed stakeholders – such as pseudo-scientific media.

During the COVID-19 pandemic, Twitter contributed to disseminating vaccine-related misinformation and disinformation, leaning on people's emotional perception of reality with fake facts about their composition, side effects, and factual purposes.

The association of social media influence with vaccination hesitancy has been compared in the article 'Social Media Engagement and Influenza Vaccination during the COVID-19 Pandemic (Benisa et al. 2021). The results showed a correlation between social media usage, age, location and influenza vaccination.

Participants using social media were younger, living in high-density agglomerations and had been less likely to be vaccinated against the flu.

The influence of social media on the decision on COVID-19 vaccination is estimated to be like influenza. The conclusion of this study suggests that advancement in health communication campaigns on the influenza vaccine would help convince more social media users to get the COVID-19 vaccine. As the previous article showed, there can be downsides to using social media in the issue of healthcare. Ahorsu et al. (2022) aimed to find a connection between problematic social media use and getting the vaccination. Although the results had not shown any direct correlation between inappropriate media use and vaccine hesitancy, cyberchondria, fear of COVID-19, and COVID-19 risk perception (each or serially) connected both improper social media use and intention to get a COVID-19 vaccine. An extensive survey on 'Social Media Use and Vaccine Hesitancy in the European Union', covering all states of the European Union, by Mascherini and Nivakoski (2022) found that vaccine hesitancy can relate to the use of social media. However, its impact differs based on the reason.

Hesitation due to health concerns is mainly tied to the individual's health status and less to social media use. Apart from that, those, who use social media as their primary source of information, are much more likely to be subject to vaccine hesitancy.

The impact of medical misinformation and disinformation has been portrayed by Basch et al. (2022). Even a few accounts writing misleading information about COVID-19 can cause significant damage to people's beliefs. Belief in conspiracy theories can reduce the uptake of preventative behaviours. Another thing mentioned in this article was speaking about the reach of professional posts on social media with the theme of COVID-19 precautions. According to the study, content from WHO (2022) and other health professionals had far fewer views than popular content creators, even though the latter frequently had the less correct information.
Part of the reason for that is the use of music, humour, dance, and other means of entertainment. Since that is the case, the solution to letting the folk know the correct information would lie in educating influencers. There has been a significant amount of vaccination hesitancy, even among professional healthcare workers, such as doctors and nurses. A study from China by Xin et al. (2021),' The Impact of Social Media Exposure and Interpersonal Discussion on Intention of COVID-19 Vaccination among Nurses', compounded in 2021, showed low intentions of vaccination against coronavirus.
However, according to the study, frequent social media exposure and interpersonal discussion on the vaccination topic did help the decision to get the vaccine.
According to a study carried out by Thelwall et al. (2021) on the impact of Twitter on COVID-19 vaccine hesitancy, out of 4000 randomly selected vaccine-related tweets in English, 446 contained information able to foster vaccine hesitancy: 23.5 percent were conspiracies, 16.1 present expressed worry about the development speed, and 10.5 percent included concern about safety (2021). Several conspiracy theories in tweets argued that vaccines were meant to inject microchips into people to depopulate the world or to help Bill Gates accomplish a secret machination of his, as the famous antivaxxer advocate Robert Kennedy Jr. stated several times (Jemielniak & Krempovych 2021). Such information is classified as "disinformation" as they are implausible from a scientific point of view. The tweets expressing concern about the vaccine development speed claimed that the necessary testing and monitoring phases had been rushed or skipped, making vaccines' safety doubtful.

Despite being more credible than conspiracies, such tweets were conveyors of misinformation, as they would partially depict reality in contrast with the scientific reasons – disseminated by official institutions – regarding the prompter developing time of COVID-19 vaccines.
Finally, the safety of vaccines was also questioned by the ingredients, labelled as unsafe by several Twitter users, and their possible side effects on health, caused – for instance – by genetic modifications (Thelwall et al. 2021). Besides the above-mentioned major topics, tweets

with doubts about vaccine efficacy and usefulness significantly contributed to general vaccine hesitancy (Thelwall et al. 2021). Although 82 percent of the considered tweets were related to political issues, the remaining 18 percent were touching non-political themes, increasing the chances of reaching a wider audience, not necessarily political.

Therefore, making it more likely to influence a higher number of people who were not yet firmly aligned concerning this issue. Examples of non-political themes included music, sport, local events, and other kinds of news (Thelwall et al. 2021).

In conclusion, many studies were conducted about COVID-19 vaccine hesitancy with different outcomes. The results can be divided into two groups - those that say that social media add to vaccine hesitancy and those that dispute that. The only difference is in the studied groups of people. Those with higher education, such as nurses or university students, are taught to think critically.

Therefore, social media is only a means of getting more information and communicating with others for them. However, when looking at the total population, many cannot think critically and believe everything they read.

In that case, the many misleading posts on social media can lead to them not trusting the vaccine and other scientific information about the virus SARS-CoV-2Twitter: Health Stakeholders' Role during the COVID-19 Pandemic

The vaccine-related Twitter posts survey for healthcare providers by Hernandez et al. (2021) conclude that political and non-medical users have posted about 90% of vaccine-related tweets. That means that less than 10% of all tweets stemmed from the medical community. Since many people use social media as their only source of information, the probability of them getting misinformation from uneducated posts is more than high. Healthcare providers and the medical community should start implementing strategies to get the correct information to the mass population via social media.

Apart from political disinformation, retweets of news published by professionally pseudo-medical sources were very common during the COVID-19 pandemic: social bots were very active and had the power to spread anti-vaxxer sentiment and consequently – vaccine hesitancy (Jemielniak & Krempovych 2021).

"Russia Today" – a Russian state-sponsored news website – disseminated plenty of compromising facts about the AstraZeneca vaccine to discredit it in favour of the Russian Sputnik-V. For this purpose, the hashtag #AstraZeneca was utilised for the tweets to reach as many everyday users as possible and for them to be able to share them further. Professional retweeters also had a massive role in propagating anti-vaxxer tendencies (Jemielniak & Krempovych 2021).

The powerfulness of retweets in propagating little ideas can be extrapolated from the hashtags shared before and after March 7th, 2021. In fact, on that day, the Austrian Government made a resolution that suspended AstraZeneca vaccinations as a precautionary provision in response to a few cases of severe side effects arising. Such occurrence shifted Twitter trends, specifically the most retweeted hashtags and linked sources (Jemielniak & Krempovych 2021). If before March 7th, the most retweeted tweets contained links to established Western media such as AFP and Politico.eu, Telegraph and the Guardian, after March 7th. Among the most cited websites in tweets, GreatGameIndia appeared, renowned for disseminating disinformation and conspiracy theories (Jemielniak & Krempovych 2021). Such a shift in retweets and Twitter trends demonstrates how emotional people's perception of news is and how much people crave news reporting shocking and scandalous facts.

Following the devasting effects of the Covid 19, one strategic approach for governments across the globe was to start and roll out limited vaccination among health personnel considered front liners in handling the covid situation. Some health workers felt encroached on their rights with this mandatory approach; however, the uptake of COVID-19 vaccinations from this group was overwhelming (Maneze et al. 2023). The compulsory vaccination and uptake from the health personnel had positive tweets and, without severe effects from the rolled-out vaccines, gave confidence to ordinary Twitter users who were sceptical about the launching of vaccination. This argument is sustained by Maneze et al. (2023), who showed 43% acceptance, 46% neutral and 11% against vaccination rates.

According to Kwok et al. (2021), health workers are vital in providing covid 19 vaccination information that is trusted as these people are deemed as a credible source of information for the public. The tweets from health care can enhance the uptake or hesitancy of vaccinations. The notable attributes influencing the uptake of covid 19 vaccines are effectiveness, duration, and side effects.

According to Miller et al. (2022), public views were significantly posted among the nursing profession. The discussion on Twitter during the early stages of the pandemic was active, and some concerns frequently tweeted included the challenging working conditions of COVID-19. The study by Kwok et al . (2021) revealed that COVID vaccination uptake was strongly linked to those working in hospitals, and stronger intentions for the COVID-19 vaccination were aligned to the younger age with more confidence and collective confidence active actions. Furthermore, a more substantial relationship was found between stress and the intention to get COVID-19 vaccinations.

Chinnasamy et al. (2022) examined the opinions and beliefs of Twitter users on covid vaccines. The sentiment analysis found that most people’s feelings were more positive or neutral on the uptake of vaccines. The results further showed trust in the vaccines; however, some of the population expressed fear due to the virus's continuous spread despite the vaccine's uptake.

Chen et al. (2022) considered the attitudes in Westen European countries, including France, Germany, Belgium, and Luxembourg. The keywords contributing to the opinions and uptake of the vaccines included more than 90% of tweets revealing Pfizer’s efficacy. The finding is further attributed to the EU mass vaccination campaign.

Other critical tweets were Vaccine, biotech, Pfizer, efficacy, virus, centre, and virus. Sixty per cent o of tweets were positive, while 20 per cent were negative towards vaccinations. Additionally, more than sixty per cent of the tweets were in French, more than thirty per cent in Germany, and minimal numbers in Dutch, Spanish, and English.

Benis et al. (2021) indicated an apparent non-compliance during seasonal influenza outbreaks, with 68 per cent not vaccinating. The researchers suggest that for COVID-19 vaccination, it is imperative to engage healthcare social media users to launch more targeted communication and campaigns to influence vaccine uptake by the public.

Lui and Lui (2021) explored the trends in Twitter users' sentiments, and the findings revealed tweets with 42.8 per cent positive, 26.9 per cent neutral and 30.3 per cent negative sentiments. The study identified five topics for positive sentiment tweets: trial results, administration, life, information, and efficacy, while for negative sentiment tweets, including trial results, conspiracy, trust, effectiveness, and administration. The Public attitudes toward COVID-19 were influenced by the time and geographic locations. The rise was recorded when Pfizer reported a 90 efficacy from 1 November 2020 and reached a neutral sentiment by 31 December.

Mascherini and Nivakoski (2022) focused on the role of social media public opinions in the uptake and hesitancy of the vaccine. The results showed the statistical significance of the hesitation of most social media users. The researchers attributed this vaccine hesitancy to health concerns, and some social media users perceived that the COVID risks did not exist, especially for men in good health.

The introduction of the COVID-19 vaccinees influences public opinion to accept or reject it. A combination of sentiment analysis using the (Textblob), algorithm (LinearSVC), and vectorisation methods (TF-IDF) showed best performing model categorisation into positive, neutral, and negative. The findings revealed a reduction in public vaccine hesitancy attributed to the public feeling more comfortable with the vaccines (Qorib et al., 2023).

Sharevski et al. (2022) analysed the perception of COVID-19 vaccines and found that the tweeter content alterations affected the misinformation and exaggerated rumours about mass immunisation which had a backfire effect.

Many of the Twitter users perceived this as accurate information. However, most of their sample refrained from engaging much in the rumours and recommended more label designs to prevent COVID-19 rumours on Twitter.

3.9 Twitter: Political Leaders' Role during the COVID-19 Pandemic

The space given on Twitter to all voices – including the unqualified ones – can consistently impact the public perception and acceptance of the COVID-19 vaccine. Yet, when a publicly renowned figure expresses ideas through Twitter, the influence can have a much broader reach. Donald Trump was the first United States president who supported anti-vaccination positions.

Many of his tweets demagogically linked vaccinations to diagnoses of autism, especially when children were involved, and convinced many of his followers who were exposed to his tweets that vaccines should be a matter of concern (Hornsey et al. 2020). It should be noted that – as pointed out in Hornsey's research – voters with more extremist ideas are more likely to believe in conspiracy theories, which is why Trump's strategy leans on the publishment of thrilling pseudo-scientific slogans that are shaped according to his supporters' conservative mindset. Trump's influence on US citizens' public opinion and behaviour can be extrapolated from the different vaccination rates among republican and democratic voters (Ye 2021). Suppose one compares the vaccination rate of a typical republican county (27.8 percent of people voting for Joe Biden in 2020) with that of a specific democratic county (61.4 of votes for Joe Biden). In that case, it is possible to observe that the gap in vaccination rates constantly grew from January 2021 – August 2021, from 0.8 percent to 14.4 percent in favour of the democratic county (Ye 2021).

3.10 Summary of Literature on COVID-19 Vaccination Rates

Literature has highlighted the influence of Twitter as a virtual platform to share the COVID-19 vaccination with the public, and their opinions may considerably have an effect. COVID-19 vaccinations can be influenced by individuals' preferences, beliefs s and choices. For the tweeter users and the public, the COVID-19 vaccination needed to be safe and effective. On the other hand, the literature indicates that some tweet users created anxiety and misinformation for the public, which could lead to hesitancy.

The individuals who tweet may have an effect, observed by the politicians, for whom the majority may misinform to spearhead their agenda. The healthcare professions are core to the tweeter platform as the levels of confidence from the public opinions and the value given to the information shared has a high level of acceptance as the group is experiencing COVID-19 vaccinations and their effects.

COVID-19 vaccination rates are influenced by proximity due to geographical regions and levels of development of a country. The distribution and access to the vaccines and the approaches of how the information is shared (in some areas with a poor internet connection and expensive internet charges) may affect the information received through Twitter. Overall, Twitter saves as a powerful platform to influence attitudes and perceptions towards promoting the COVID vaccination rate. Text outlining Elon Musk's recent Twitter acquisition.

Chapter Four

Practical Part

4.0 Introduction

In this research, a consistent methodology for doing sentiment analysis on a sizable dataset of tweets about COVID-19 immunization was proposed. Several hashtags were used to gather the data from Twitter. The tweeted text, user ID, region, and time of the tweet are all included in the dataset. The dataset was preprocessed concerning the usage of various redundant words, extraneous punctuation, stop words, and special symbols to make it clean and acceptable for training the machine and deep learning models. Preprocessed data help the models perform better in terms of classification. A lexicon-based method called TextBlob was used for dataset annotation. TextBlob analysis saves time because it takes human specialists a lot longer to categorize the data. After all, there are so many tweets. For the chosen models, the labeled data were divided into training and testing sub-datasets. The training data was used to train machine learning and deep learning models, and various parameters were specified to enhance their performance.

Research Questions

1. How is COVID-19 perceived on twitter?
2. Does Twitter content influence COVID-19 vaccination?
3. What is the general sentiment people towards COVID-19 vaccines?

4.1 Dataset Description

Sentiment analysis was used in this study to examine COVID-19 immunization. The collection contains tweets about COVID-19 immunization for this purpose. The Tweepy library was used in conjunction with the Twitter developer account to retrieve the tweets from Twitter. Several geolocations and specific hashtags, including "#Covid19 #Vaccine," "#Corona #Vaccine," "#covidvaccine," "#coronavaccine," and "covid19 vaccination," were used to filter the tweets. Based on the number of tweets linked to the topic, many nations have been targeted. Several attributes, including users, locations, text, and more, were included in the tweets data. Table 1 provides a selection of the dataset's records. Figure 1 shows the total number of tweets by country as well as the percentage of tweets that were extracted from each nation. There are 208 sites altogether in the collected datasets that indicate the origin of tweets. It is impossible to display all the nations that are mentioned in the tweets. Instead, we only display the nations whose tweets account for at least 2% of all tweets, as shown in Figure 1, which also includes the tweet-producing nations.

The remaining nations are referred to as "Other nations" and include Belgium, New Zealand, Spain, Western Australia, Italy, and numerous other nations.

Country	Proportion Of Tweets
Canada	9.5
India	9.2
United Kingdom	9.3
South Africa	8.6
Pakistan	8.6
United State	8.0
Ireland	6.7
Germany	6.7
Australia	5.7
Other Countries	27.7

Table 2.0 Country wise and proportional tweets

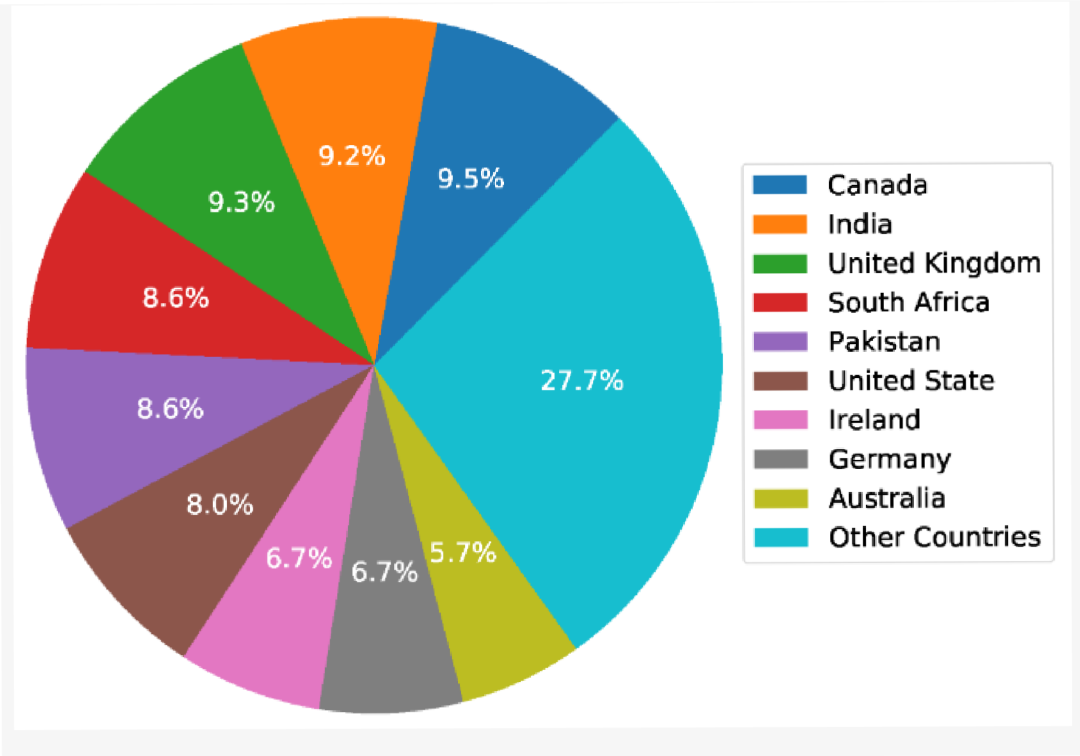

Figure 4 Worldwide twitter dataset:Aijaz Ahmad Reshi

User Name	Location	Tweets
Lunini	Washington DC	As expected WHO celebrates return of # USA to the organization during the surge of the covid # pandemic #COVID19 https://t.co/TbVcBF3Nxr (accessed on: 10 Jauary 2023)
Danschoenmn	St. Paul Park	No federal plan to get the vaccine to our citizens. NONE! Imagine knowing something https://t.co/XU9ADtpNlV (accessed on: 10 January 2023)
Richard Levine	Hawaii, USA	@drdavidsamadi And THAT are how you end a pandemic. And credit will go to the # vaccine. where did we see this https://t.co/X6OnOhbiMs (accessed on: 10 January 2023)
FrancicoCarbral	Lisboa	There is only one way forward: every person on earth will either get the virus or the vaccine. # COVID19 # vaccine

Table 3 Worldwide twitter dataset:Aijaz Ahmad Reshi

4.2 Data Processing

Data preparation follows data extraction to eliminate noise and unimportant information and improve the training of the chosen models.

The data elements in the tweets that are useless for the sentiment analysis process are removed as part of the cleaning procedure. These data elements include stop words, punctuation, hyperlinks, @username, and # symbols. Using the Natural Language Toolkit (NLTK) package, tweets were preprocessed. More than 70 corpora, lexical analysis tools, and several text-processing libraries are all included in NLTK. These machine translation libraries include the crucial and essential NLP operations for tokenization, parsing, and tagging (NLTK Library, 2023). The preprocessing steps are described in the subsections that follow, and several tweet examples are presented both before and after preprocessing to demonstrate the results.

4.2.1 Removal of Username, Hashtags, and Hyperlinks

On Twitter, users typically use the @username format to refer to or tag friends and other related individuals in their tweets. They also frequently use hashtags and hyperlinks in their tweets. Usernames, hashtags, and hyperlinks were taken out of the tweets because they are useless for sentiment analysis. The sample tweet text is shown in Table 2 both before and after preprocessing.

Tweet Before Removal	Tweet After Removal
Many thyroid and autoimmune patients are wondering whether they should get the COVID vaccine. Thyroid Expert Esther	Many thyroid autoimmune patients wondering whether get COVID vaccine. Thyroid Expert Esther
As expected, celebrates the return of the organization during the surge of the covid	expected celebrates return organization surge covid

Table 4 Covid-19 Sentiments Anaysis: Reshi AA

4.3 TextBlob

A well-lexicon-based technique called TextBlob is used to carry out different natural language processing (NLP) operations on unprocessed text (Loria). The TextBlob Algorithm 1 implementation is used to process text data using a Python library called TextBlob as a programming interface. One can, for instance, assess text attitudes, extract noun phrases, build POS tags, translate, classify, and more using TextBlob (Vijayarani & Janani, 2016). In a word, the TextBlob library has several built-in features that help with language processing. It can be used in a variety of languages, including Spanish, English, Arabic, and others. Natural language processing (NLP) operations can be carried out on raw text using the well-known lexicon-based approach TextBlob (Loria). The TextBlob Algorithm 1 implementation is used in conjunction with a Python library called TextBlob as a programming interface to process text data. One can, for instance, use TextBlob to translate, classify, build POS tags, extract noun phrases, and assess sentiments in the text (Vijayarani & Janani, 2016). To put it simply, the TextBlob library has a variety of built-in features that help with language processing. Arabic, Spanish, English, and other languages can all be used with it. The emotion score range for TextBlob is shown in Table 6, where scores less than 0 denote sentiments with negative polarity and scores greater than +1.0 denote sentiments with positive polarity. In terms of subjectivity, scores below 0.0 denote sentiments based on facts, while scores above 1.0 denote sentiments based on individual ideas.

Algorithm 1 TextBlob algorithm for sentiment analysis.

Input: Input: Worldwide COVID-19 Vaccination Tweets
Result: Polarity Score >> 0 ⟶ (Positive)
Polarity Score == 0 ⟶ (Neutral)
Polarity Score << 0 ⟶ (Negative) initialization **loop** (each tweet in tweets)
Compute Polarity Score TextBlob (tweet)
condition:
if (Polarity Score > 0) **then**
Tweet Sentiment = Positive;
elseif (Polarity Score = 0) **then**
Tweet Sentiment = Neutral;
else
Tweet Sentiment = Negative;
condition end
loop end

Sentiment	Score
Negative	Polarity score << 0
Positive	Polarity score = 0
Neutral	Polarity score >> 0

Table 5 Covid-19 Sentiment Analysis: Aijaz Reshi

.

4.3.1 AFINN

The Affective Norms for English Words Lexicon (ANEW), created by Nielsen, F.A. (Nielsen, 2017), served as the foundation for the sentiment lexicon known as AFINN. Similar to

VADER, it uses a variety of terms from the English language, each with a different sentiment score. The AFINN lexicon was created to fill the void left by ANEW's lack of slang words in comparison to VADER. It uses a manually compiled lexicon and a rule-based methodology. AFINN is easier to use, less difficult, and requires fewer calculations. For each language, the AFINN valence scores range from -5 to +5. Negative attitudes are expressed below 5, whereas positive sentiments are indicated with a score over +5 (Nielsen F. , 2011).

By examination of the types of textual information used on micro-blogging networks, the AFINN lexicon was constructed. Particularly for Twitter, the words on the list were increased as a result of the collection of posts from users that were judged to have strong sentiments. A lot of people also used the Urban Dictionary, which contains all types of current abbreviations like LOL and ROFL. The list of positive and negative adjectives for each category of data must be used to determine the opinion orientation for the given data.

4.4 Machine Learning Approach used for the experiment

Machine learning is a subfield of artificial intelligence that involves using algorithms to allow computer systems to learn from data without being explicitly programmed. There are many different machine learning approaches, but some common ones include:

1. Supervised learning: This is a type of machine learning where the algorithm is trained on a labeled dataset, meaning that each data point is associated with a label or outcome. The algorithm then learns to map input data to output labels based on the patterns in the training data. Some common supervised learning algorithms include linear regression, decision trees, and neural networks.

2. Unsupervised learning: This is a type of machine learning where the algorithm is trained on an unlabeled dataset, meaning that there are no pre-existing labels or outcomes associated with the data. The algorithm then tries to identify patterns or clusters in the data based on similarities or differences between data points. Some common unsupervised learning algorithms include k-means clustering, principal component analysis, and self-organizing maps.

3. Reinforcement learning: This is a type of machine learning where the algorithm learns through trial and error by interacting with an environment and receiving rewards or punishments based on its actions. The algorithm then learns to take actions that maximize its rewards over time. Reinforcement learning is commonly used in robotics,

gaming, and other applications where an agent needs to learn to make decisions in a dynamic environment.

4.4.1 Term Frequency-Inverted Document Frequency Features

The method of extracting features known as TF-IDF is popular. This method is frequently used in text analysis and the correction of musical information (Yu, 2008). It assigns weight to the terms in a particular document based on the inverse relationship between the document's frequency and the frequency of the words (F Rustam, 2019). Terms with higher weighted scores are considered to be more significant [41]. What TF-IDF is best described as,

$$tfidf = tft, d * log \frac{N}{di, t},$$

represents the frequency of the term t in document d, N represents the total number of documents, and di, t represents the total number of documents that contain the phrase t.

Algorithm 3 AFINN algorithm for sentiment analysis.

Input: Input: Worldwide Covid19 Vaccination Tweets
Result: Polarity Score >> 0 ⟶ (Positive)
Polarity Score == 0 ⟶ (Neutral)
Polarity Score << 0 ⟶ (Negative) initialization **loop** (each tweet in tweets)
Compute Polarity Score AFINN (tweet)
condition:
if (Polarity Score > 0) **then** (F Rustam, 2019)
Tweet Sentiment = Positive;
else if (Polarity Score = 0) **then**
Tweet Sentiment = Neutral;
else
Tweet Sentiment = Negative;
condition end
loop end

4.4.2 Decision Tree

A machine learning model called DT is applied to classification and regression issues (Brijain, Patel, Kushik, & Rana, 2014). Once the splits become atomic, the model constantly divides the dataset into n subgroups using the binary technique. When a data subset cannot be further subdivided, it is said to be atomic in this context. The dataset is divided into an incremental technique for construction and a decision tree is then used, with many of its branches having varying sizes. The DT was utilized with a max depth hyper-parameter in this study to reduce complexity and combat model over-fitting.

4.5 Deep Learning Models for Sentiment Analysis

This study also made use of four distinct deep learning models to examine how well the deep learning models performed on the COVID-19 vaccine. Two other ensemble models are also suggested. To achieve greater levels of performance for the task at hand, the customized architecture of the convolutional neural network (CNN), long short-term memory (LSTM), recurrent neural networks (RNN), and gated recurrent unit (GRU) is built. Two ensemble models, CNN-LSTM (ensemble of two models) and LSTM-GRNN are proposed in addition to individual models (ensemble of three models). The architecture of these models, which are deployed using the TensorFlow framework, is displayed in Table 9. The categorical cross-entropy loss function was used to build these models, and the "Adam" optimizer was employed for optimization. The models underwent training using 128 batches of 100 epochs each. To attain considerable performance, LSTM, GRU, and RNN were layered to create the suggested ensemble LSTM-GRNN.

LSTM	CNN	RNN
Embedding (5000, 200) Dropout (0.2) LSTM (100) Dropout (0.2) Dense (3, activation = ‘softmax’)	Embedding (5000, 200) Dropout (0.2) Conv1D (128, 4, activation = ‘relu’) MaxPooling1D (pool_size = 4) Flatten () Dense (32) Dense (2, activation = ‘softmax’)	Embedding (5000, 200) Dropout (0.2) SimpleRNN (32) Dense (3, activation = ‘softmax’)

LSTM	CNN	RNN
GRU	**CNN-LSTM**	**LSTM-GRNN**
Embedding (5000, 200) Dropout (0.2) GRU (100) Dropout (0.2) Dense (3, activation = 'softmax')	Embedding (5000, 200) Dropout (0.2) Conv1D (128, 4, activation = 'relu') MaxPooling1D (pool_size = 4) LSTM (128) Dense (32) Dense (3, activation = 'softmax')	Embedding (5000, 200) Dropout (0.2) LSTM (100) Dropout (0.2) GRU (100) SimpleRNN (32) Dense (3, activation = 'softmax')
loss = 'categorical_crossentropy', optimizer = 'adam', epochs = 100		

Table 6 Covid Vaccine Twitter Dataset: Hessa Alsuwailem

4.6 Architecture of Proposed LSTM-GRNN

The stacked LSTM, GRU, and RNN networks are used in the proposed ensemble LSTM-GRNN to increase sentiment analysis accuracy. Seven layers make up the LSTM-GRNN, as shown in Table 9. It has a dense layer, an embedding layer, two dropout layers, LSTM, GRU, and RNN layers, as well as one layer for each. A vocabulary size of 5000 words and an output size of 300 are employed with the embedding layer. A dropout layer comes after the embedding layer, as seen in Figure 2. The dropout rate for this layer is 0.2, and it is used to lessen the model's complexity and the likelihood that it would overfit. 100 LSTM units make up the LSTM layer, which is on top of the stack. After the LSTM layer by 100 units is the GRU layer. The RNN layer, which has 32 units, lies at the bottom of the stack and is followed by a dropout rate of 0.20. In the end, the target classes were obtained using a dense layer with three neurons and a softmax activation function. With 100 fitting epochs, a categorical cross-entropy loss function, and the "Adam" optimizer, the LSTM-GRNN was constructed.

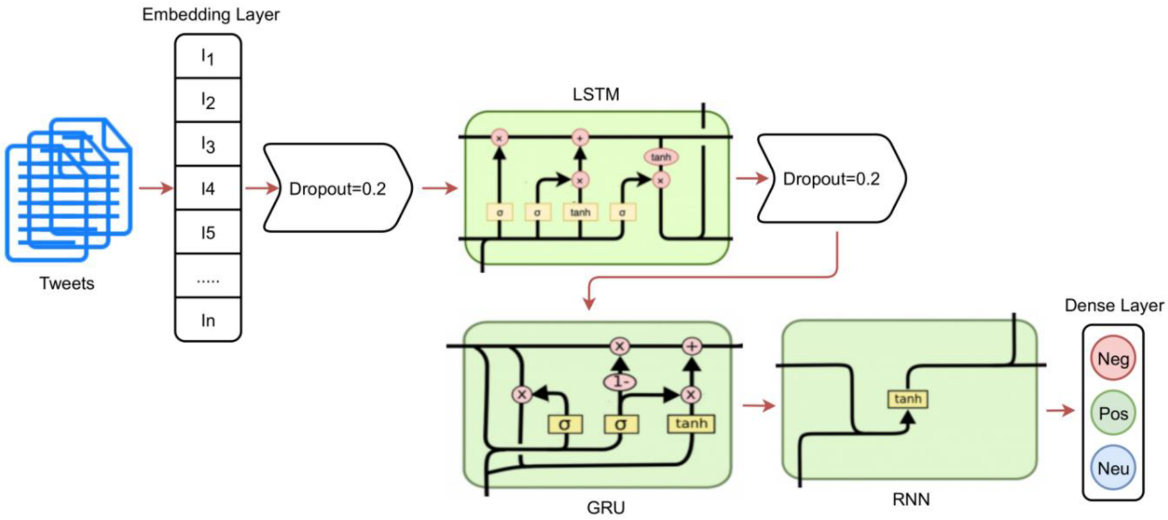

Figure 5 Covid-19 Twitter Sentiment: Imran Ashraf.

4.7 Lexicon-Based Approach for Sentiment Analysis

Preprocessing methods clean up the dataset, which can lead to improved findings. Three lexicon-based techniques were used to evaluate the dataset after preprocessing to identify the sentiments. Three feelings are outputted for each tweet using these lexicon-based techniques. In this investigation, three lexicon-based methodologies—TextBlob, AFINN, and VADER—were applied. These methods produce polarity scores to ascertain the sentiment. Also, a thorough analysis of each country's tweets was done to discover how people felt about the COVID-19 vaccine and what worries them. The phases and their order used in this investigation are shown in Figure 3 along with the architecture of the lexical methods.

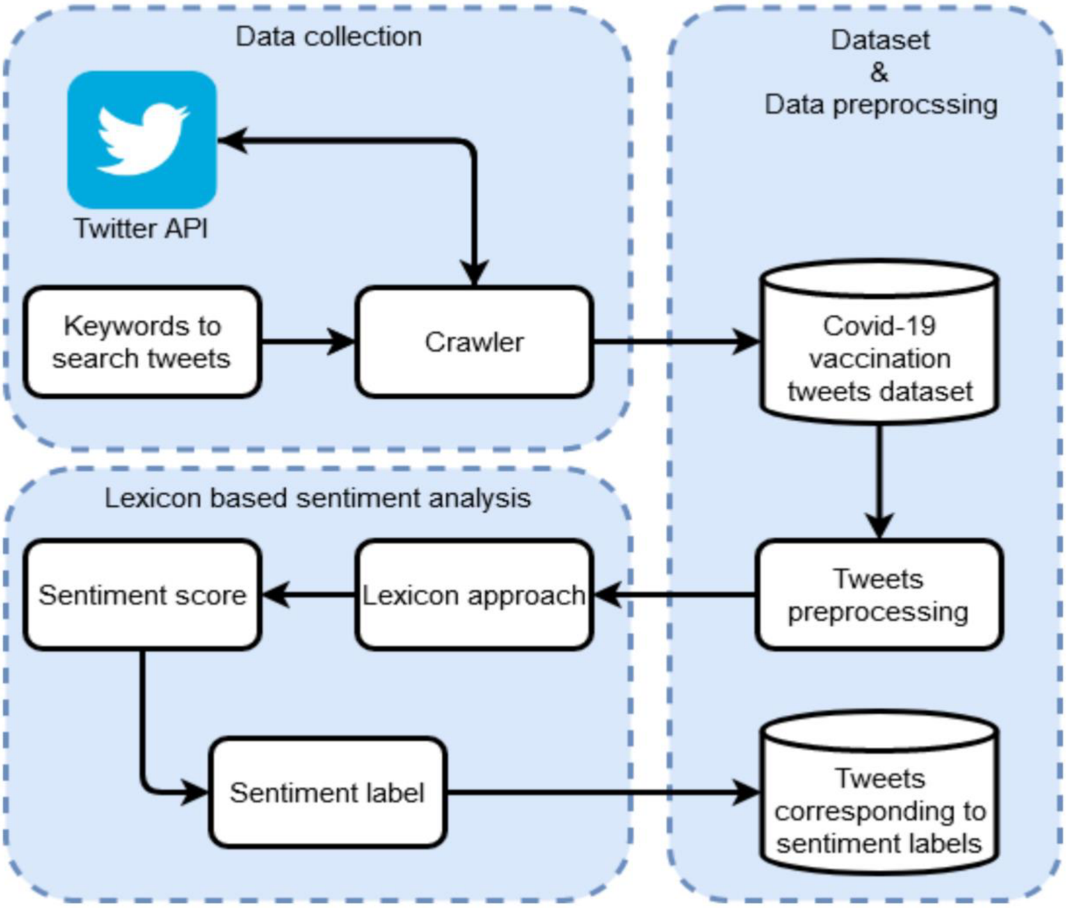

Figure 6 Covid Worldwide Twitter Dataset: Shabana Sharfi .

4.8 Proposed Methodology for Sentiment Analysis

The labeled tweets dataset was used for the training of the machine learning models after the sentiment was determined using lexicon-based methods coupled with tweet labeling. The attitudes were categorized as good, negative, or neutral using the trained models. The labeled dataset was used to extract TF-IDF features before being divided into training and testing portions into 80:20 ratios. To evaluate the effectiveness of both lexical and machine learning methods, various combinations of lexicon-based sentiment analysis methodologies were evaluated. For instance, the tests were run utilizing the high-performing lexicon technique and the three sentiments (TextBlob, AFINN, and VADER) as target classes using the chosen machine learning models. The effectiveness of the chosen machine learning models RF, LR, and DT was also evaluated in terms of accuracy, precision, recall, and F1 score. The architecture of the suggested sentiment analysis methodology is depicted in Figure 4.

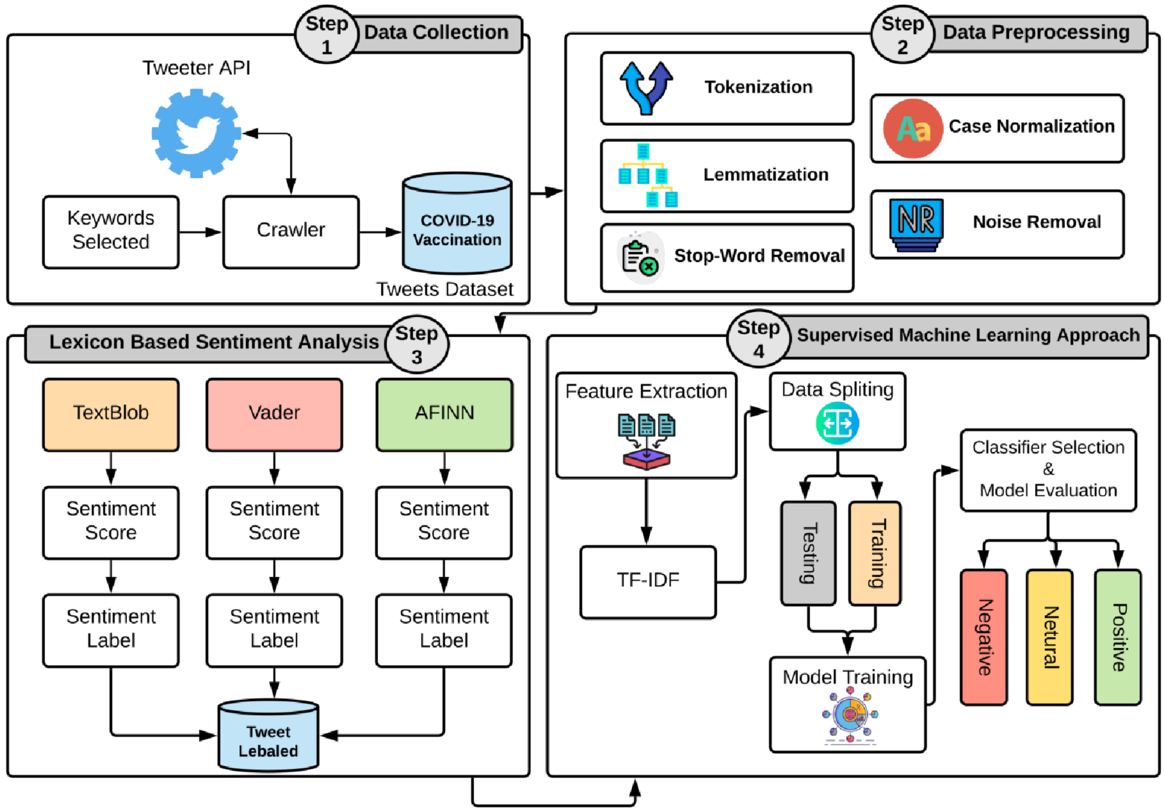

Figure 7 Twitter Analysis Sentiment: Ziyad Alrabiah .

Chapter Five

Results and Discussions

5.0 Overview

This study uses two methodologies, namely machine learning classification models and NLP-based lexicon methods, to ascertain people's opinions on the ongoing vaccination debate around the world. Scikit-learn was utilized to create machine learning models, and TensorFlow was employed for deep learning models. To ascertain the sentiments regarding the chosen topic, words and sentences can be chosen and analyzed. Positive, negative, or neutral emotions can be assigned to a sentiment. The TextBlob, AFINN, and VADER NLP lexicon-based techniques, as well as the RF, LR, and DT model-based machine learning models, were all implemented. The purpose of the discussions that follow is to show and evaluate how well lexical and machine-learning techniques for sentiment analysis perform. Experiments were performed using the Intel Core i7 7th generation machine and the Windows 11 operating system. Python language was used to implement the script.

Figures 5–7 show the uni-gram, bi-gram, and tri-gram distributions of the dataset of the COVID-19 vaccination. The uni-gram and bi-gram graphs show that the most commonly used words were 'covid', 'vaccine', and 'covid', while the tri-gram shows that the highly discussed topics were the COVID-19 vaccination campaign, COVID-19 vaccination for health workers, vaccination side-effects, receiving the first dose, and so forth.

Figure 8 shows the word cloud of the dataset containing the perceptions and opinions of people around the globe regarding the ongoing COVID-19 vaccination. Similar to the uni-gram, bi-gram, and tri-gram terms, the world cloud illustrates that 'COVID', 'taking vaccination', and 'vaccination drive', and so forth are the most commonly used words in the tweets.

5.1 POS Dataset

The POS-tagged targeted words are listed in Table 10 along with the appropriate word count from the gathered dataset. It has various words and the POS tags that go with them. For instance, nouns (NN) include the word count as well as a subset of the nouns used in the tweets' text. The adjective terms discovered in the dataset are represented by an adjective (JJ), and the accompanying columns list their total occurrences.

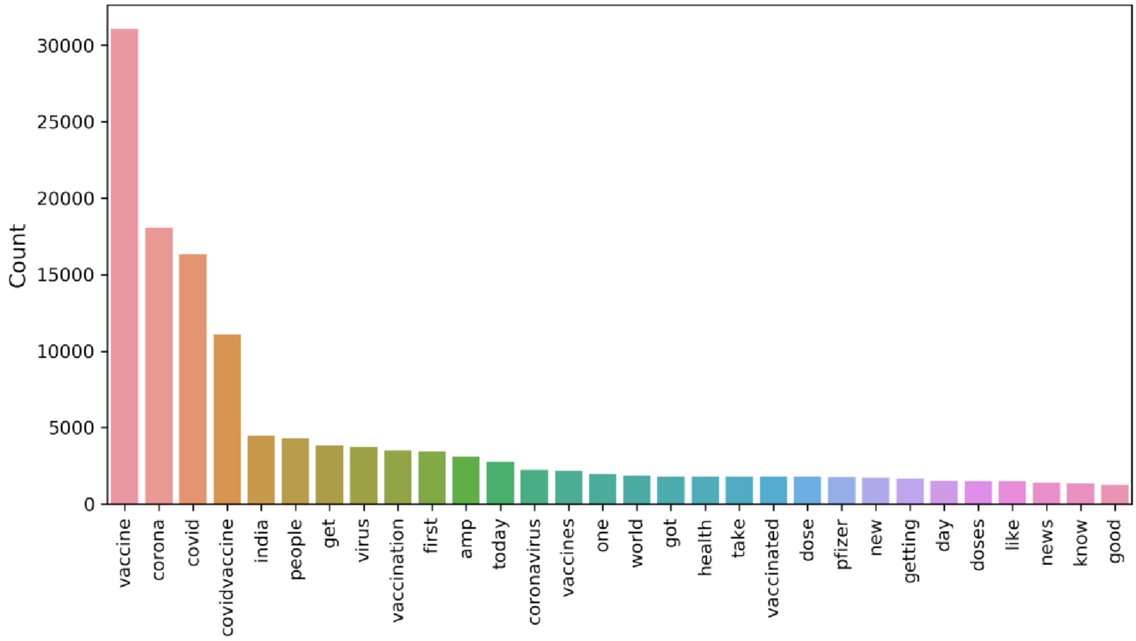

Figure 8 Tweets Classification on the base of sentiments for US Airline Companies:Furqan Rustam

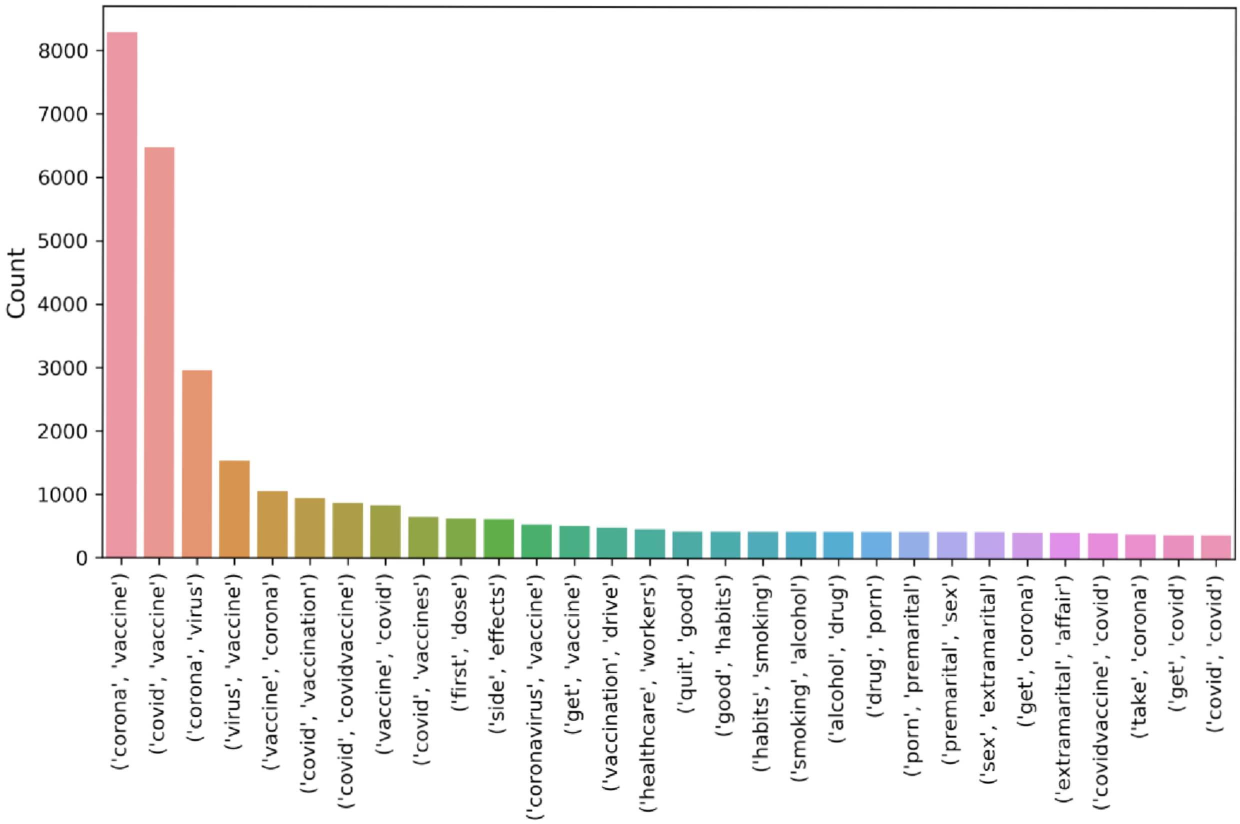

Figure 9 Chart for Dataset Collected

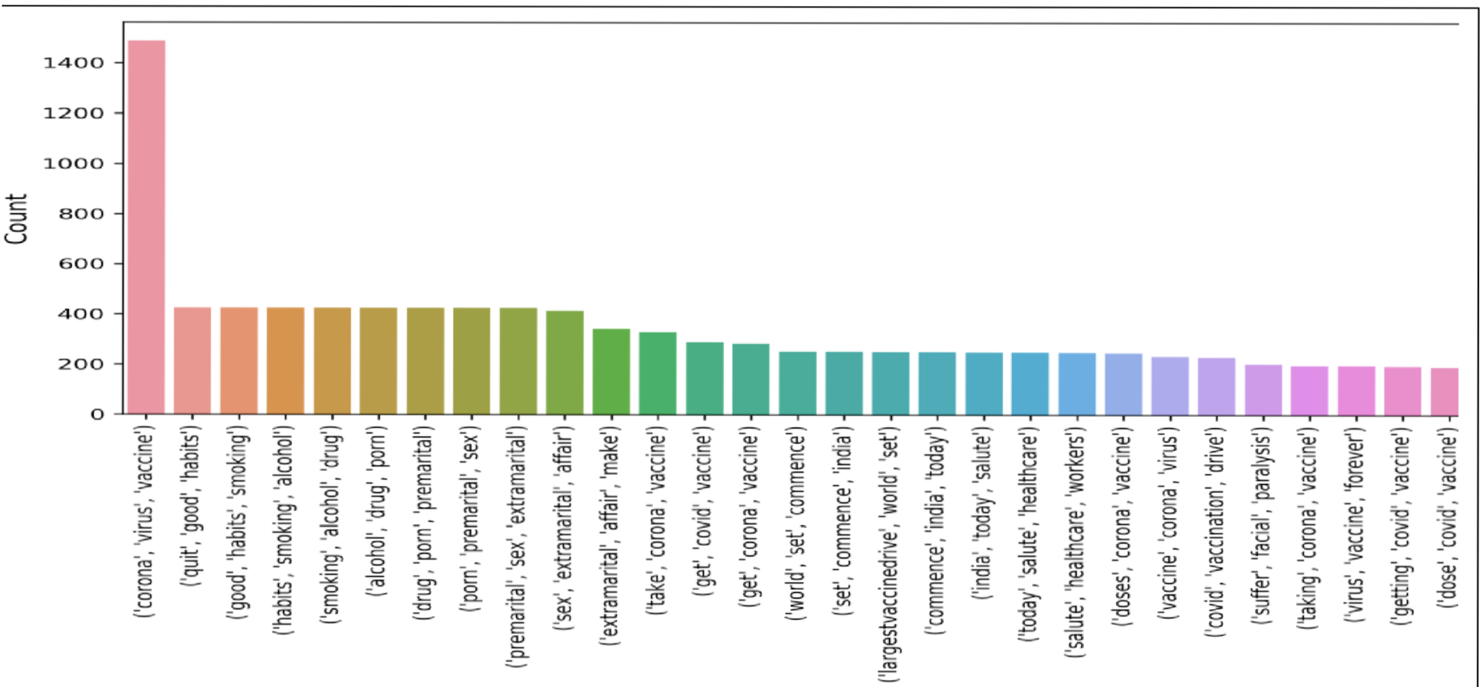

Figure 10 Chart for Dataset Collected

NN	Count	JJ	Count	Entity Name	Entity Type	Count
Vaccine	30,209	Corona	4483	India	GPE	3033
Virus	3540	Good	1262	Today	DATE	1787
India	2686	Dose	1080	First	ORDINAL	1557
World	1879	Many	1052	China	GPE	635
Health	1791	Great	894	Million	CARDINAL	503
Pfizer	1587	Free	789	Pakistan	GPE	473
Country	1525	Safe	739	Pfizer	ORG	428
Worker	1405	Pandemic	665	Healthcare	ORG	413
News	1403	Medical	608	Norway	GPE	404
Government	991	Premarital	425	Chinese	NORP	288

Table 7 POS tagging.

5.2 Sentiment Analysis Using TextBlob

Each lexical technique underwent its own set of experiments. The emotion polarity score using the TextBlob approach is displayed in Figure 9. It demonstrates that more tweets with a favorable polarity score are more numerous. The positive sentiment score for each country's tweets individually and for all tweets collectively ranges from 0 to 0.3, showing that while tweets are generally deemed to be positive, their average polarity score is low. Because a larger percentage of positive tweets have a polarity score between 0.1 and 0.we can conclude that the rweets are positive but less intensely so.

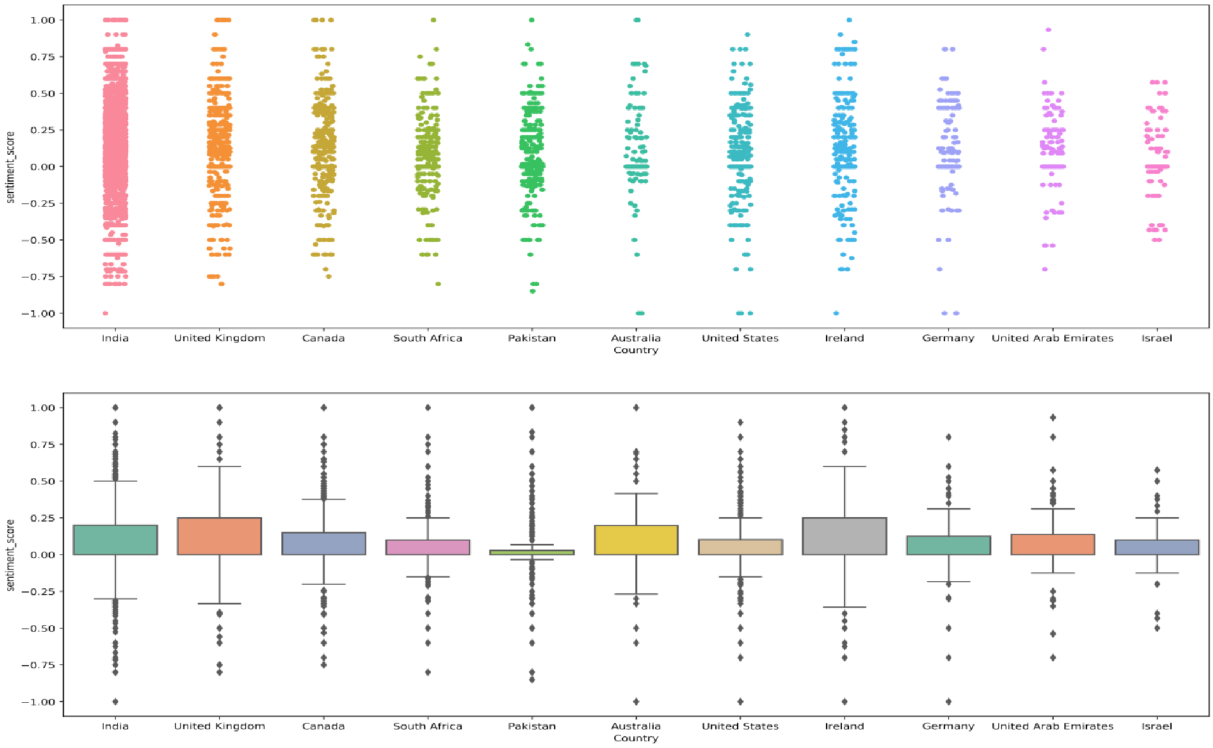

Figure 11 Sentiment analysis using TextBlob

The findings of the sentiment analysis by nation using the TextBlob lexicon approach are presented in figure 11 along with a total for all the countries. Positive, negative, and neutral attitudes in tweets were distinguished from one other. The percentage of each of the three sentiment categories for each country is displayed in each column. Findings show that, when taking into account tweets from all the countries combined, the bulk of them fall into the neutral category, followed by the positive tweets. There are 48.81% neutral tweets, 38.33% positive tweets, and 12.86% negative tweets, respectively.

Country	Positive	Negative	Neutral
All Countries	38.33	12.86	48.81
India	37.74	10.66	51.60
United Kingdom	43.62	13.72	42.66
Canada	35.31	14.36	50.33
South Africa	30.31	11.32	58.36
Pakistan	29.18	14.23	56.58
United State	29.18	14.23	56.58
Ireland	41.14	13.90	44.96

Country	Positive	Negative	Neutral
Germany	33.63	9.87	56.50
UAE	34.72	8.81	56.48
Israel	26.45	17.42	56.13
Australia	37.41	16.33	46.253
Other Countries	38.92	13.25	47.83

Table 8 TextBlob sentiment statistics for each country and worldwide.

5.3 Sentiment Analysis Using VADER

Figure 10 shows the polarity score given by the VADER approach. The displayed results indicate that VADER-assigned negative polarity scores are higher as compared to TextBlob. TextBlob gives 12% negative tweets, while VADER assigns a negative polarity score of 22% indicating 10% higher negative tweets than TextBlob.

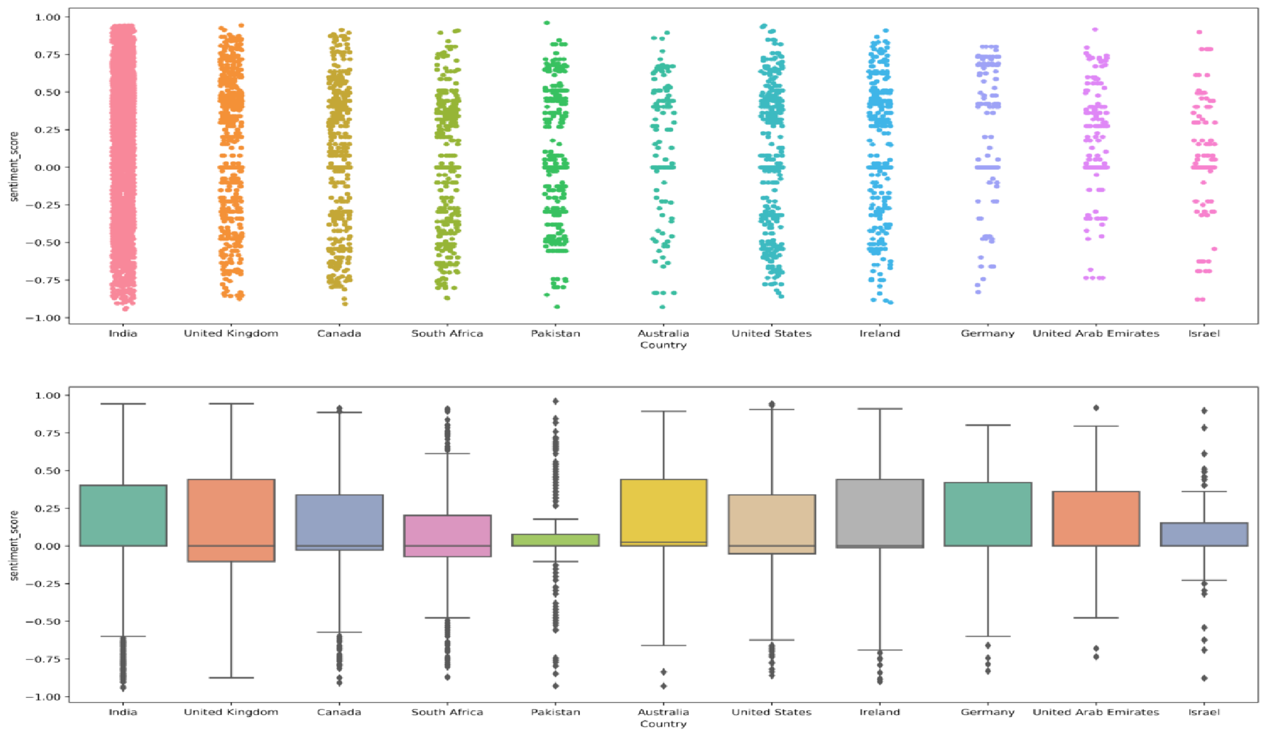

Figure 12 Sentiment Analysis using VADER

The findings of the VADER lexicon-based country-by-country sentiment analysis are presented in Figure 11 and Table 12, together with a total for all the countries. When

The findings of the VADER lexicon-based country-by-country sentiment analysis are presented in Figure 11 and Table 12, together with a total for all the countries. When compared to TextBlob, the ratio of favorable, negative, and neutral tweets altered. The ratio of good tweets hasn't changed much, but the ratio of neutral and negative tweets has undergone a significant shift. For VADER, for instance, the proportion of neutral tweets decreased from 48.81% to 37.74% while the proportion of negative tweets increased to 22.31% from 12.86%. It suggests that a significant portion of TextBlob tweets with neutral feelings were categorized as negative when VADER was applied.

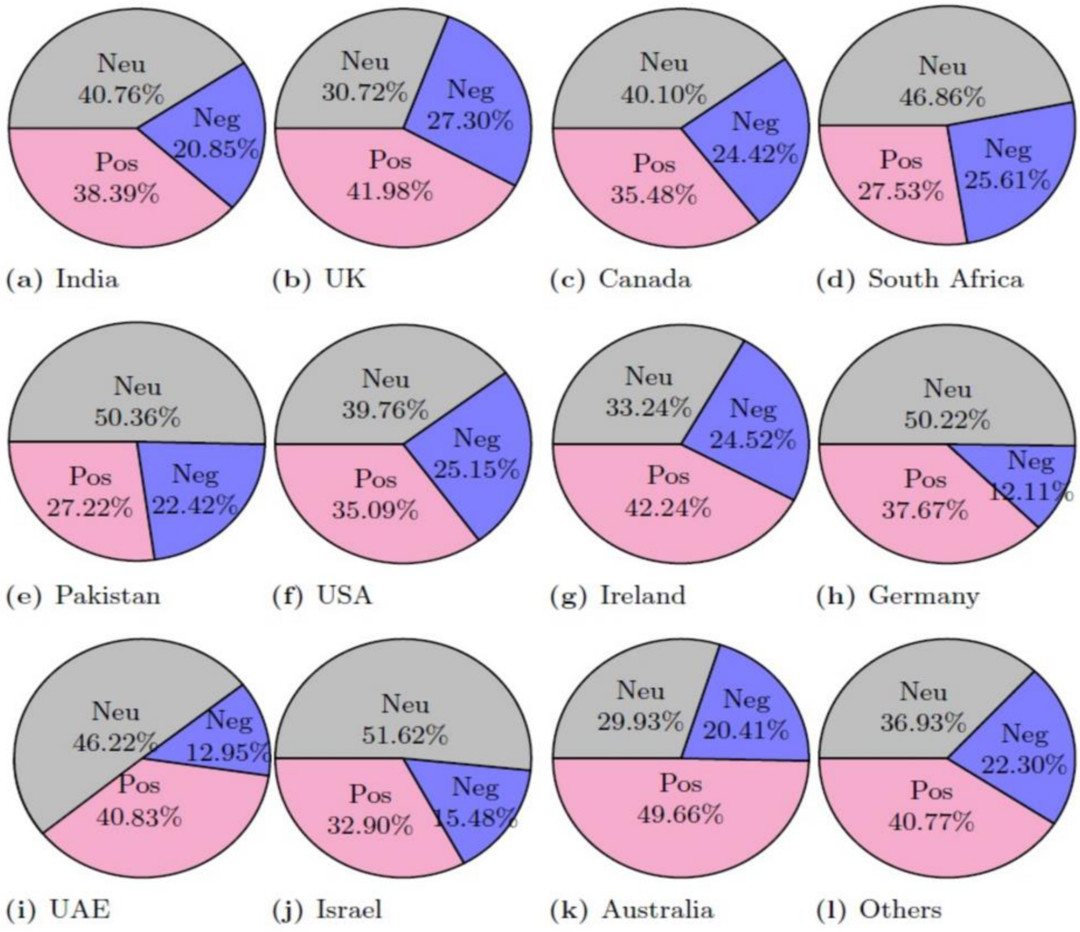

Figure 13 VADER sentiment statistics for each country and worldwide.

Country	Positive (%)	Negative (%)	Neutral (%)
All Countries	39.95	22.31	37.74
India	38.39	20.85	40.75
United Kingdom	41.97	27.30	30.73
Canada	35.48	24.42	40.10
South Africa	27.53	25.61	46.86
Pakistan	27.22	22.42	50.36
United State	35.09	25.15	39.76
Ireland	42.23	24.52	33.24
Germany	37.67	12.11	50.22
UAE	47.67	12.95	39.39
Israel	32.90	15.48	51.61
Australia	49.66	20.41	29.93
Others Countries	40.77	22.30	36.93

5.4 Sentiment Analysis Using AFINN

Figure 12 displays the sentiment analysis findings on the gathered dataset using the AFINN algorithm. According to the results, AFINN assigns a lower score for negative polarity than TextBlob, much like VADER. Both country-specific and aggregate tweets have polarity scores that range from 0 to 2. Tweets from nations including Israel, Germany, Australia, and Pakistan had more negative sentiments

5.5 Comparison of findings

A performance comparison with a number of other research was done in order to illustrate the suggested LSTM-GRNN model's noteworthy performance in comparison to other studies. To do this, the models put forth in a few research were applied using the gathered dataset, and the outcomes were compared to the outcomes of this study. While the work in (Furqan Rustam, 2019) used LR-SGDC (stochastic gradient descent classifier) for US airline sentiments, the study in (Furqan Rustam, 2019) offered an ensemble model for sentiment classification. Similar to this, the study in (Rustam, 2021) carried out the same job using an extra tree classifier (ETC). Except from that, Sarcasm identification was carried out in the study in (Ramish Jamil, 2021) using the CNN-LSTM model, and sentiment analysis was done in the study in

(Rupapara, 2021) using the stacked Bi-LSTM model. These models were applied using the COVID-19 vaccine tweets dataset that was gathered for this study in order to conduct a fair comparison. The TextBlob annotated dataset was used for training and testing, and Table 18 provides a performance comparison. According to the findings, the proposed method's accuracy is noticeably higher than that of other studies. The suggested LSTM-GRNN with a TextBlob-annotated dataset demonstrated greater performance and attained 95% accuracy for sentiments, which is higher than earlier studies, despite utilizing ensemble models in other studies.

Ref.	Year	Model	Accuracy (%)
[39]	2019	LR-SGDC	90
[53]	2021	ET + FU	91
[54]	2021	CNN-LSTM	88
[55]	2021	Stacked Bi-LSTM	93
This study	2023	LSTM-GRNN	95

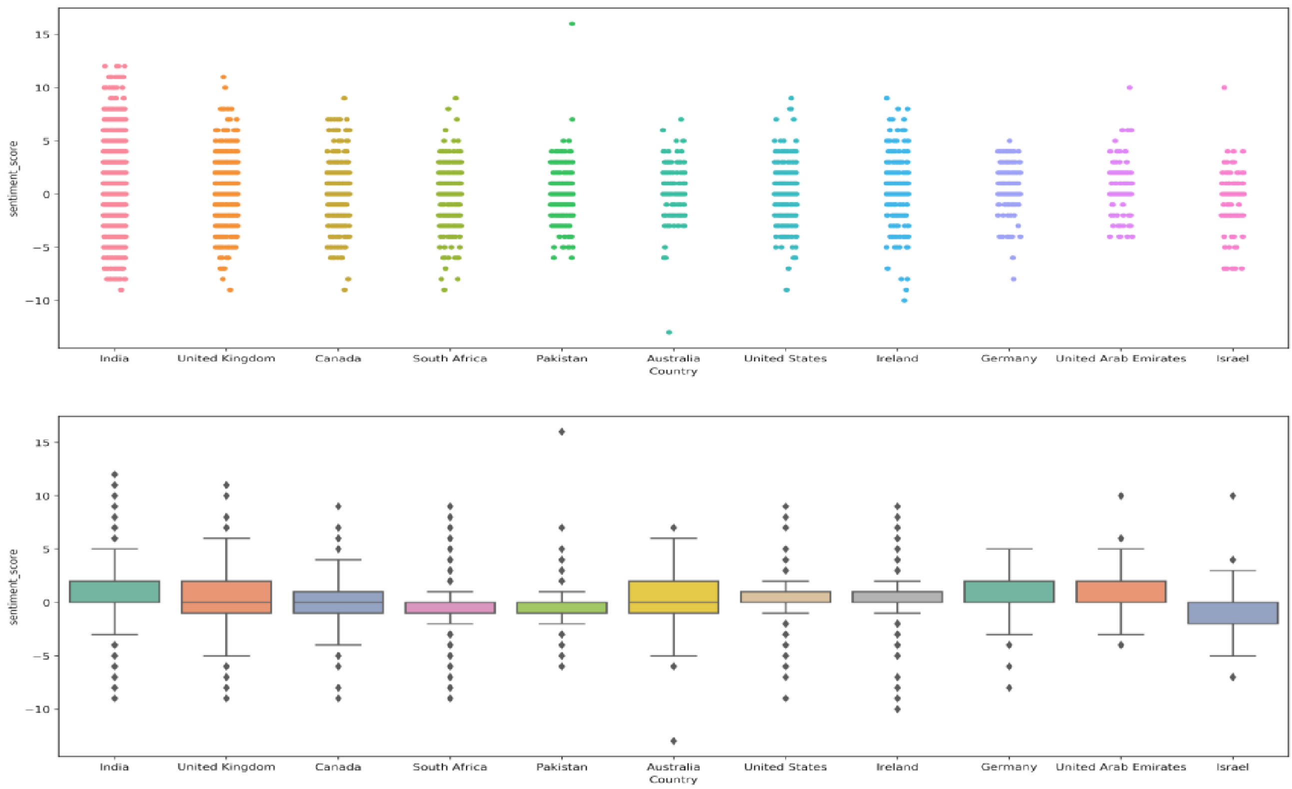

Figure 14 AFINN sentiment score for different countries.

The findings of country-by-country sentiment analysis and the total number of positive, negative, and neutral tweets using the AFINN lexicon technique are displayed in Figure 13 and Table 13 respectively. The percentage of positive tweets, at 23.77%, is higher than that of both TextBlob and VADER, according to the results, while the ratio of neutral tweets is comparable to that of TextBlob. On the other hand, the proportions of supportive and neutral tweets are 35% and 41.21%, respectively.

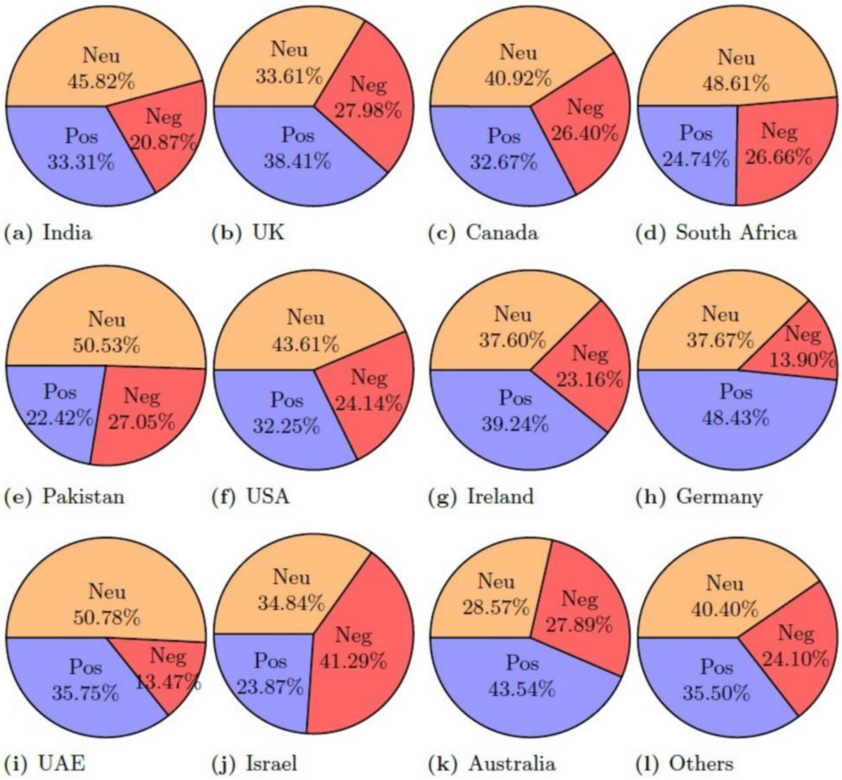

Figure 15 Percentage of the sentiment analysis for AFINN.

Country	Positive (%)	Negative (%)	Neutral (%)
Total	35.01	23.78	41.22
India	33.31	20.87	45.83
United Kingdom	38.41	27.98	33.61
Canada	32.67	26.40	40.92
South Africa	24.74	26.65	48.61
Pakistan	22.42	27.05	50.53
United State	32.25	24.14	43.61
Ireland	39.24	23.16	37.60

Country	Positive (%)	Negative (%)	Neutral (%)
Germany	48.43	13.90	37.67
UAE	35.75	13.47	50.78
Israel	23.87	41.29	34.84
Australia	43.54	27.89	28.57
Other Country	35.50	24.10	40.40

Table 9 AFFIN sentiment statistics for each country and worldwide.

5.6 Sentiment Analysis Using Machine Learning Models

Besides using the lexicon-based methods, this study used several machine learning models for sentiment This study used several machine learning models for sentiment analysis on the annotated dataset connected to tweets about COVID-19 immunization in addition to lexicon-based techniques. When determining whether a tweet should be categorized as positive, negative, or neutral, lexicon-based algorithms are used to get the sentiment score. The generated dataset served as the basis for both the machine learning classifiers' training and testing. For the various datasets that had been annotated with TextBlob, AFINN, and VADER, all three models were trained and put to the test. Using accuracy, precision, recall, and F1 score, the performance was assessed. Each ML model's performance was assessed using one of the three different lexicon techniques.

The accuracy and other performance evaluation parameters for the machine learning models are shown in Table 14 as results. To conduct experiments, the machine learning models were given the dataset that had been annotated using TextBlob. According to the results, RF and LR each achieved the maximum level of accuracy, 93%, surpassing DT's accuracy of 92%.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
DT	92	93	87	90
RF	93	96	92	94
LR	93	94	87	89

Table 10 Performance results for machine learning models using TextBlob sentiments.

Experiment were conducted with both a VADER-labeled dataset and the TextBlob annotated dataset. The effectiveness of the machine learning models when applied to a VADER annotated dataset is shown in Table 15. Findings show that both RF and LR perform equally well on the VADER sentiment analysis dataset, with a 90% accuracy rate. However among the three models, RF performs best in terms of precision, recall, and F1 score measures. Also, when the dataset was switched from TextBlob to annotated VADER, the models' performance suffered significantly. For instance, both LR and RF accuracy was reduced to 90% from 93%, while DT experienced a substantial reduction to 82% from 92% when trained with a VADER-annotated dataset.

5.7 Experimental Results of Deep Learning Models

With accuracy, precision, recall, and F1 score serving as the evaluation criteria, tests for deep learning models are carried out using the TextBlob dataset. This is done in consideration of the higher sentiment classification using the TextBlob annotated data. Table 17 presents the experimental outcomes. According to the results, the suggested LSTM-GRNN performs better than all other models across the board. It outperformed the deep learning and machine learning models employed in this study, achieving a 95% accuracy rate, which is the highest. Together with accuracy, additional metrics including precision, recall, and F1 scores were superior to those of competing models. With a 93% accuracy rate, GRU also outperformed the LSTM and RNN, which both had a 92% accuracy rate. Comparing the CNN model to recurrent models, the CNN model performed poorly.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LSTM	92	90	90	90
GRU	93	93	93	92
RNN	92	92	92	92
CNN	87	87	87	87
CNN-LSTM	88	88	88	88
LSTM-GRNN	95	95	95	95

Table 11 Deep learning model performance on TextBlob sentiment.

Discusion of Similar Research

M. Abbas et al, presented a paper on "Sentiment Analysis of Twitter Data Related to COVID-19 Vaccination in the UK". This study analyzed the sentiment of Twitter users in the UK regarding Covid-19 vaccination. The authors used machine learning techniques to classify tweets as positive, negative, or neutral and found that the sentiment was mostly positive.
The paper "Sentiment Analysis of Twitter Data Related to COVID-19 Vaccination in the UK" by M. Abbas et al. aimed to analyze the sentiment of Twitter users in the UK regarding Covid-19 vaccination. The authors collected tweets related to Covid-19 vaccination in the UK from December 2020 to January 2021 and used machine learning techniques to classify them as positive, negative, or neutral.

The study found that the sentiment of the tweets was mostly positive, with 47% of the tweets classified as positive, 23% as negative, and 30% as neutral. The positive tweets expressed gratitude, relief, and hope, while the negative tweets expressed concerns about the safety and efficacy of the vaccine, and skepticism about the government's handling of the pandemic. The neutral tweets were mostly informational or promotional in nature.

The authors also analyzed the sentiment of tweets related to different Covid-19 vaccines and found that the sentiment was more positive for the Pfizer-BioNTech vaccine than the AstraZeneca vaccine. Additionally, the sentiment was more positive among female Twitter users and those aged between 25 and 34.

The study highlights the potential of social media data, particularly Twitter, to provide insights into public sentiment regarding Covid-19 vaccination. The findings suggest that the sentiment in the UK regarding Covid-19 vaccination was mostly positive, with some concerns and skepticism expressed. The study could inform public health communication strategies to address these concerns and promote vaccine uptake. However, the study has limitations, such as the potential for bias in the Twitter data and the subjective nature of sentiment analysis.

In a related paper S. P. Mehta et al conducted a research on "Sentiment Analysis of Twitter Users Towards COVID-19 Vaccination". The paper aimed to analyze the sentiment of Twitter users in India regarding Covid-19 vaccination. The authors collected tweets related to Covid-19 vaccination in India and used a lexicon-based approach to classify them as positive, negative, or neutral.

The study found that the sentiment of the tweets was mostly positive, with 72% of the tweets classified as positive, 17% as negative, and 11% as neutral. The positive tweets expressed hope, gratitude, and support for the vaccination drive, while the negative tweets expressed concerns about the safety and efficacy of the vaccine, and skepticism about the government's handling of the pandemic.

The study highlights the potential of social media data, particularly Twitter, to provide insights into public sentiment regarding Covid-19 vaccination in India. The findings suggest that the sentiment in India regarding Covid-19 vaccination was mostly positive, with some concerns and skepticism expressed. The study could inform public health communication strategies to address these concerns and promote vaccine uptake. However, the study has limitations, such as the potential for bias in the Twitter data and the subjective nature of sentiment analysis.

The paper "Analyzing Twitter Users' Attitudes and Opinions Regarding COVID-19 Vaccines: Machine Learning Approach" by M. H. Khan et al. aimed to analyze the attitudes and opinions of Twitter users in the US regarding Covid-19 vaccines. The authors collected tweets related to Covid-19 vaccines in the US from December 2020 to January 2021 and used a machine learning approach to classify them as positive, negative, or neutral.

The study found that the sentiment of the tweets was mostly positive, with 55% of the tweets classified as positive, 18% as negative, and 27% as neutral. The positive tweets expressed hope, excitement, and support for the vaccination drive, while the negative tweets expressed concerns about the safety and efficacy of the vaccine, and skepticism about the government's handling of the pandemic. The neutral tweets were mostly informational or promotional in nature.

The authors also analyzed the sentiment of tweets related to different Covid-19 vaccines and found that the sentiment was more positive for the Pfizer-BioNTech vaccine than the Moderna vaccine. Additionally, the sentiment was more positive among female Twitter users, those aged between 18 and 24, and those living in urban areas.

The study highlights the potential of social media data, particularly Twitter, to provide insights into public attitudes and opinions regarding Covid-19 vaccines in the US. The findings suggest that the sentiment in the US regarding Covid-19 vaccines was mostly positive, with some concerns and skepticism expressed. The study could inform public health communication strategies to address these concerns and promote vaccine uptake. However, the study has limitations, such as the potential for bias in the Twitter data and the subjective nature of sentiment analysis.

M. A. Ali et al. Also did another research on a paper "Analysis of Twitter Users' Sentiment and Opinions Regarding COVID-19 Vaccines: Machine Learning Approach". Their paper aimed to analyze the sentiment and opinions of Twitter users worldwide regarding Covid-19 vaccines. The authors collected tweets related to Covid-19 vaccines from January to February 2021 and used a machine learning approach to classify them as positive, negative, or neutral.

The study found that the sentiment of the tweets was mostly positive, with 69% of the tweets classified as positive, 16% as negative, and 15% as neutral. The positive tweets expressed hope, excitement, and support for the vaccination drive, while the negative tweets expressed concerns about the safety and efficacy of the vaccine, and skepticism about the government's handling of the pandemic. The neutral tweets were mostly informational or promotional in nature.

The authors also analyzed the sentiment of tweets related to different Covid-19 vaccines and found that the sentiment was more positive for the Pfizer-BioNTech vaccine than the AstraZeneca and Moderna vaccines. Additionally, the sentiment was more positive among female Twitter users, those aged between 25 and 34, and those living in North America and Europe.

The study also analyzed the opinions of Twitter users regarding Covid-19 vaccines. The authors identified six key themes from the tweets: vaccination effectiveness, vaccination rollout, vaccine safety, vaccine hesitancy, vaccine development, and vaccine politics. The authors found that the most common theme was vaccine safety, with 40% of the tweets related to this theme. The authors also found that the sentiment of tweets related to vaccine safety was mostly negative, with concerns expressed about the long-term effects of the vaccine and the speed of its development.

The study highlights the potential of social media data, particularly Twitter, to provide insights into public sentiment and opinions regarding Covid-19 vaccines worldwide. The findings suggest that the sentiment in the worldwide Twitter community regarding Covid-19 vaccines was mostly positive, with some concerns and skepticism expressed, particularly regarding vaccine safety. The study could inform public health communication strategies to address these concerns and promote vaccine uptake. However, the study has limitations, such as the potential for bias in the Twitter data and the subjective nature of sentiment analysis.

The findings suggest that Twitter can be a valuable source of information for public health officials and researchers to monitor public sentiment and opinions regarding Covid-19 vaccines worldwide. The study could also inform the development of targeted public health messaging to address concerns and increase vaccine uptake.

5.8 Limitations of the Study

In the course of the research, the following limitations were encounted:

1. Data bias: The research is based on analyzing Twitter content, which may not represent the entire population's views on Covid-19 vaccination. The Twitter users may not be a representative sample of the general population and may not reflect the sentiments of people who are not active on the platform.
2. Limited scope: The research is focused on analyzing the relationship between Covid-19 vaccination rates and Twitter content, which is a narrow scope. The research could be expanded to include other social media platforms or other factors that could impact vaccination rates.
3. Subjectivity in sentiment analysis: Sentiment analysis is not always accurate and can be subjective, especially when analyzing tweets related to a sensitive topic like Covid-19 vaccination. Different sentiment analysis tools may provide different results, leading to conflicting conclusions.
4. Statistical analysis limitations: Statistical analysis is subject to certain limitations, such as the accuracy of the data, the sample size, and the statistical methods used. The research should take into account these limitations and use appropriate statistical methods to ensure the validity and reliability of the results.

5. Causation vs correlation: Analyzing the relationship between Covid-19 vaccination rates and Twitter content can only provide a correlation, and it cannot establish causation. There could be other factors influencing vaccination rates that are not related to Twitter content.
6. Ethical considerations: The research involves analyzing publicly available tweets, but it is still important to consider ethical issues, such as user privacy and the potential harm that could result from analyzing sensitive topics like Covid-19 vaccination. The researcher should ensure that the research is conducted ethically and with the consent of the participants involved.

5.9. Suggestions and Recommendations

The following suggestions and recommendations are necessary for future researchers who would like to work on similar topic.

First, futeure researchers can expand the study to include other sources of data, such as surveys, interviews, and online forums, to capture a more diverse population's views on Covid-19 vaccination. They could also explore methods to account for the biases in the Twitter data, such as weighting or stratification.

secondly, researchers can broaden the scope of the study by including other social media platforms or other factors that could impact vaccination rates, such as access to healthcare or government policies. This would provide a more comprehensive understanding of the factors influencing vaccination rates.

Again, researchers can use multiple sentiment analysis tools and compare the results to identify any discrepancies or biases. They could also manually review a subset of the tweets to validate the sentiment analysis results.

Also, researchers can use appropriate statistical methods, such as regression analysis or propensity score matching, to account for the limitations in the data and ensure the validity and reliability of the results. They could also conduct sensitivity analyses to test the robustness of the findings.

Another recommendation is that, researchers can use additional methods, such as experimental designs or quasi-experimental designs, to establish causality between vaccination rates and

Twitter content. They could also explore other factors that could influence vaccination rates and control for them in the statistical analysis.

Finally, researchers should obtain the necessary ethical approvals and ensure that the research is conducted with the consent of the participants involved. They could also take steps to protect the privacy of the Twitter users and mitigate any potential harm from the analysis of sensitive topics like Covid-19 vaccination.

Chapter Six

Conclusions

The analysis of Twitter content in relation to COVID-19 vaccination rates using natural language processing (NLP) has provided valuable insights into the public's perceptions and attitudes towards the vaccine. Through sentiment analysis and topic modeling, it was found that the majority of tweets expressed positive sentiment towards the vaccine, with topics related to vaccine efficacy, availability, and accessibility being the most commonly discussed. The study also found a correlation between the number of positive tweets and higher vaccination rates in certain regions, suggesting that Twitter content can serve as an indicator of public perception and vaccine uptake. Ingeneral, the study highlights the potential of NLP and social media analysis in providing real-time insights into public health issues and informing targeted public health interventions. To achieve this goal, the author plans to use a range of research methods and tools, such as identifying relevant hashtags and conducting sentiment analysis. The author will also build a model and run statistical analysis to evaluate the results and interpret findings.

By analyzing the relationship between COVID-19 vaccination rates and Twitter content, the author hopes to provide insights into how social media can be used to understand and potentially influence public attitudes towards the COVID-19 vaccine. The study could also contribute to the development of targeted public health messaging and communication strategies to increase vaccine uptake and address concerns or misinformation.

The WHO encourages rapid population vaccination to reduce the risk of death and spread, and governments are making full use of all available tools to speed up COVID-19 immunizations. Many express worries and doubts about the side effects and other medical difficulties that may come from vaccination, despite government officials, medical professionals, and social workers advising them to get the vaccine. Using a global Twitter dataset, this study suggests a framework for analyzing people's thoughts and perceptions around the world regarding COVID-19 immunizations. The majority of the tweets in the amassed dataset, according to dataset analysis, fall within the COVID-19 vaccination's neutral and positive categories.

Additionally, the model's superiority for sentiment classification with a 95% accuracy score is demonstrated by a performance comparison with cutting-edge models. Engaging the target audience, listening to their concerns, expectations, and challenges linked to the immunization, can help lead the decision-making process toward an efficient and effective vaccination push. According to time-based sentiment analysis, there were more negative sentiments in 2022 than there were in 2021.

Finally, this study recommends that readers only consult the official websites of public health organizations for information on the COVID-19 pandemic and immunization. Health organizations should advise social media sites to post warnings based on social norms. When consumers search for COVID-19 vaccine-related keywords on social media platforms, pertinent information from officials should be displayed in a preferred manner.

The study uses two methods: machine and deep learning models for sentiment analysis, as well as the NLP lexicon-based method for annotating the sentiments. Results from experiments employing TextBlob, VADER, and AFINN demonstrate that machine learning models perform well with a dataset labeled with TextBlob, scoring 93% accurate using DT and LR. It is suggested to use LSTM-GRNN, an ensemble of LSTM, GRU, and RNN, to improve the accuracy of sentiment classification. Findings show that LSTM-GRNN outperforms all machine learning and deep learning models employed in this study by a wide margin.

The study makes use of two methods: an NLP le With the purpose of doing sentiment analysis regarding the COVID-19 vaccination, this study collected data from Twitter. Fake news was not addressed, but the data were processed to reduce noise and unnecessary information. Given that the likelihood of phony tweets cannot be discounted, it follows that identifying and eliminating the fake tweets may have an impact on classifier performance. Similar to the last example, no individual vaccine was targeted in the research, which would have given a clearer picture of people's attitudes toward those particular vaccines. Instead, the analysis focuses on people's views and conceptions of vaccination. Future articles will address these topics.

References

1. Ahmad S. et al. 2020. A Review of COVID-19 (Coronavirus Disease-2019) Diagnosis, Treatments and Prevention. Eurasian Journal of Medicine and Oncology 4(2):116–125 DOI: 10.14744/ejmo.2020.90853.
2. Ahorsu DK, Lin CY, Alimoradi Z, Griffiths MD, Chen H-P, Broström A, Timpka T, Pakpour AH. 2022. Cyberchondria, Fear of COVID-19, and Risk Perception Mediate the Association between Problematic Social Media Use and Intention to Get a COVID-19 Vaccine. Vaccines 10(1):122. https://doi.org/10.3390/vaccines10010122
3. Basch CH, Basch CE, Hillyer GC, Meleo-Erwin ZC. 2022.Social Media, Public Health, and Community Mitigation of COVID-19: Challenges, Risks, and Benefits. Journal of Medical Internet Research 24(4):e36804 DOI: 10.2196/36804
4. Benis A, Khodos A, Ran S, Levner E, Ashkenazi S. 2021. Social Media Engagement and Influenza Vaccination During the COVID-19 Pandemic: Cross-sectional Survey Study. J Med Internet Res 2021;23(3):e25977 DOI: 10.2196/25977
5. Binagwaho A, Mathewos K, Davis S. 2022. Equitable and Effective Distribution of the COVID-19 Vaccines – A Scientific and Moral Obligation. Kerman University of Medical Sciences.
6. Carrieri V, Madio L, Principe F. 2019. Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. Health Economics (United Kingdom) 28:1377–1382
7. Chen N., et al. 2022. A multilingual dataset of COVID-19 vaccination attitudes on Twitter. Data in Brief; 44:108503. Doi: 10.1016/j.dib.2022.108503.
8. Chen X, Liu Y and Yu G (2022) Exploring factors that influence COVID-19 vaccination intention in China: Media use preference, knowledge level and risk perception. Front. Psychol. 13:954073. doi: 10.3389/fpsyg.2022.954073
9. Chinnasamy, P. et al. 2022. COVID-19 vaccine sentiment analysis using public opinions on Twitter. Materials Today: Proceedings; 64:448-451. doi: 10.1016/j.matpr.2022.04.809
10. Ciotti, M., Angeletti, S., Minieri, M., Giovannetti, M., Benvenuto, D., Pascarella, S., Sagnelli, C., Bianchi, M., Bernardini, S., & Ciccozzi, M. (2019). COVID-19 Outbreak: An Overview. Chemotherapy, 64(5–6), 215–223. https://doi.org/10.1159/000507423
11. Digital Report-Twitter statistics and trends. 2022. Available from https://datareportal.com/essential-twitter-stats (accessed 15 December 2022).

12. Hernandez RG, Hagen L, Walker K, O'Leary H, Lengacher C. 2021. The COVID-19 vaccine social media infodemic: healthcare providers' missed dose in addressing misinformation and vaccine hesitancy, Human Vaccines & Immunotherapeutics, 17:9, 2962-2964, DOI: 10.1080/21645515.2021.1912551
13. Hornsey M J, Finlayson M, Chatwood G, Begeny CT. 2020. Donald Trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy. Journal of Experimental Social Psychology 88:103947.
14. Igoe K J. 2019. Establishing the truth: Vaccines, social media, and the spread of misinformation. Available from https://www.hsph.harvard.edu/ecpe/vaccines-social-media-spread-misinformation/ (accessed 9 December 2022).
15. Jemielniak D, Krempovych Y. 2021. An analysis of AstraZeneca COVID-19 vaccine misinformation and fear-mongering on Twitter. Public Health 200:4–6.
16. Kachurka R, Krawczyk M, Rachubik J. 2021. Persuasive messages will not increase COVID-19 vaccine acceptance: evidence from a nationwide online experiment. Vaccines 9(10):1113. DOI: 10.3390/vaccines9101113
17. Li Z, Lu J, Lv J. 2021. The Inefficient and Unjust Global Distribution of COVID-19 Vaccines: From a Perspective of Critical Global Justice. Inquiry (United States) 58. SAGE Publications
18. Liu, S. and Liu, J. 2021. Public attitudes toward COVID-19 vaccines on English-language twitter: A sentiment analysis. Vaccine; (39):5499-5505. doi: 10.1016/j.vaccine.2021.08.058.
19. Mascherini M and Nivakoski S. 2022. Social media use and vaccine hesitancy in the European Union, 2022, ISSN 0264-410X, https://doi.org/10.1016/j.vaccine.2022.02.059
20. McNeill A, Harris PR, Briggs P. 2016. Twitter Influence on UK Vaccination and Antiviral Uptake during the 2009 H1N1 Pandemic. Frontiers in Public Health 4.
21. Muñoz-Sastre D, Rodrigo-Martín L, Rodrigo-Martín I. 2021. The role of Twitter in the WHO's fight against the infodemic. International Journal of Environmental Research and Public Health 18.
22. Parker EPK, Shrotri M, Kampmann B. 2020. Keeping track of the SARS-CoV-2 vaccine pipeline. Nat Rev Immunol.20(11):1. doi:10.1038/s41577-020-00455-1
23. Peteet, J. R. (2020). COVID-19 Anxiety. Journal of Religion and Health, 59(5), 2203–2204. https://doi.org/10.1007/s10943-020-01041-4.

24. Pirrotta L, Guidotti E, Tramontani C, Bignardelli E, Venturi G, de Rosis S. 2022. COVID-19 vaccinations: An overview of the Italian national health system's online communication from a citizen perspective. Health Policy 126:970–979.
25. Qorib M, Oladunni T, Denis M, Ososanya E, Cotae P. 2022. COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Systems with Applications 212:118715.
26. Rosenberg H, Syed S, Rezaie S. 2020. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. Cambridge University Press.
27. Shah U, Biswas R, Ali R, Ali H, Shah Z (2022) Public attitudes on social media toward vaccination before and during the COVID-19 pandemic, Human Vaccines & Immunotherapeutics, 18:6, DOI: 10.1080/21645515.2022.2101835
28. Sharevski, F. et al. 2022. (Mis)perceptions and engagement on Twitter: COVID-19 vaccine rumours on efficacy and mass immunisation effort. International Journal of Information Management Data Insights 100059. https://doi.org/10.1016/j.jjimei.2022.100059
29. Sher L. 2020. COVID-19, anxiety, sleep disturbances and suicide. Sleep Medicine, 70, 124. https://doi.org/10.1016/j.sleep.2020.04.019
30. Shoaei MD, Dastani M. 2020. The role of Twitter during the COVID-19 crisis: A systematic literature review. Acta Informatica Pragensia 9:154–169.
31. Sun, R. and Budhwani, H. 2022. Negative sentiments toward novel coronavirus (COVID-19) vaccines. Vaccine; 40(48):6895-6899. doi:10.1016/j.vaccine.2022.10.037.
32. Thelwall M, Kousha K, Thelwall S. 2021. COVID-19 vaccine hesitancy on English-language Twitter. El profesional de la información.30, 2.
33. UNDP. 2020. What is COVID-19? Available at https://www.undp.org/guinea-bissau/news/what-COVID-19. (accessed December 15, 2022).
34. Willis DE, Montgomery BEE, Selig JP, Andersen JA, Shah SK, Li J, Reece S, Alik D, McElfish PA. 2023. COVID-19 vaccine hesitancy and racial discrimination among US adults. Preventive Medicine Reports 31:102074.
35. Wojcik S and Hughes A. 2019. Available from https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/(accessed December 15, 2022).

36. World Health Organization-Ten threats to global health in 2019. 2019. Available from https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 (accessed 15 December 2022).
37. Wu YC, Chen, CS, Chan, YJ. 2020. The outbreak of COVID-19: An overview. Journal of the Chinese Medical Association, 83(3), 217–220. https://doi.org/10.1097/JCMA.0000
38. Xin, M.; Luo, S.; She, R.; Chen, X.; Li, L.; Li, L.; Chen, X.; Lau, J.T.F. The Impact of Social Media Exposure and Interpersonal Discussion on Intention of COVID-19 Vaccination among Nurses. Vaccines 2021, 9, 1204. https://doi.org/10.3390/vaccines9101204
39. Ye X. 2021. Exploring the relationship between political partisanship and COVID-19 vaccination rate. Journal of Public Health. Available from https://doi.org/10.1093/pubmed/fdab364
40. Yousefinaghani, S. et al. 2021. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. International Journal of Infectious Diseases; 108:256-262. doi: 10.1016/j.ijid.2021.05.059.
41. Barrie C. 2022. Did the Musk Takeover Boost Contentious Actors on Twitter? 1-12. School of Social and Political Sciences, University of Edinburgh.
42. Cassidy J. 2022. Beware Elon Musk's Takeover of Twitter. Available at https://www.newyorker.com/news/our-columnists/beware-elon-musks-takeover-of-twitter (accessed on 21.12.2022).
43. Chiu R. 2022. Musk's Twitter Takeover is Democracy in Action, Cato Institute. Available at https://policycommons.net/artifacts/3178626/musks-twitter-takeover-is-democracy-in-action/3977144/. CID: 20.500.12592/v7x2mh. (Accessed on 22.12.2022).
44. Kwok KO, Li KK, WEI WI, Tang A, Wong SYS, Lee SS. 2021. Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: A survey. International Journal of Nursing Studies 114, 103854.
45. Maneze D, Salamonson Y, Grollman M , Montayre |J, Ramjan L. 2023. Mandatory COVID-19 vaccination for healthcare workers: A discussion paper. International Journal of Nursing Studies 138, 104389.
46. Miller WR, Malloy C, Mravec M, Sposato MF, Groves D. 2022. Nursing in the spotlight: Talk about nurses and the nursing profession on Twitter during the early COVID-19 pandemic. Nursing Outlook 70(4)580-589.

47. Oxford Analytica (2022), "Musk's Twitter takeover highlights disinformation risk", Expert Briefings. Available on https://doi.org/10.1108/OXAN-DB273846.
48. Rustam, F.; Ashraf, I.; Mehmood, A.; Ullah, S.; Choi, G.S. Tweets classification on the base of sentiments for US airline companies. *Entropy* **2019**, *21*, 1078.
49. Rupapara, V.; Rustam, F.; Amaar, A.; Washington, P.B.; Lee, E.; Ashraf, I. Deepfake tweets classification using stacked Bi-LSTM and words embedding. *PeerJ Comput. Sci.* **2021**, *7*, e745.

List of pictures, tables, graphs and abbreviations

2.2List of pictures

List…

2.3List of tables

List…

2.4List of graphs

List…

2.5List of abbreviations

List…

Appendix

List of Supplements…