THESIS

PANDEMIC PERCEPTIONS: ANALYZING SENTIMENT IN COVID-19 TWEETS

Submitted by

Shadaab Kawnain Bashir

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2023

Master's Committee:

Advisor: Indrakshi Ray
Co-Advisor: Hossein Shirazi

Haonan Wang

ABSTRACT

PANDEMIC PERCEPTIONS: ANALYZING SENTIMENT IN COVID-19 TWEETS

Social media, particularly Twitter, became the center of public discourse during the COVID-19 global crisis, shaping narratives and perceptions. Recognizing the critical need for a detailed examination of this digital interaction, our research dives into the mechanics of pandemic-related Twitter conversations. This study seeks to understand the many dynamics and effects at work in disseminating COVID-19 information by analyzing and comparing the response patterns displayed by tweets from influential individuals and organizational accounts. To meet the research goals, we gathered a large dataset of COVID-19-related Tweets during the pandemic, which was then meticulously manually annotated. In this work, task-specific transformers and LLM models are used to provide tools for analyzing the digital effects of COVID-19 on sentiment analysis. By leveraging domain-specific models $\text{RoBERTa}_{\text{Twitter}}$ fine-tuned on social media data, this research improved performance in critical task of sentiment analysis. Investigation demonstrates individuals express subjective feelings more frequently compared to organizations. Organizations, however, disseminate more pandemic content in general.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to everyone who helped me along the way in school and helped me finish my thesis.

First and foremost, I want to express my sincere gratitude to Indrakshi Ray and Hossein Shirazi, who served as my thesis advisor and co-advisor respectively, for their unwavering guidance, patience, and invaluable insights. Their knowledge and commitment were essential in determining the focus and quality of this study.

My family's continuous support, inspiration, and patience during the trying parts of this project are all things I am grateful for. Their confidence in me has served as a constant inspiration.

Last but not least, this work was supported in part by funding from NSF under Award Numbers DMS 2123761, CNS 1822118, and from AFRL, ARL, Statnett, AMI, NewPush, and Cyber Risk Research and from NIST under Award Number 60NANB23D152.

This thesis is the result of many months of effort, and it would not have been possible without the assistance of the people and organizations indicated above. I sincerely appreciate all of their help, and I hope that this work will be a good representation of how much they believe in my skills.

DEDICATION

I would like to dedicate this thesis to my parents and my spouse

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The COVID-19 pandemic caused by the novel Severe Acute Respiratory Syndrome Coronavirus 2 ($SARS - CoV - 2$) shocked the world and created a lot of uncertainty in the normal course of life. This virus originated in Wuhan, China [41], and spread like wildfire, resulting in massive destruction in the world. Many lives were lost (6,887,000), as well as thousands of reported cases (761,402,282) of COVID-19, which were reported every single day [45]. Medical and healthcare professionals and even government officials had to play catch up to combat this state of emergency. World Health Organisation (WHO) declared COVID-19 a pandemic on March 11, 2020 [44], with a lot of mandates including lockdowns, the wearing of masks, sanitizing hands, etc. In the meantime, researchers worked night and day to come up with a vaccine for this virus, which itself was a challenge in such a state of emergency.

The weight of developing plans for managing the population and efficiently utilizing the available resources fell unexpectedly on each nation as a result of the rise in COVID-19 cases. Fear, anxiety, and distress among people have increased as a result of exponentially rising cases across the globe. Many people have lost their employment as a result of the pandemic, been forced to work from home, or need to enroll in distance learning programs. Researching how people feel is essential to understand how the COVID-19 pandemic has impacted people's mental health. There is evidence that this pandemic is closely related to the physical and mental health of the world's population [5, 18].

Understanding the effect that COVID-19 had is crucial for protecting communities from misery, anxiety, and mental disease. Twitter and other social media has become a popular outlet for people to voice their worries, opinions, and thoughts about the global pandemic during these challenging circumstances [1]. Social media data, gathered by live-streaming public tweets using an Application Programming Interface (API) search, is a useful resource for computer scientists and researchers to understand how people feel about the pandemic. Due to false information shared on social media, COVID-19 is now becoming a source of despair, stress, and worry in addition to being an infectious disease that is spread through contact and by minute droplets formed when people cough, sneeze, or talk. The quick dissemination of misleading information on social media has a direct impact on mental health. Individuals' primary reliance on the Internet and the highest activity have been noted due to the state of lockdown and social isolation [35]. Many outbreaks and pandemics could have been swiftly controlled if experts had taken social media data into account, according to numerous research publications [37].

Sentiment analysis, also known as opinion mining, is a contextual text-mining technique that locates and pulls subjective data from texts. It aids in the analysis of people's attitudes and feelings toward things like goods, people, and situations. Text mining has been a useful way to get datasets while working on social media databases, as noted by [39]. Social media text mining appears to be an appropriate alternative for conducting public opinion research on COVID-19 [27]. As a result, different perspectives can be provided by gathering and monitoring data from social media sites like Twitter. Also, using Twitter large-scale datasets can be collected at a low cost [7]. In this research, we used tweets from influencers, and by an influencer, we refer to a person or organization with at least 10K followers or an account that has been verified. Since influencers have a large following, information may propagate more quickly. Additionally, the fact that these influencers are tweeting from verified accounts suggests that the information is trustworthy. Furthermore, the influencers have a large following and their tweets have an effect on how those followers behave because they want to imitate the influencers they are following.

Studies that have recently been presented concentrated on the automatic detection of tweets posted concerning COVID-19 [22, 47]. However, there has not been much research done on examining the sentiment elements of these tweets related to COVID-19. By evaluating the development of sentimental content in COVID-19 tweets from influencers, our work seeks to close this gap. The following are the main goals of our study: (i) To investigate the sentiment that influencers expressed in their tweets about COVID-19. (ii) To examine the global sentiment trend over the course of 14 months.

To answer these queries, we created a novel labeled dataset of COVID-19 and conducted an experimental investigation utilizing Natural language processing (NLP) techniques for sentiment analysis. We address our effort on developing a labeled dataset for COVID-19 that includes labeled instances of sentiment expressed in tweets. Each tweet's sentiment is classified into three categories: positive, negative, or neutral. It is essential to emphasize that the sentiment annotation method focuses exclusively on the conveyed sentiment within the tweet by the author, rather than how it makes the annotator feel. To that end, we used multiple types of transformers and Large Language Model (LLM)s to train and test the dataset in order to provide baselines for these tasks. Our findings contribute to a deeper understanding of COVID-19 sentiment on Twitter and serve as a helpful resource for researchers exploring this topic.

This study develops NLP models to assess sentiment from Tweets, which increases our comprehension of the COVID-19 controversy. Important findings show that while organizations broadcast larger amounts of pandemic-related content overall, individuals express greater subjective feelings than organizations. Furthermore, we detected marked polarity in individual tweets (positive or negative feelings), in contrast to organizations' more neutral attitudes. However, the sheer number of tweets from organizations surpasses those from individuals. By empowering nuanced investigation across sentiment, this research significantly enhances the capacity to decode complex perspectives and narratives shaping the social media crisis discourse.

The rest of the paper is divided into the following sections: In Chapter 2, we explore the literature review and examine earlier studies on sentiment analysis in tweets on COVID-19. The methodology of our work is described in depth in Chapter 3, including the models used in our experiments. Our experimental results are reported in Chapter 4 along with trend analysis shown in graphical charts. Chapter 5 wraps up our work and initiates discussions regarding future directions.

Chapter 2

Literature Review

In this section, we analyze and discuss major contributions in the area of sentiment analysis, highlighting the evolution of sentiment analysis methodology, and current state-of-the-art solutions that have influenced the landscape of research. Several studies on the sentimental impact of COVID-19 have been conducted as a result of this pandemic using Twitter data. In the Table 2.1, we have listed the dataset, sentiment detection technique, and the algorithm that was used in the existing literature.

2.1 TextBlob

TextBlob is a Python package that may be used to determine the polarity of a tweet. Polarity refers to the expression of positive, negative, or neutral sentiment in text, which can be recognized using NLP tasks to obtain insight into subjective ideas and feelings. For COVID-19-related tweets, it is frequently utilized by Chandrasekaran and Hemanth [6], Khan et al. [19], Manguri et al. [25], Mishra et al. [26], Raza et al. [31], Vijay et al. [42] to extract the sentiment from the tweet's text.

Manguri et al. [25] looked at the COVID-19 outbreak and used TextBlob to focus on two keywords to determine the polarity and subjectivity surrounding the pandemic. On the other hand, in order to extract thoughts connected to COVID-19, we used 42 keywords in our study. Furthermore, the authors in this study collected data for only 7 consecutive days, but we collected data for 14-months and examined the pattern over this extended period of time. Their findings revealed that there were more than 50% neural sentiments, which is consistent with our findings.

Mishra et al. [26] in their research examined the COVID-19 vaccinations produced by several organizations and discovered that Moderna was the vaccine that Twitter users preferred the most. However, in our work, we not only considered vaccine keywords but also looked at several topics that are directly relevant to this pandemic.

Similarly, in the study by Raza et al. [31], the authors focused on vaccination hesitancy, using TextBlob, and reached the best accuracy, 93.15%, by employing Support Vector Machine (SVM) classifier with TF-IDF vectorization. The difference between this and our work is that because we used transformer-based models, our system was able to keep the tweet's semantics after word embedding. However, in their study, the models will be unable to preserve the semantics of the tweets.

Vijay et al. [42] have demonstrated that when a lockdown is announced, people's attitudes change from negative to positive and the number of tweets posted increases. In this work, the authors have collected the data for 7 months to perform their analysis. For our trend analysis task, we did not notice similar observations as the authors have claimed. However, we discovered that the number of negative tweets was always more than the number of positive tweets across a 14-month period.

Similar outcomes of positive sentiment about the COVID-19 restriction were also observed in the work Khan et al. [19]. In this task, the writers intended to see how individuals in India reacted when the government issued a tweet. The authors employ TextBlob to mark the polarity of the tweet in this case. The primary difference between their work and ours is that we used classification to detect sentiment from tweets. Furthermore, because we employed a transformer-based model, we do not need to perform any text normalization to clean our data.

Chandrasekaran and Hemanth [6] employed machine learning and deep learning techniques to examine the sentiment of Twitter data, with the outcome indicating that individuals showed optimism towards the pandemic's recovery.

2.2 VADER

Valence Aware Dictionary and sEntiment Reasoner ($VADER$) is another program that uses a rule-based methodology to ascertain the sentiment of tweets. To assess the sentiments of a tweet, $VADER$ is employed in numerous research works, including Chakraborty et al. [5], Liu and Liu [20], Luo and Kejriwal [24], Wrycza and Maślankowski [46].

The Liu and Liu [20], and Luo and Kejriwal [24] investigate vaccine hesitancy, following the announcement of the vaccination by Pfizer, there was an improvement in public opinion, as evidenced by the Liu and Liu [20]. The authors of this paper employed LDA to find themes for attitudes that contribute to distinct types of sentiment so that vaccine information can be conveyed properly. However, we have not looked into the themes behind each class of sentiment that we have used in our analysis. On the other hand, Luo and Kejriwal [24] demonstrates that vaccine hesitancy remained a significant issue even after the government marketed the vaccine's advantages. The authors of this article conducted a comparative analysis of vaccine hesitancy during the 2020 US presidential election. According to their findings, there was considerably less apprehension toward the vaccine following the election. However, in the course of our research, we concentrated not just on the topic of vaccines, but also on the 42 keywords associated with COVID-19.

Wrycza and Maślankowski [46] used Naïve Bayes classifier and the $VADER$ lexicon to examine people's sentiment towards remote work during COVID-19. The authors noted that in March 2020, when this issue reached its peak, individuals began to accept the idea of working from home. Similar trends regarding work from home have been observed during our experiment, with a largely positive sentiment toward the topic of work from home.

Finally, Chakraborty et al. [5] employed fuzzy inference using the $VADER$ in their study from the classification of sentiment into three categories, and their proposed model obtained an F1-score of 79$\%$. However, in our research, we have seen that our classifier provided better results with an F1-score of 84$\%$. The author also found in their investigation that more neutral tweets were posted on Twitter, which is similar to what we discovered in our study. Furthermore, the authors examined and discovered that people retweeted more messages with a negative attitude. However, we have not conducted any retweet analysis in our research.

2.3 TextBlob and VADER

Sattar and Arifuzzaman [34], and Qorib et al. [29] employ both TextBlob and $VADER$ for sentiment analysis. Sattar and Arifuzzaman [34] used both tools to label the sentiment connected to vaccination awareness.This study uses sentiment analysis on seven distinct types of vaccines to investigate the issue of vaccine awareness. Researchers also conducted a time predicting analysis in their work, looking at when the first dose would be completed in the United States, as well as when most states will be fully vaccinated. It was also discovered that their forecast corresponds to the government's forecast. However, in our work, we examined not only the issue of vaccinations but also various other themes for our sentiment analysis task.

Using several ML algorithms, Qorib et al. [29] classified the sentiment of each tweet. Researchers employed TextBlob and $VADER$ to examine the polarity of the tweets. The authors of this study assessed 42 models and found that TextBlob with TF-IDF vectorization and the LinearSVC classification model gave them the best results. The authors of this work, however, did not conduct their analysis using any transformer-based models.

2.4 Machine Learning

After reviewing numerous distinct approaches, we examined the various machine-learning algorithms utilized for sentiment analysis.

Syuzhet is an R package used for sentiment analysis and emotion detection in textual data. Samuel et al. [33] analyzed the sentiment of tweets where the authors used machine learning algorithms to classify the tweet into two classes negative and positive sentiments. This study examines how the fear of COVID-19 has evolved over time. To perform this classification they divided the tweets into long and short tweets and they saw that classifying short tweets with Naïve Bayes gave them an accuracy of 91$\%$. However, in our work, we have not categorized our tweets into long or short tweets for classification.

Sahir et al. [32], use the Naïve Bayes algorithm to examine the public opinion on COVID-19 and divide it into three categories: positive, neutral, and negative for online learning for October 2020 in Indonesia. According to the author, the infrastructure for online education was not accessible at the start of COVID-19, which resulted in greater negative sentiment against online education in Indonesia. However, we have not addressed the topic of online education in our research.

Imran et al. [16], applied Deep Neural Network (DNN) and Long Short-Term Memory ($LSTM$) with word embeddings to classify the sentiments of covid tweets. The experimental results show that using Bidirectional Long Short-Term Memory ($Bi-LSTM$), the authors were able to achieve an accuracy of 81.64% on the Sentiment140 dataset. The authors also used Transformer models, Bidirectional Encoder Representations from Transformers ($BERT$), in this work, although their results were poor when this Transformer model was used. On the other hand, we found our $BERT$ model, we were able to achieve better outcomes.

Jang et al. [17] employed ABSApp, a portable system for weakly-supervised aspect-based sentiment extraction, to assess aspect-based sentiments on COVID-19-related public health interventions or problems and generated two labels (positive/negative) for the sentiments. The author conducted this analysis based on the region of Canada, and they chose the subject of face masks and social distancing for this task. They used LDA to conduct a trend analysis, looking at how COVID-19-related subjects progressed over time.

Rajput et al. [30] also used NLP methods determined the word frequency and sentiments of each tweet using Unigram, Bigram, and Trigram frequencies modeled by a power-law distribution. They separated the tweets into two categories: those posted by WHO and those made by the general public. In their investigation, they discovered that WHO has posted more positive tweets than the general public, which is similar to what we discovered in our individual versus organization analysis.

Contreras Hernández et al. [9] used transformers to identify sentiments from Twitter data where the tweets were written in Spanish and originated in Mexico; in this work, the authors found greater precision when utilizing the Spanish-based $BERT$ transformers model. However, the authors can also look at the tweets from a global perspective using different languages in their analysis.

The methods that were used to detect the sentiments are different from our approach because in our research we tried to analyze the sentiment that the author was trying to convey through the tweets and for that, we have used gold standard data. However, in the literature, most of the researchers used a lexicon-based approach to analyze the sentiments of the tweets. Our research intends to add to a thorough knowledge of user sentiment on social media platforms, particularly during the COVID-19 pandemic, by building on the corpus of prior research. Through this analysis, we will have a deeper knowledge of the variables affecting user behavior and be able to look into the trends around particular topics during the COVID-19 period.

Table 2.1: Overview of the existing approaches on COVID-19 Sentimental Analysis

Ref	**Dataset size**	**Sentiment Techniques**	**Method**
[25]	530,232 tweets are retrieved from Twitter for seven days from 09-04-2020 to 15-04-2020	TextBlob	Naïve Bayes
[26]	Every day, 6000 tweets are retrieved	TextBlob	
[31]	206,465 were collected from Twitter	TextBlob	TF-IDF using SVM
[42]	140,000 tweets were collected for month-wise analysis and 250,000 tweets were collected for day-wise analysis from December 2019 to May 2020.	TextBlob	-
[19]	From February through April, 50,000 tweets on the COVID-19 pandemic were collected	TextBlob	Naïve Bayes
[6]	-	TextBlob	$Bi-LSTM$
[20]	2,678,372 tweets were scraped from Twitter	$VADER$	-
[24]	API search was used to obtain 371,940 tweets. There are 128,472 tweets left after deleting non-English and retweets	$VADER$	-
[46]	523,000 tweet were collected from 1 February 2020 to 10 October 2020	$VADER$	Naïve Bayes
[5]	A dataset of 226,668 tweets collected between December 2019 and May 2020	$VADER$	Fuzzy Inference
[34]	1.2M original tweets from Twitter were collected and the tweets were gathered over the course of five weeks, commencing on 10 April 2021 and concluding on 17 May 2021.	TextBlob and $VADER$	SVM
[29]	Created dataset of 61102 positive tweets, 39195 neutral tweets, and 23987 negative tweets.	TextBlob and $VADER$	TF-IDF with LinearSVC
[33]	From February to March of 2020, 900,000 tweets were collected	Syuzhet	Naïve Bayes

[32]	159,045 Indonesian tweets were retrieved from Twitter	-	Naïve Bayes
[16]	For the time period of the 3rd to the 29th of February 2020, and a dataset from Kaggle [38] was also used.	-	$Bi-LSTM$
[17]	Collected a total of 319,524 English tweets from the USA (293,929) and Canada (25,595).	-	ABSApp
[30]	From 11 March 2020 to 30 March 2020, 4M tweets were scraped in English, Spanish, and French.	Unigram, Bi-gram, and Trigram frequencies	Power-law distribution
[9]	Compiled and analyzed 10,619 Spanish tweets during the months of August-October of 2021.	-	$BERT$

Chapter 3

Methodology

The COVID-19 pandemic has had a significant impact on societies all around the world, and social media platforms have developed crucial places for people to discuss their thoughts and views about the ongoing issue. Analyzing user-generated information from platforms such as Twitter can help to understand the public's thoughts, feelings, and sentiments concerning COVID-19. This methodology section describes the processes for conducting sentiment analysis on tweets concerning COVID-19.

3.1 Data collection

We use Twitter Streaming API to get data by capturing real-time tweets. Using a collection of constantly updated keywords and hashtags to fetch the appropriate tweets in the English language from the API regarding COVID-19. For instance, we included the COVID-19 variants Delta and Omicron to our list when they were discovered. Using the Twitter API, we collected tweets containing keywords and hashtags, as shown in Table 3.1. We filtered our data using the English language, and it came from all around the world. In addition, we removed the retweeted data from their tweets. Finally, we generated our dataset using these two filters, and the data collecting processing is shown in Figure 3.1.

The summary of the dataset size and data collection period, which began in September 2020 and lasted for 14 months until October 2021, is shown in Table 3.2. This resulted in a total of

Table 3.1: COVID-19 keywords. 42 keywords used to gather COVID-19-related Tweets.

booster, china, corona, coronavirus, coronavirusoutbreak, coronaviruspandemic, coronavirusupdates, (covid-19), covid, covid-19, covid__19, covid_19, covid—19, covid19, covid19https, covid19outbreak, covid19pandemic, covid2019, covidoutbreak, covidpandemic, delta variant, johnson and johnson, lockdown, masks4all, moderna, omicron, omicronvariant, outbreak, pandemic, pfizer, quarantine, sanitizer, selfIsolation, socialdistancing, stayhome, staysafe, vaccine, vaxxed, virus, washyourhands, wfh, workfromhome

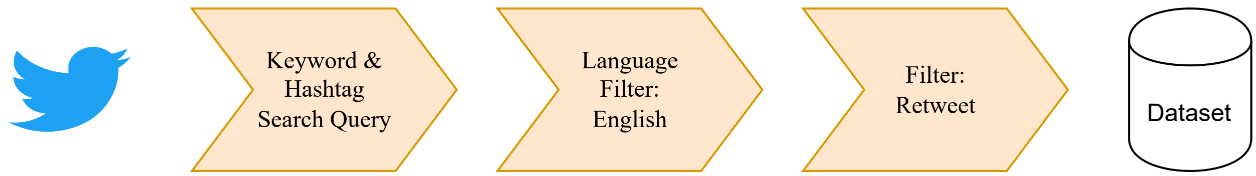

Figure 3.1: Twitter Data collection pipeline

over 447M tweets. In each month, we specified how many days the algorithm was completely functioning, for example, 31 days in October 2020 and just 11 days in March 2021. A total of 324 days were covered by our algorithm's full operation, resulting in an average of 1.38M tweets every day.

Table 3.2: Summary of monthly data collection: **#Tweets:** total number of tweets per month, **#Samples:** number of randomly selected sample tweets for each month, **#Days:** operational days, **Avg./Day:** average number of tweets per month. Columns marked with * indicate values in millions.

Data	#Tweets*	#Samples*	#Days	Avg./Day*
Sep-2020	14.60	0.14	18	0.81
Oct-2020	49.97	0.49	31	1.61
Nov-2020	58.51	0.58	30	1.95
Dec-2020	70.53	0.70	30	2.35
Jan-2021	55.91	0.55	30	1.86
Feb-2021	39.87	0.39	20	1.99
Mar-2021	23.77	0.23	11	2.16
Apr-2021	38.53	0.37	27	1.43
May-2021	25.93	0.21	31	0.84
Jun-2021	12.61	0.10	18	0.70
Jul-2021	7.49	0.06	10	0.75
Aug-2021	20.93	0.18	27	0.78
Sep-2021	18.24	0.16	25	0.73
Oct-2021	10.70	0.09	16	0.67
Total	**447.58**	**4.26**	**324**	**1.38**

3.2 Data sampling

Due to the large size of the dataset, which included over 447M tweets, we down-sample our dataset to choose and annotate a subset of the tweets for our task. To do this, we randomly selected several Twitter accounts from our dataset. We used an iterative sampling procedure to make sure that the representation sentiment was more evenly distributed. We carefully choose data from our original dataset for the sampling procedure, concentrating on terms like *face mask, vaccine, social distancing, quarantine, and lockdown*. By doing this, we make sure the picked tweets are relevant to the context of our analysis and offer insightful information about the sentiment people have toward these subjects. We eventually collected more than 2.2K tweets using this iterative sampling strategy. In addition, we only considered influential Twitter users, defined as accounts that are verified and or accounts that have more than 10K followers. Table 3.4 illustrates some examples of the tweets based on the theme that we have selected to sample our data.

Table 3.3: Data sampling examples on the topic of face mask, vaccine, social distancing, quarantine, and lockdown from 447M tweets.

Topic	**Tweet**
Social distancing	*Hi @esswhywon, We have passed your comment onto our COVID 19 Social Distancing Team. Many Thanks Steve*
Face mask	*The masks aren't 100% and some people do not wear them correctly nor do they have filters, this virus is TOUGH so a flimsy cheap face masks isn't as good as one with a filter -> https://mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-mask/art-20485449*
Lockdown	*The ones calling for lockdown, without risk or injury to themselves, should pay up.*
Quarantine	*Last year my neighborhood in SF had parents outside their apartments on the sidewalks handing out candy to kids. No door knocking. I'm kind of hoping that happens again this year, just with everyone wearing masks. Will have to quarantine the treats for 24 hours, of course.*
Vaccine	*@JKNC_ president and Lok Sabha MP Farooq Abdullah, who had tested positive for Covid-19 few days back, has been hospitalised. Abdullah had got first dose of Covid vaccine on March 2. @NewIndianXpress @TheMornStandard*

3.3 Task Description

Our primary goal in this study is to determine the author's intention during the composing of the tweet. We are looking for true emotions, whether positive, negative, or neutral that shaped the author's choice of word. This analysis will provide vital insights into the text's emotional underpinnings, helping us to better comprehend the motivations and goals behind the tweet and, as a result, contribute to a deeper understanding of social media communication.

Table 3.4: Examples of the intent of Influencers versus what the Tweet's polarity

Tweet	**True Intention**	**Tweet's Polarity**
166 labconfirmed cases of COVID19 have been reported in Lancaster County today, bringing the community total to 9,330. LNK Lancaster County COVID19 Update for October 29:	Neutral	Negative
21,600 doses of the AstraZeneca vaccine to arrive into the country this weekend.	Neutral	Positive

3.4 Data Annotation

To appropriately label the tweets, we used a gold-standard manual annotation effort. A team of two annotators worked on this, and in the event of a disagreement, a third annotator was consulted to ensure agreement. We use Cohen Kappa which is a statistical measure of inter-rater reliability that accounts for chance agreement. It is a widely used statistic for quantifying the level of agreement between two raters or measuring methods. For our experiment, the Cohen Kappa value was 0.46, which shows moderate agreement between two annotators for the annotation task. In Table 3.5, the question posed for our task and its related label distribution is displayed. The complicated nature of the sentiment in the tweets may be the cause of the challenge of achieving a high score of Cohen Kappa value. Emojis, slang, abbreviations, and casual language are frequently used in tweets, making it challenging to describe the meaning accurately. Additionally, some tweets may use complex language like sarcasm or irony that annotators may find difficult to understand some examples of conflict for this task are shown in Table 3.6.

Table 3.5: Questions asked from annotators during annotation: *What is the sentiment conveyed by the of the Tweet?* and label distribution

Total Tweet	**Label**	**Distribution**
2245	Positive	15.37%
	Negative	20.09%
	Neutral	64.54%

Table 3.6: Example of conflicts

Conflicting Tweet	**Annotator 1**	**Annotator 2**
Of the past 148 days, Ive spent nearly onethird of them in quarantine: in Beijing, Hong Kong and Tokyo. Traveling internationally during this global pandemic has been both exhausting and liberating. CNN	*Neutral*	*Negative*
Wearing face masks can help slow the spread of the COVID19 virus. Mask up. This is just one of many reasons why community members and employees wear masks. Share your WHY? Select submissions will be used.	*Positive*	*Neutral*
It will be crucial to wear face masks and practice physical distancing for a good while still. By failing to check the spread of the virus, Americas leaders have ensured that no vaccine not even a 95 effective one will quickly contain it.	*Negative*	*Positive*

Table 3.7: Data pre-processing examples

Original Tweet	**Pre-processed Tweet**
👏 👏 👏 👏 *Jay MBE! We salute you - King of the lockdown pub quiz. #QueensBirthdayHonours #MBE*	*: ::clapping_hands:: ::clapping_hands:: ::clapping_hands:: ::clapping_hands:: ::clapping_hands:: jay mbe! we salute you king of the lockdown pub quiz. queensbirthdayhonours mbe*
Hi @esswhywon, We have passed your comment onto our COVID 19 Social Distancing Team. Many Thanks Steve	*hi USER_HANDLE, we have passed your comment onto our covid 19 social distancing team. many thanks steve*
Hello, masks and social distancing are required in order to use public transit. We have been using more high capacity & extra buses on routes, and cleaning our buses with hospital-grade disinfectant, to mention a few. Please visit https://rtcsnv.com/coronavirus/.	*hello, masks and social distancing are required in order to use public transit. we have been using more high capacity amp; extra buses on routes, and cleaning our buses with hospitalgrade disinfectant, to mention a few. please visit URL_TOKEN.*

3.5 Data pre-processing

For our study, we classified the tweets into three categories: neutral, negative, and positive sentiments. We used a Python tool called Demoji [28] to remove the emoticons, and the hashtags were also taken out of the tweets. Additionally, we changed usernames to "USER_HANDLE" and URLs to "URL_TOKEN," replacing special characters with NULL characters. To preserve uniformity, the complete text of the tweets was converted to lowercase. These pre-processing methods attempted to clean the data and standardize the text for subsequent analysis. The steps are shown in Figure 3.2 and data pre-processing examples are given in Table 3.7.

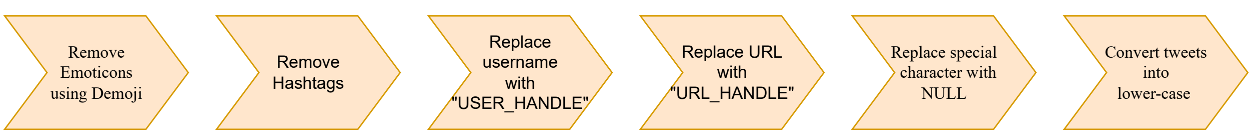

Figure 3.2: Data pre-processing

3.6 Models

For sentiment analysis, we employed a variety of Transformer models, including $BERT$[10], $\text{RoBERTa}_{\text{Twitter}}$[2], and a few shot learning models, like SetFit[40]. To assess how well our models perform compared to the current sentiment analysis techniques that we have mentioned in Chapter 2, we have also used packages like TextBlob and $VADER$. Additionally, we have used LLM models like GPT3.5 to infer the sentiment from the tweets. In this section, we will discuss the models and Python packages that we utilized for our experiments along with the parameters we used to tune our models.

3.6.1 BERT

$BERT$ is pre-trained using text data without labels. As a pre-trained deep bidirectional model, $BERT$ can benefit from enhancements made to the output layer, making it suitable for a range of tasks including sentiment analysis and language inference. $BERT$ uses Masked Language Model (MLM), where the MLM's objective is to estimate a masked word's original vocabulary id only based on its context after randomly masking some of the tokens from the input, reducing the requirement for unidirectionality and increasing its robustness for inference-based tasks. Pre-training and fine-tuning are the two stages of the $BERT$ model's operation. The model is trained on unlabeled data over various pre-training tasks. The pre-trained parameters are used to establish the $BERT$ model, and labeled data from the downstream tasks is used to fine-tune each parameter [10].

In our study, we made an essential change to the output layer of $BERT$, allowing the model to distinguish between and categorize tweet sentiment into three different categories namely positive, neutral, and negative. To do this, we carefully adjusted our models over the course of two epochs, using a batch size of 16 and a learning rate of 0.00001. This intentional and systematic approach allowed us to fully utilize $BERT$'s pre-trained architecture while adapting its capabilities to the unique requirements of our categorization work.

3.6.2 RoBERTa$_{\text{Twitter}}$

The Robustly Optimized BERT Pretraining Approach ($RoBERTa$) model is an enhanced version of the $BERT$ architecture that uses random tokens to predict hidden tokens in the input sequence. The model also features dynamic masking, which makes learning more efficient by changing the masking pattern dynamically during training [21].

In our experiment, we used the RoBERTa$_{\text{Twitter}}$ form of the $RoBERTa$ model, which was trained on 58M tweets and adjusted for sentiment analysis using the TweetEval benchmark [2], we fine-tuned this model using a learning rate of 0.00001, along with the batch size of 16 over two epochs.

3.6.3 SetFit

A branch of machine learning called few-shot learning involves working with training models to learn new tasks with a few numbers of labeled data. The main goal is to make it possible for models to extrapolate from a small sample size or instance of a given class or activity to make precise predictions on brand-new, unforeseen occurrences. We utilized SetFit as a few-shot learning model because it has the benefit of not requiring any specially created prompts. SetFit has the additional benefit of requiring less training time than other few-shot models. Sentence Transformers' capacity to produce dense embeddings based on paired sentences is utilized by SetFit. It utilizes limited labeled input data during the initial fine-tuning phase stage by contrastive training, where positive and negative pairs are produced through in-class and out-of-class selection. Following training on these pairs, the Sentence Transformer model produces dense vectors for each sample. The classification head is trained on the encoded embeddings with the corresponding class labels in the second stage. The unseen example is supplied into the refined Sentence Transformer at the time of inference, which creates an embedding that, when fed to the classification head, produces a class label prediction [40].

Eight different samples for each sentiment category were used as the training sample. Our model was trained using these examples, with a batch size of 16 samples for each iteration. We carried out a total of 20 iterations within a single epoch, methodically tweaking our model's comprehension of numerous sentiment categories.

3.6.4 TextBlob

TextBlob is a Python-based open-source NLP program built on the Natural Language Toolkit ($NLTK$) toolkit. It can be used for a variety of NLP tasks, including sentiment analysis, part-of-speech tagging, and noun phrase extraction, among others. For sentiment analysis, it assigns a polarity score to a particular text, and the sentiment of that text can be inferred from the polarity score [23].

In our experiment, we used a sentiment analysis package from TextBlob to evaluate the tweet, which gave us a polarity score reflecting the sentiment attached to the text. Any sentiment that had a polarity score less than 0.0 was considered to be negative. In contrast, we classified a score as positive if it was higher than 0.0. The tweet was classified as having a neutral sentiment when the polarity score was equal to 0.0, indicating that there were no particularly strong positive or negative emotions present in the text. One such example is provided in Figure 3.3.

```
tweet = " : all of us, when quarantine is over"
TextBlob(tweet).sentiment.polarity
```
```
0.0
```

Figure 3.3: TextBlob execution

3.6.5 VADER

$VADER$ is another Python-based NLP tool that uses the $NLTK$ package as its foundation similar to TextBlob. $VADER$ is pre-trained on lexical and grammatical heuristics as well as sentiment-rich lexical features to analyze sentiments expressed from social media such as Tweets. The polarity and intensity of the sentiment in the tweets can be evaluated using this tool [15].

Similar to TextBlob, we evaluated the tweet in our experiment using a sentiment analysis package from $VADER$, which provided us with a compound polarity score expressing the sentiment associated with the text. Positive tweets are those that have a compound score greater than 0.5, neutral tweets are those that have a score between -0.5 and 0.5, and negative tweets are those that have a compound score less than -0.5. Figure 3.4 illustrates how $VADER$ assigns polarity to a tweet.

```
tweet = " : all of us, when quarantine is over"
sia.polarity_scores(tweet)
```

```
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Figure 3.4: VADER execution

3.6.6 GPT-3.5

GPT-3.5 improves its inferencing capabilities through reinforcement learning from human feedback [8]. The model learns to generate responses that are more closely aligned with human expectations and intended outcomes through trial and error. GPT-3.5 also leverages prompt engineering techniques to aid in reliable inference. These prompts are carefully crafted to elicit significant insights or predictions from the model's vast knowledge base. GPT-3.5 excels in a wide range of inference tasks due to the combination of reinforcement learning and strategic prompt engineering, making it a versatile and powerful tool for natural language understanding and creation [4, 13, 14].

In our experimentation with GPT-3.5, we harnessed the power of prompt engineering to effectively discern sentiments, employing a temperature value of 0.0, which means that the randomness of the inference decreases. These precisely prepared prompts allow us to query the model with maximum specificity, employing both zero-shot and few-shot prompting examples to derive subtle information. Zero-shot prompting is a NLP strategy in which a model is given a prompt and instructed to provide a response without any extra training on that particular task. Additionally, few-shot prompting is a method that falls somewhere between full fine-tuning of language models and zero-shot prompting. Few examples in this technique enable the model to specialize and adapt to the task/domain, without needing the thousands of examples necessary for thorough fine-tuning. We carefully provided a collection of examples that represented various classes in order to effectively adjust this model to our task-specific requirements, allowing it to fit effortlessly into the framework of our specific task. Table 3.8 represents the prompt we used in our experiment.

Table 3.8: Prompts used for experiment

GPT3.5 Model	Prompts
Zero-shot	*Classify the sentiment of the author behind the tweet as "Negative", "Neutral", or "Positive".* *tweet: {tweet}*
Few-shot	*Classify the sentiment of the author behind the tweet as "Negative", "Neutral", or "Positive".* *Label the sentiments as follows:* *1. Sky News host says the coronavirus authoritarianism* *has become evil after a young family in the US was forced off a plane because their crying* *twoyearold daughter would not wear a face mask. ->Negative* *2. COVID19 FEMA Quarantine camps on the horizon? SATAN SOLDIERS* *AT THE CDC ARE PREPARING QUARANTINE CAMPS! Trump MAGA TWGRP TheMighty100 ->Negative* *3. Has Anyone Noticed That the Worlds Billionaires During the Covid Fraud Are Getting Richer While* *Everyone Else Is Facing Poverty? ->Negative* *4. With the impact of COVID and the rapid digitalization of B1B, the need for personalized interactions with buyers* *has dramatically increased. Find out whats working in personalization ::down_arrow:: ->Neutral* *5. Over 14 lakh tests were conducted for detection of coronavirus in last 14 hours ->Neutral* *6. LIVE Victoria has recorded zero new local cases of coronavirus or new cases in hotel quarantine. ->Neutral* *7. But yes livin the lockdown dream. ::penguin:: ::wine_glass:: ::sailboat:: ->Positive* *8. London may be in nearlockdown conditions, but it still looked beautifully festive on Christmas night ::snowflake::* *not to mention the traffic! ->Positive* *9. Bills paid, Im my own boss, booked weeks in advance, corona free, been drinking my water and Im around a lot of* *genuine love. Its a few bumps in the road. But Im blessed.->Positive* *tweet: {tweet}*

Chapter 4

Experiment and Result

We will present a complete summary of our experimental setup workflow in this section, providing light on the meticulous procedures and techniques we used to perform our research. Furthermore, we will go over the findings of our experiments in detail. Our investigation will include the results and will shed light on how the model was used to reach our conclusion.

4.1 Datasets

We used three datasets to set up our experiments, from which we constructed one dataset to analyze the sentiment of tweets about COVID-19, which we are referring to as DS1 that was included in Chapter 3.2. We then used two more datasets from Hugging Face to train and test our models which we have labeled as DS2 [3] and DS3 [11]. The Table 4.1 contains comprehensive information about the dataset. We used 80% of the original data from all datasets for training our models, and the remaining 20% was used for testing.

Table 4.1: Datasets for experiment

Dataset	**Size**				**Description**
	Negative	**Neutral**	**Positive**	**Total**	
DS1	451	1449	345	2245	COVID-19 Tweets
DS2	903	903	903	2709	Tweets
DS3	8782	12548	8968	30298	Tweets

4.2 Experimental Models

We have trained numerous models throughout our experiments, including $\text{RoBERTa}_{\text{Twitter}}$, $BERT$, and SetFit. To train and test our models, we employed DS1, DS2, and DS3 and observed that DS1 produced the best results for our sentiment analysis task, as shown in Table 4.2.

Table 4.2: Experimental Result using all datasets

Train	Test	BERT				RoBERTa$_{\text{Twitter}}$				SetFit			
		P	**R**	**A**	**F1**	**P**	**R**	**A**	**F1**	**P**	**R**	**A**	**F1**
DS1	DS1	**0.77**	**0.77**	**0.77**	**0.77**	**0.84**	**0.84**	**0.84**	**0.84**	**0.79**	**0.78**	**0.78**	**0.78**
	DS2	0.57	0.47	0.47	0.43	0.76	0.72	0.72	0.77	0.70	0.68	0.68	0.68
	DS3	0.52	0.52	0.52	0.52	0.71	0.71	0.71	0.71	0.66	0.64	0.64	0.63
DS2	DS1	0.73	0.51	0.51	0.53	0.75	0.63	0.63	0.65	0.70	0.64	0.64	0.65
	DS2	0.65	0.64	0.64	0.63	0.75	0.75	0.75	0.75	0.71	0.71	0.71	0.71
	DS3	0.59	0.51	0.51	0.47	0.71	0.69	0.69	0.67	0.69	0.69	0.69	0.68
DS3	DS1	0.67	0.63	0.63	0.65	0.71	0.69	0.69	0.70	0.70	0.71	0.71	0.70
	DS2	0.61	0.57	0.57	0.58	0.71	0.67	0.67	0.68	0.68	0.60	0.60	0.61
	DS3	0.78	0.78	0.78	0.78	0.81	0.81	0.81	0.81	0.78	0.77	0.77	0.77

We discovered through this experiment that our fine-tuned models perform better on DS1 since we are looking at the author's sentiment on COVID-19 as expressed in a tweet. The other two datasets we used merely looked at the sentiment from the tweet itself and did not take into account how the author feels; thus, utilizing DS1 and the fine-tuned models, we achieved a better result for our particular task.

We then selected DS1 and ran experiments on the models, $BERT$, RoBERTa$_{\text{Twitter}}$, and SetFit. We also used TextBlob and $VADER$, as we have seen in the earlier work by other researchers, we intended to compare our fine-tuned model performance with widely used Python toolkits like TextBlob and $VADER$. The result of this experiment is given in table Table 4.3. RoBERTa$_{\text{Twitter}}$ produces the highest value of the F1-score for all labels. SetFit, on the other hand, performs marginally worse, with the F1-score of 0.65, 0.86, and 0.74 for negative, neutral, and positive labels respectively. $BERT$ exhibits competitive performance in sentiment analysis but does not outperform the outcomes of the RoBERTa$_{\text{Twitter}}$ model. In this experiment, we have seen that using TextBlob and $VADER$ the performance of the models is not as good as our fine-tuned models as both TextBlob and $VADER$ analyze the tweet based on the polarity of the text and hence cannot take into account the author's sentiment. Additionally, we have used LLM models, GPT-3.5. We have utilized both zero-shot and few-shot prompting techniques for GPT-3.5 to complete this task. Our experiments show that fine-tuned RoBERTa$_{\text{Twitter}}$ performs better than GPT-3.5 models.

Table 4.3: Performance of Algorithms using task-specific dataset

Label	BERT				RoBERTa$_{\text{Twitter}}$				SetFit				TextBlob				VADER				Zero-Shot GPT3.5				Few-Shot GPT3.5			
	P	R	A	F1	P	R	A	F1	P	R	A	F1	P	R	A	F1	P	R	A	F1	P	R	A	F1	P	R	A	F1
Negative	0.68	0.53	0.53	0.60	**0.77**	0.68	0.67	**0.72**	0.64	0.66	0.66	0.65	0.32	0.31	0.31	0.31	0.37	0.51	0.51	0.43	0.47	**0.73**	**0.73**	0.57	0.49	0.66	0.66	0.56
Neutral	0.80	**0.93**	**0.93**	0.86	0.86	0.91	0.91	**0.89**	0.86	0.86	0.87	0.86	0.77	0.32	0.32	0.46	0.84	0.33	0.33	0.48	**0.88**	0.54	0.54	0.67	0.80	0.71	0.71	0.75
Positive	0.76	0.46	0.46	0.58	**0.81**	0.75	0.75	**0.78**	0.75	0.72	0.72	0.74	0.25	0.86	0.86	0.38	0.27	0.83	0.83	0.41	0.48	**0.88**	**0.88**	0.62	0.60	0.61	0.61	0.60

Table 4.4: COVID-19 keywords for 6 topics

Topic	Keywords
Face Mask	*mask, face, N95, KN95, covering, nose wire, adjustable strap, breathable fabric*
Travel	*travel, border, closures, hotel, passport*
WFH	*wfh, home, remote, hotel, telecommuting, virtual, video, zoom*
Social Distancing	*social, distancing, six feet, physical, close contact*
Lockdown & Quarantine	*lockdown, stay home, curfew, restricted movement, quarantine, isolation*
Vaccine	*vaccine, vaccination, vaccinated, pfizer, moderna, astraZeneca, johnson, sinopharm, sputnik*

We have observed an interesting pattern when looking into the result from this experiment, we have seen that using zero-shot prompting GPT-3.5 was able to infer the classes reasonably, and for neutral labels it was not able to perform well however when using few-shot prompting, GPT-3.5 it was able to achieve a higher F1-score for the neutral class, which indicates that few-shot prompting approach can perform better on neutral class compare to the zero-shot approach which is also observed by the authors in [43].

4.3 Sentimental Trend Analysis

Additionally, to do a sentimental trend investigation, we also wanted to conduct trend analyses for individuals as well as organizations to examine how the sentiment changed between September 2020 and October 2021 for COVID-19 related tweets. To observe these trends, we built a new dataset for a selection of topics, which is shown in Table 4.4. We employed a coarse-grain method to capture every significant phrase in this dataset by using comparable words related to a topic. Our top-performing model, RoBERTa$_{\text{Twitter}}$ was used to categorize the tweets' sentiments, and a general pattern was found.

The pattern for the sentimental shift for both individuals and organizations for all of the topics specified in Table 4.4 during 14 months is shown by the Figure 4.1, Figure 4.2, Figure 4.3. Among all the topics, the least positive sentiments were conveyed during the timeframe compared to neutral and negative sentiments. From mid-July to mid-August 2021, there was an increase in the number of negative tweets on all topics, this may be caused by the increase in Delta variant cases, which included numerous hospitalizations and fatalities during this period [12]. As seen in Figure 4.1, the topic of face masks consistently received more negative sentiment than the other topics. Travel was the most neutral topic of discourse among influencers on Twitter for neutral sentiment, whereas face masks were the least neutral subject, as demonstrated in Figure 4.2. The influencers expressed the most positive sentiment for the topic of working from home as shown in Figure 4.3.

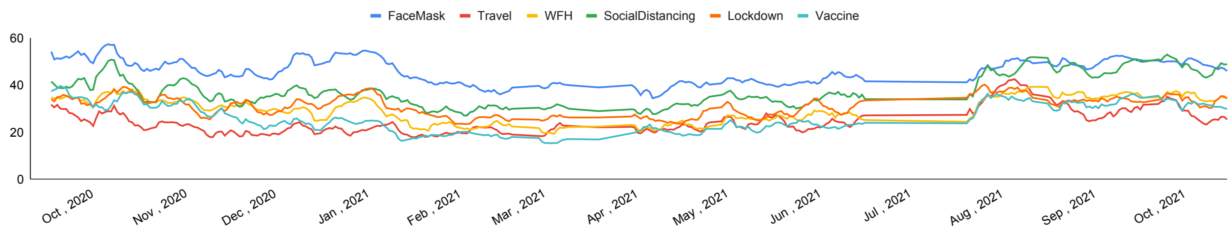

Figure 4.1: Negative Trend on Significant Topics Related to COVID-19

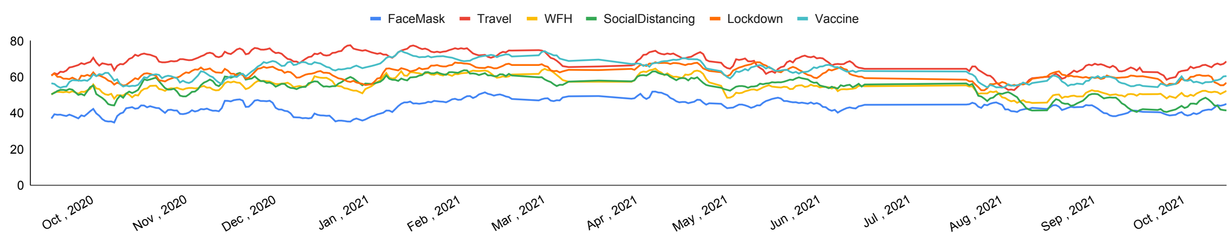

Figure 4.2: Neutral Trend on Significant Topics Related to COVID-19

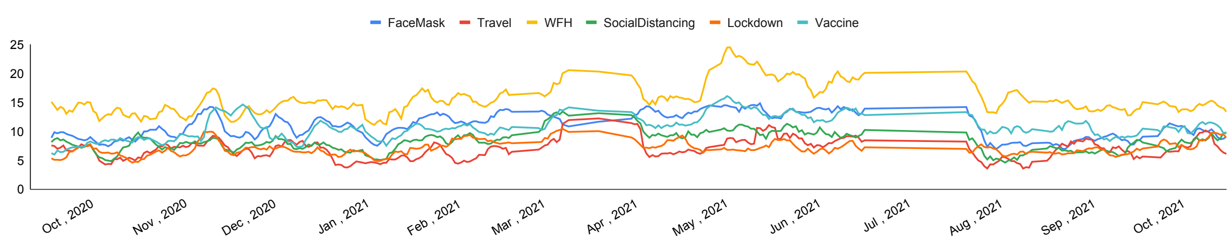

Figure 4.3: Positive Trend on Significant Topics Related to COVID-19

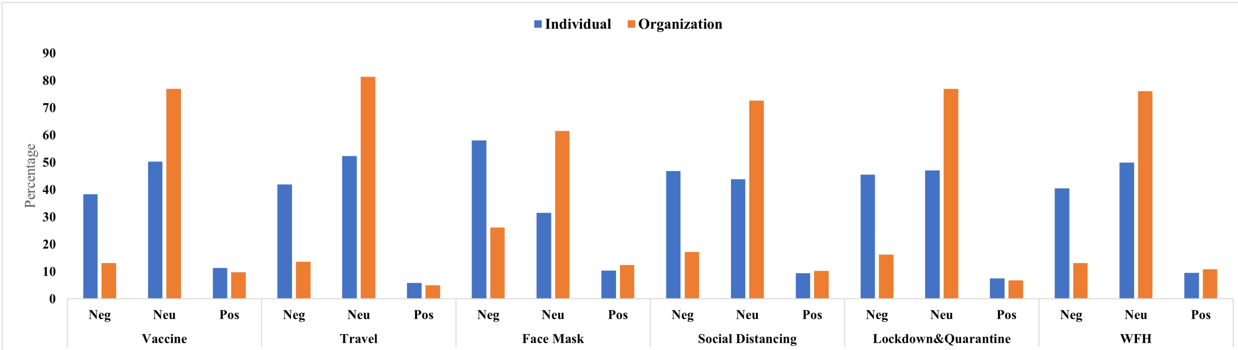

Figure 4.4: Individual versus Organization Sentiment Polarity on Significant Topics Related to COVID-19.

4.3.1 Individual versus Organization's Sentimental Analysis

In our earlier research, we divided influencers into the categories of individual and organization, where we defined an individual as someone who manages their own account while an organization was defined as an account managed by a company, business, brand, non-profit, government agency, or other organization [36][1]. Figure 4.4 depicts the polarity of sentiments expressed by influencers over a 14-months period on six crucial themes provided in Table 4.4. The graph reveals that influencers, including individuals and organizations, predominantly express neutral feelings for all key topics, indicating that influencers are primarily focused on distributing information to their followers during a crisis. Individuals tend to display more negative attitudes than organizations, as organizations represent a collective entity with goals, objectives, and brand images. They often maintain a professional demeanor to avoid negativity in discussions. On the other hand, individuals are more subjective towards their feelings. Furthermore, it can be noted that the topic of face masks is rated the least positively by both individuals and organizations. There were instances of misinformation and conspiracy theories involving face masks throughout the pandemic and influencers may have been afraid to market face masks for fear of endorsing or spreading inaccurate information.

[1] Under Review by The 16th ACM International WSDM Conference

Chapter 5

Conclusion and Future Work

We investigated the COVID-19-related tweets in this research work with the main goals of identifying the authors' sentiments and conducting an extensive trend analysis that included both organizations and individuals. Notably, our research takes a novel approach by concentrating not just on the sentiment expressed in the tweets but also on the underlying emotions that the authors were feeling when they composed the tweets. To achieve this, we harnessed the power of transformer models, including the cutting-edge GPT-3.5, a unique feature previously unexplored in the context of COVID-19 sentiment analysis within the existing literature. Through careful testing, we discovered that the fine-tuned transformer model $\text{RoBERTa}_{\text{Twitter}}$ performed best, attaining an F1-score of 84%. Additionally, this study demonstrates significant disparities between individuals and organizations when it comes to expressing sentiment. Individuals, in contrast to organizations, take more subjective positive or negative attitudes, according to the analysis. Organizations, on the other hand, broadcast a greater volume of Tweets overall. The current research represents a substantial advancement in our knowledge of the complex sentimental connection that surrounds the COVID-19 discourse on social media.

In our future research initiatives, we intend to conduct a thorough examination of the cutting-edge LLM. To facilitate this research, we will first focus on fine-tuning GPT-3.5 with our dataset, DS1. This decision is motivated by cost-effectiveness, as fine-tuning GPT-3.5 is a more cost-effective option than the more recent GPT-4.0. We expect that this fine-tuning effort will produce better results for our particular task. Furthermore, for inference tasks, we intend to leverage the capability of GPT4.0, which benefits from training on the most recent data and may outperform GPT-3.5 in inference-based tasks. To widen our perspective, we aim to investigate different LLMs, such as Llama-2, and evaluate their effectiveness on our task-specific dataset.

Bibliography

[1] Gil Appel, Lauren Grewal, Rhonda Hadi, and Andrew T Stephen. The future of social media in marketing. *Journal of the Academy of Marketing science*, 48(1):79–95, 2020.

[2] Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: Empirical Methods in Natural Language Processing(EMNLP) 2020*, pages 1644–1650, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.148.

[3] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266, Marseille, France, June 2022. European Language Resources Association.

[4] Samuel Bestvater. Using machine learning to infer real-world political attitudes and behaviors from social media data. Indiana University Workshop in Methods, 2023.

[5] Koyel Chakraborty, Surbhi Bhatia, Siddhartha Bhattacharyya, Jan Platos, Rajib Bag, and Aboul Ella Hassanien. Sentiment analysis of covid-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, 97:106754, 2020.

[6] Ganesh Chandrasekaran and Jude Hemanth. Deep learning and textblob based sentiment analysis for coronavirus (covid-19) using twitter data. *International Journal on Artificial Intelligence Tools*, 31(01):2250011, 2022.

[7] Sungwoon Choi, Jangho Lee, Min-Gyu Kang, Hyeyoung Min, Yoon-Seok Chang, and Sungroh Yoon. Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks. *Methods*, 129:50–59, 2017.

[8] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30, 2017.

[9] Salvador Contreras Hernández, María Patricia Tzili Cruz, José Martín Espínola Sánchez, and Angélica Pérez Tzili. Deep learning model for covid-19 sentiment analysis on twitter. *New Generation Computing*, 41(2):189–212, 2023.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[11] Hugging Face. Tweet sentiment extraction. https://huggingface.co/datasets/mteb/tweet_sentiment_extraction.

[12] Centers for Disease Control and Prevention. Covid data tracker. https://covid.cdc.gov/covid-data-tracker. (Accessed on 04/23/2023).

[13] Saadia Gabriel, Skyler Hallinan, Maarten Sap, Pemi Nguyen, Franziska Roesner, Eunsol Choi, and Yejin Choi. Misinfo reaction frames: Reasoning about readers' reactions to news headlines. *arXiv preprint arXiv:2104.08790*, 2021.

[14] Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597*, 2023.

[15] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 216–225, 2014.

[16] Ali Shariq Imran, Sher Muhammad Daudpota, Zenun Kastrati, and Rakhi Batra. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. *IEEE Access*, 8:181074–181090, 2020.

[17] Hyeju Jang, Emily Rempel, Giuseppe Carenini, and Naveed Janjua. Exploratory analysis of covid-19 related tweets in north america to inform public health institutes. *arXiv preprint arXiv:2007.02452*, 2020.

[18] Harleen Kaur, Shafqat Ul Ahsaan, Bhavya Alankar, and Victor Chang. A proposed sentiment analysis deep learning algorithm for analyzing covid-19 tweets. *Information Systems Frontiers*, pages 1–13, 2021.

[19] Rijwan Khan, Piyush Shrivastava, Aashna Kapoor, Aditi Tiwari, and Abhyudaya Mittal. Social media analysis with ai: sentiment analysis techniques for the analysis of twitter covid-19 data. *Journal of Critical Reviews*, 7(9):2761–2774, 2020.

[20] Siru Liu and Jialin Liu. Public attitudes toward covid-19 vaccines on english-language twitter: A sentiment analysis. *Vaccine*, 39(39):5499–5505, 2021.

[21] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[22] Christian E Lopez, Malolan Vasu, and Caleb Gallemore. Understanding the perception of covid-19 policies by mining a multilanguage twitter dataset. *arXiv preprint arXiv:2003.10359*, 2020.

[23] Steven Loria. Textblob: Simplified text processing. https://textblob.readthedocs.io/en/dev/, 2020.

[24] Yuesheng Luo and Mayank Kejriwal. Understanding covid-19 vaccine reaction through comparative analysis on twitter. In *Intelligent Computing: Proceedings of the 2022 Computing Conference, Volume 1*, pages 846–864. Springer, 2022.

[25] Kamaran H Manguri, Rebaz N Ramadhan, and Pshko R Mohammed Amin. Twitter sentiment analysis on worldwide covid-19 outbreaks. *Kurdistan Journal of Applied Research*, pages 54–65, 2020.

[26] Amrita Mishra, Mohd Saif Wajid, and Upasana Dugal. A comprehensive analysis of approaches for sentiment analysis using twitter data on covid-19 vaccines. *Journal of Informatics Electrical and Electronics Engineering (JIEEE)*, 2(2):1–10, 2021.

[27] Oladapo Oyebode, Chinenye Ndulue, Ashfaq Adib, Dinesh Mulchandani, Banuchitra Suruliraj, Fidelia Anulika Orji, Christine T Chambers, Sandra Meier, Rita Orji, et al. Health, psychosocial, and social issues emanating from the covid-19 pandemic based on social media comments: text mining and thematic analysis approach. *JMIR Medical Informatics*, 9(4): e22734, 2021.

[28] Python Package Index. demoji 1.1.0. https://pypi.org/project/demoji/, March 2016.

[29] Miftahul Qorib, Timothy Oladunni, Max Denis, Esther Ososanya, and Paul Cotae. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on covid-19 vaccination twitter dataset. *Expert Systems with Applications*, 212:118715, 2023.

[30] Nikhil Kumar Rajput, Bhavya Ahuja Grover, and Vipin Kumar Rathi. Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. *arXiv preprint arXiv:2004.03925*, 2020.

[31] Ghulam Musa Raza, Zainab Saeed Butt, Seemab Latif, and Abdul Wahid. Sentiment analysis on covid tweets: an experimental analysis on the impact of count vectorizer and tf-idf on sentiment predictions using deep learning models. In *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, pages 1–6. IEEE, 2021.

[32] Syafrida Hafni Sahir, Raden Sri Ayu Ramadhana, Muhammad Fauzi Romadhon Marpaung, Shabrina Rasyid Munthe, and Ronal Watrianthos. Online learning sentiment analysis during the covid-19 indonesia pandemic using twitter data. In *IOP Conference Series: Materials Science and Engineering*, volume 1156, page 012011. IOP Publishing, 2021.

[33] Jim Samuel, GG Md Nawaz Ali, Md Mokhlesur Rahman, Ek Esawi, and Yana Samuel. Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6):314, 2020.

[34] Naw Safrin Sattar and Shaikh Arifuzzaman. Covid-19 vaccination awareness and aftermath: public sentiment analysis on twitter data and vaccinated population prediction in the usa. *Applied Sciences*, 11(13):6128, 2021.

[35] Broadband Search. How has the internet held up during the covid-19 pandemic? the story so far. https://www.broadbandsearch.net/blog/internet-story-covid19-pandemic, 2023. (Accessed on 05/29/2023).

[36] Hossein Shirazi, Noushin Salek Faramarzi, Thomas Harris, Shadaab Kawnain Bashir, Ashmita Shishodia, Indrakshi Ray, and Hajar Homayouni. Investigating influential covid-19 perspectives: A multifaceted analysis of twitter discourse. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 2023.

[37] Rameshwer Singh, Rajeshwar Singh, and Ajay Bhatia. Sentiment analysis using machine learning technique to predict outbreaks and epidemics. *International Journal of Advanced Science and Research*, 3(2):19–24, 2018.

[38] Shane Smith. Coronavirus (covid19) tweets. https://www.kaggle.com/datasets/smid80/coronavirus-covid19-tweets.

[39] Dandan Tao, Pengkun Yang, and Hao Feng. Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety*, 19(2):875–894, 2020.

[40] Lewis Tunstall, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat, and Oren Pereg. Efficient few-shot learning without prompts. *arXiv preprint arXiv:2209.11055*, 2022.

[41] Thirumalaisamy P Velavan and Christian G Meyer. The covid-19 epidemic. *Tropical Medicine & International Health*, 25(3):278, 2020.

[42] Tanmay Vijay, Ayan Chawla, Balan Dhanka, and Purnendu Karmakar. Sentiment analysis on covid-19 twitter data. In *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–7, 2020. doi: 10.1109/ICRAIE51050.2020.9358301.

[43] Zengzhi Wang, Qiming Xie, Zixiang Ding, Yi Feng, and Rui Xia. Is chatgpt a good sentiment analyzer? a preliminary study. *arXiv preprint arXiv:2304.04339*, 2023.

[44] WHO:World Health Organization. Coronavirus disease (covid-19) pandemic. https://www.who.int/europe/emergencies/situations/covid-19, 2023. (Accessed on 04/04/2023).

[45] WHO:World Health Organization. Coronavirus disease (covid-19) pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019, 2023. (Accessed on 04/04/2023).

[46] Stanisław Wrycza and Jacek Maślankowski. Social media users' opinions on remote work during the covid-19 pandemic. thematic and sentiment analysis. *Information Systems Management*, 37(4):288–297, 2020.

[47] Jianlong Zhou, Shuiqiao Yang, Chun Xiao, and Fang Chen. Examination of community sentiment dynamics due to covid-19 pandemic: A case study from australia. 2020.