M.Sc. Engg. (CSE) Thesis

A BIMODAL LONGITUDINAL INVESTIGATION ON CHANGES IN SENTIMENTS OVER SOCIAL MEDIA INTERACTIONS OWING TO COVID-19 PANDEMIC

Submitted by

Md. Saidul Hoque Anik

0417052044

Supervised by

Dr. A. B. M. Alim Al Islam

Submitted to

Department of Computer Science and Engineering

Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

August 2022

Candidate's Declaration

I, do, hereby, certify that the work presented in this thesis, titled, "A BIMODAL LONGITUDINAL INVESTIGATION ON CHANGES IN SENTIMENTS OVER SOCIAL MEDIA INTERACTIONS OWING TO COVID-19 PANDEMIC", is the outcome of the investigation and research carried out by me under the supervision of Dr. A. B. M. Alim Al Islam, Professor, Department of CSE, BUET.

I also declare that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Md. Saidul Hoque Anik

Md. Saidul Hoque Anik

0417052044

The thesis titled "**A BIMODAL LONGITUDINAL INVESTIGATION ON CHANGES IN SENTIMENTS OVER SOCIAL MEDIA INTERACTIONS OWING TO COVID-19 PANDEMIC**", submitted by Md. Saidul Hoque Anik, Student ID 0417052044, Session April 2017, to the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, has been accepted as satisfactory in partial fulfilment of the requirements for the degree of Master of Science in Computer Science and Engineering and approved as to its style and contents on August 31, 2022.

Board of Examiners

1.	______________________ Dr. A. B. M. Alim Al Islam Professor Department of CSE, BUET, Dhaka	Chairman (Supervisor)
2.	______________________ Dr. Mahmuda Naznin Professor and Head Department of CSE, BUET, Dhaka	Member (Ex-Officio)
3.	______________________ Dr. Anindya Iqbal Professor Department of CSE, BUET, Dhaka	Member
4.	______________________ Dr. Sadia Sharmin Assistant Professor Department of CSE, BUET, Dhaka	Member
5.	______________________ Dr. Md. Mahbubur Rahman Professor Department of CSE Military Institute of Science and Technology (MIST), Dhaka, Bangladesh	Member (External)

Acknowledgement

To Almighty Allah be the glory for allowing me to come up with this thesis dissertation. Thanks to my mentor and supervisor, Dr. A. B. M. Alim Al Islam, for his unfailing encouragement and guidance, I have been able to complete this work successfully. I am thankful to the respected members of my Board of Examiners Dr. Mahmuda Naznin, Dr. Anindya Iqbal, Dr. Sadia Sharmin, and Dr. Md. Mahbubur Rahman.

I am also eternally obliged to my spouse for her selflessness and assistance, particularly amid the challenges of her motherhood. My dissertation would not have been complete without the nice comments and helpful remarks I received from the respected faculties on my board of examiners.

Dhaka
August 31, 2022

Md. Saidul Hoque Anik
0417052044

Contents

Candidate's Declaration i

Board of Examiners ii

Acknowledgement iii

List of Figures vii

List of Tables viii

List of Algorithms ix

Abstract x

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Motivation behind This Study on Sentiment Analysis . . . 2
  - 1.2.1 Roles of Social Media . . . 2
  - 1.2.2 Sentiment Analysis in Social Media . . . 2
  - 1.2.3 Scopes of Sentiment Analysis . . . 3
  - 1.2.4 COVID-19 Sentiments in Twitter . . . 4
- 1.3 Our Research Questions . . . 5
- 1.4 Our Approach . . . 6
- 1.5 Organization of This Study . . . 7
- 1.6 Research Ethics and Anonymization . . . 7

2 Related Work 8
- 2.1 Approaches to Sentiment Analysis . . . 8
  - 2.1.1 Location and Topic-based Study . . . 8
  - 2.1.2 Knowledge Graph and Frameworks . . . 9
  - 2.1.3 Image Sentiment Detection . . . 9
  - 2.1.4 Change-point Detection . . . 10
  - 2.1.5 Analysis Duration . . . 10

Dataset Crafting 10
Summary 10
2.2 Research Gaps 11

3 Methodology **13**
Data Collection 13
Sentiment Detection 15
Text Sentiment Classification 15
Image Sentiment Classification 20
User Sentiment Group Migration 20
Plotting and Change-point Detection 21
CUSUM (Cumulative Sum) 22
PELT Search 22
Dynamic Programming (DP) Search 23
Series Decomposition 23
Effect Size of Sentiment Change 23
Correlation between Text and Image Sentiment 24
Settling Trend Identification 24
LOWESS Smoothing Filter 24
DTW Calculation 24
Series Reconstruction 25
Prophet 25
Orbit 25
Effect of Smoothing Filter for Forcasting 25

4 Findings **26**
Sentiment Library Evaluation and Heuristic Development 27
User Sentiment Group Migration 27
Change-point Identification 30
Text Sentiments 30
Image Sentiments 31
Effect Size of Sentiment Change 31
Settling Trend Identification 32
Correlation between Text and Image Sentiment 32
Series Reconstruction 33
Text Sentiments 33
Image Sentiments 34
Effect of Smoothing Filter 35

5 Discussion **37**

Quantification of Sentiment Change ... 37

Text Sentiment ... 37

Image Sentiment ... 37

Effect Size of Change for Sentiments ... 38

Correlation in Text and Images ... 38

Settling Trend and Graph Reconstruction ... 38

Threats to Validity ... 39

Implications of Our Study ... 40

Comparison with Other Similar Research Studies ... 41

6 Future Work **42**

Image Captioning ... 42

Topic Modeling ... 42

Geographical Analysis ... 43

7 Conclusion **44**

References **45**

List of Figures

2.1 History of pandemic alongside the emergence of social media. Image Courtesy: #AFPGraphics [1] and WidSix [2] 9

Steps of our methodology 14
Listing users based on days active in Twitter 15
Active User Selection Process 16

Migration of users between various sentiments in 2020........ 26
Change-point (marked using red vertical line) for Text and Image Sentiments . 31
Settling trend after the change-point for six time series 32
Three correlation coefficients (left: Pearson's relation, middle: Spearman's relation, and right: Kendall's relation) denote similarity patterns between text and image sentiments over eight quadrants (2019 to 2020). Each color represents the corresponding sentiment (green for positive, gray for neutral, and red for negative). 33
Series Reconstruction using Prophet and Orbit library 34

List of Tables

Baseline text sentiment detection libraries 17
Six time series generated from the posts of 569 users spanning over 724 days . 22
Precision and Recall score of the five libraries and the three heuristics used 28
Evaluation of five popular libraries and a few heuristics based on human perspective. The best and second-best results are highlighted in green and yellow color 29
Change-points detected by the three algorithms and the final change-point considered 29
Effect size in sentiment trend after the change-points and the interpretation 30
Effect of various smoothing algorithms on series forecasting 35

5.1 Interpretation of Pearson's r correlation between text and image sentiments over the quarters 39

List of Algorithms

1 Majority voting on all five classifiers (Heuristic-1) 19
2 Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-1 (Heuristic-2) 20
3 Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-2 (Heuristic-3) 21

Abstract

Changes in human sentiments over cyberspace with the emergence of a pandemic were unheard of before COVID-19, given that the last recorded pandemic occurred decades before interactive cyberspace existed. Accordingly, the opportunities and dimensionalities pertinent to the human sentiments and their changes that surfaced over social media interactions demand an in-depth analysis, which we perform in this study. Existing related research studies found to date focus on analyzing sentiments covering only text-based social media posts with limited contexts. These studies generally use out-of-the-box libraries to classify sentiments, which are often not suited for social media posts as they possess different styles compared to regular text bodies. To go beyond these studies, first, we collect public thoughts and images shared on Twitter by those who showed their interest in COVID-19. We then explore the existing sentiment classifier libraries and their potential blend for developing a new classification technique to better analyze sentiments over text-based tweets. Afterward, we perform exploratory data analysis on these collected thoughts and images to find the patterns inherently embedded within these changes of sentiments, owing to the COVID-19 outbreak, expressed over social media. Our findings through a bimodal investigation subsuming both text and images reveal a correlation between the two modalities of expression, pointing to changes in sentiments over two years spanning pre- and during COVID phases, identifying change-points for each type of sentiment during the different phases, etc. These findings unveil new dimensionalities of human interactions over cyberspace during a pandemic period.

Chapter 1

Introduction

The emergence of a global pandemic and its widespread transmissions is exhibiting far-reaching consequences in almost every field including social, economic, political, and virtual landscapes. Faced with a crisis that has rendered billions of people around the world unable to safely commute outside of their homes and interact physically, we have been experiencing a parallel increase in human interactions over cyberspaces in recent years. Such a high level of human-computer interactions over virtual spaces was never seen before in history, which opens up new avenues for investigations. As a result, in-depth analyses are required to comprehend the interactions so that relevant entities can strategize future decisions and implement efficient policies based on the findings of the analyses. In this context, in this study, We conduct mixed-method research on patterns inherently embedded within changes of human sentiments expressed over cyberspaces in the light of COVID-19.

Background

In late December 2019, a novel coronavirus causing severe respiratory diseases such as pneumonia was discovered in Wuhan, China. Since then, the disease has spread globally, resulting in an ongoing pandemic. The World Health Organization (WHO) has named the virus that caused the infection Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) and the disease that resulted from the infection as *CO*rona *VI*rus *D*isease 2019 or *COVID*-19. COVID-19 is classified as an airborne disease, with several variants varying in contagiousness. The global spread of this disease can be tracked using the available dashboards, which are updated daily based on WHO reports.

Fortunately, at least a third of people who are infected do not develop noticeable symptoms. Most develop mild to moderate symptoms (fever, headache, and up to mild pneumonia), while some (14%) develop severe symptoms that involve lung infections. Since this disease is caused by a viral infection, antibiotics are not an effective treatment. As of April 2022, there are several vaccinations available for COVID-19. All these aspects impact the physical perspectives of human beings. Beyond these perspectives, COVID-19 also exhibits significant impacts on other perspectives covering human sentiments, social interactions, etc.

Motivation behind This Study on Sentiment Analysis

On a worldwide scale, the commencement of COVID-19 had a significant influence on humanity. The physical aspect has the most visible and large-scale influence, and here is where the majority of the studies have been done. Additionally, COVID-19 has a substantial effect on mental health. The majority of people brought to the hospital with serious sickness were also found to have long-term neurological and psychological issues. Because the pandemic is occurring in the digital age, COVID-19 must have an impact on social media trends, which are inextricably entwined with human lives. Investigating the engagement across social media accounts and analyzing the sentiments there could present an effective way to discover this effect.

Roles of Social Media

Social media have become a valuable source of data for scientists, as social media houses a large amount of raw user data mostly in textual form along with the image, audio, and video formats that get updated daily. We can quickly gain insights from the different formats of data through sentiment analysis and topic modeling. This is especially common among content makers and product owners. A product or movie review sentiment can be analyzed to better understand the customer and make the necessary steps to further improve the product or services. Sentiment analysis is pivotal in terms of understanding people's perceptions and helps in decision-making [3].

During catastrophes and disasters, social media has played an increasingly essential role and has emerged as an important alternative information channel to traditional media in the last five years, ranking as the fourth most popular source of emergency information. Individuals and communities have utilized social media for a variety of purposes, ranging from warning others about dangerous regions to fundraising for disaster assistance. People utilized Twitter, Facebook, Flickr, blogs, and YouTube to publish their experiences during earthquakes in the form of texts, images, and videos, resulting in a donation of $8 million to the Red Cross, demonstrating the effectiveness of social media in spreading information during calamities [4].

Sentiment Analysis in Social Media

Sentiment data from social media might also be used to better project information about the destruction and recovery scenario, as well as charity requests, to the crowd. Sentiment analysis also aids in the comprehension of people's views on many online subjects. It has a lot of uses in social media, and there have been a lot of recent studies that apply sentiment analysis methods to social media data. Movie reviews, product reviews, App reviews, stock market predictions, and trend findings are just a few examples of these uses. These often result in the revelation of new findings.

For example, fear and anxiety are two feelings that people report immediately over social media after the occurrences of earthquakes, although calm and unpleasantness are not expressed as clearly during small earthquakes but are after huge tremors. A few researchers looked into the feelings of over 50 million tweets before, during, and after Hurricane Sandy to see how people's behavior changed on Twitter based on the number of published posts and stated sentiments in their tweets at the time the hurricane hit different cities. They notice that the sentiment of tweets differs from that of typical tweets, and they conclude that evaluating sentiments from tweets, coupled with other data, allows for the use of sentiment sensing for disaster detection and location. An interesting research direction here would be to look into the impact of other available emotional indicators in social media such as product ratings and reviews. Another possible study direction may be to use the findings of psychological and sociological studies on people's behavior (e.g., hope, fear, etc.) during disasters. This additional knowledge could assist decision-makers in better comprehending people's behaviors and feelings during disasters, as well as how to deal with the problem. Another potential future research direction is to look into how this information can be reprocessed so that it is immediately usable by the appropriate authorities [5].

Scopes of Sentiment Analysis

Sentiment analysis can aid in the resolution of many problems and provide numerous indicators in areas such as elections, public opinion and advertising, health care, and public satisfaction. Sentiment analysis aids in bio-informatics, such as cancer detection, as well as forecasting future stock market trends through the analysis and mining of social media posts. Sentiment analysis is widely used in social applications. Monitoring violence by detecting violence polarity in tweets, predicting election results and public attention, determining satisfaction of places and recommending those places accordingly, and monitoring and tracking students' opinions in education are some of the existing applications. Another application for sentiment analysis is to improve machine translation quality by detecting the implicit emotion of the text, such as sarcasm. Using sentiment analysis on medical data can help define and predict suicide and depression rates, monitor and track healthy and unhealthy areas based on tweets, and rank doctors based on patient satisfaction (via posts) and experience levels. Sentiment analysis has been used in the industry for brand monitoring, stock market prediction, predicting box office results based on user tweets, and measuring user satisfaction level [6].

Sentiment Analysis can also be used to track the performance of a brand. One of the most essential purposes of sentiment analysis is to get a complete 360-degree perspective of how customers perceive a product, organization, or brand. Companies frequently employ sentiment analysis to assess the impact of their goods and campaigns on customers and stakeholders. Brand monitoring gives us a lot of information on a brand's conversions in the market. Sentiment analysis allows us to categorize the importance of all brand mentions and route them to the appropriate team. Customer service firms frequently utilize sentiment analysis to classify incoming calls from users into "urgent" and "not urgent" categories. The categorization is based on email comments or proactively recognizing calls from disgruntled consumers. Stocks and the stock market are always a risk, but they can be reduced if proper research is conducted before investing. The comments and reviews of the goods are frequently displayed on social media. It is much easier to evaluate the client retention rate when we have access to sentiment data about a firm and new items. Brand Monitoring provides us with raw, unfiltered data on customer sentiment. This analysis can, however, be applied to customer service encounters and surveys. Sentiment analysis can allow us to do any type of market research or competitor study [7].

COVID-19 Sentiments in Twitter

According to many researchers [8–11], Twitter is an effective communication tool for gaining a better grasp of public concern and awareness regarding COVID-19. Sentiment analysis and topic modeling on Tweets can provide important information regarding trends in the discussion of the COVID-19 pandemic on social media, as well as alternate opinions on the COVID-19 catastrophe, which has sparked widespread public concern. The findings from this tool may aid health agencies in disseminating information to allay public fears about the disease. Twitter data may be utilized to investigate public awareness and feelings regarding the COVID-19 epidemic, which should be recognized by policymakers. It is crucial to highlight that public awareness levels are dynamic, as seen by [12], the two or three awareness peaks identified in this study within a period of only a few months. The findings of the same study also demonstrated that people express negative feelings and transmit both true and false information via various social media platforms at various phases of the condition. These findings allow for a wide variety of inferences to be made. For instance, people are typically afraid whenever there is a pandemic going on. This concern can be alleviated if the government works to synchronize the flow of information and combats the spread of ”fake news” related to the epidemic. According to the research, the government should also adopt countermeasures and build national surveillance systems to analyze web-based content, including social media, in order to get a better understanding of how the general public feels. In addition, the propagation of false information on the internet has the potential to trigger widespread concern and have catastrophic results. It is necessary to have a public health presence on social media that is more proactive in order to mitigate the effects of this.

Throughout the COVID-19 pandemic, accurate and timely data on the delivery and use of health services can be used to drive key decisions and actions. As COVID-19 spreads, governments and public health authorities will need to step up their efforts to address unanswered problems. Twitter users are mostly concerned with discussing and reacting to health issues, public health interventions, and pandemic control. This type of data assists governments in determining which public health messages are effective. All levels of government must enhance their responses. Governments and public health organizations must guarantee that healthcare systems are prepared to handle rising caseloads. To combat the COVID-19 epidemic, community-based health care is a critical component of primary care. Recognizing public concern and awareness can assist governments in gaining a better understanding of how the public feels about the disease at any given time. When the outcomes are linked, a valuable healthcare resource can be created to help construct a long-term strategy [12].

To conclude, the primary application areas for sentiment analysis include: monitoring Social media Hypes, brand monitoring and managing reputation, understanding public emotion during a significant event, product viability analysis, understanding public sentiment on a specific topic, bench-marking, and market research, etc. Since sentiment analysis is a widely popular method of assessing public sentiment, any change in the sentiment will impact human life on a large scale. Measuring the changes in these sentiment trends is thus a crucial task in light of the COVID-19 pandemic.

Our Research Questions

Many significant studies have been conducted to analyze the sentiments found in shared social media posts. However, these studies often focus on a specific hashtag-related collection of posts, or the timeline of the analysis is not broad enough to show the full context. Their experiments often also lack in exploring other modalities such as exploring image or video along with text for the purpose of sentiment analysis. Considering all these gaps in the literature, this research work focuses on finding answers to the following research questions for the users who have shown interest in COVID-19.

RQ1. How have the sentiments shared over social media got changed owing to COVID-19? After the outbreak of COVID-19, there has been a shift in public opinion. However, what exactly got changed here? Did the positive, negative, or neutral sentiment rise? If so, by how much? How about the sentiments expressed over images?

RQ2. Was the change in image sentiment similar to text? If not, are they correlated by any means? What if users express sentiment differently in text and image posts? This could lead to a bimodal study for sentiment analysis.

RQ3. To what degree did the changes happen for all the sentiments? Quantifying the changes is crucial if we want to measure the impact. Simply stating ‘rise’ and ‘fall’ are not sufficient if we are trying to perform an in-depth scientific study.

RQ4. What amount of time did it take for the changes in sentiments to “settle down”? What happened after the outbreak - did everything go back to normal, and if so, how quickly did it happen? Was it the same for all sentiments?

Our Approach

In order to address these questions, we carefully selected 569 users who have posted tweets about COVID-19 during the pandemic and collected all of their public tweets and images shared during the years 2019 and 2020. We carried out several statistical analyses to provide a conclusive answer to the research questions at hand. In the process of answering the research questions, we focus on three different research objectives.

1. To develop a new classification technique for text-based Tweets leveraging existing sentiment classifiers,
2. To perform a longitudinal[1] bimodal (text and image) study on trends of sentiment spanning over two (prior and during COVID) years, and
3. To find potential similarity patterns between trends of sentiments over text and images for different periods (covering prior and during COVID years).

Following the objectives of this study, we envision the following three research outcomes.

- A new sentiment classification technique for analyzing sentiments over text-based tweets,
- A longitudinal bimodal study (using our developed technique) analyzing the change-points over time series of the trends in positive, negative, and neutral sentiments relating to COVID-19, and
- Similarity patterns between the trends in sentiments, expressed through text and images, over different time periods covering before and during the pandemic.

[1]A study strategy in which the same variables are observed repeatedly over short or long periods.

Organization of This Study

The remainder of the study is structured as follows. Chapter 2 summarizes the techniques others have followed for researching similar topics. Chapter 3 provides the steps and methodology we follow and the challenges that we face while conducting this research. Chapter 4 presents our key research findings. In Chapter 5, we compare our findings with the related studies and summarize the key findings. Chapter 6 shows limitations of our work. Finally, Chapter 7 features our concluding remarks based on this study.

Research Ethics and Anonymization

We ensured the confidentiality and anonymity of the participants in this research work. We got the approval of the study and data collection from the Ethics Committee of the institution of the author. We store our collected data in a private Google Drive.

Chapter 2

Related Work

In the course of an endemic, epidemic, or pandemic, researchers have investigated the ways in which people have been displaying their characteristics through cyberspace. During an outbreak, it is also usual practice to investigate the role that social media may play in regulating an individual's life. The three subsections that follow will talk about other studies that researched cyberspace, the typical impact on social media, and the normal approach that these researchers used.

Approaches to Sentiment Analysis

Since the outbreak of COVID-19 took place, much scientific research has been conducted in this field. Most research focused on the sentiment of the COVID-19-related text tweets. A few studies include topic modeling using Latent Dirichlet allocation (LDA) [11,12] which is generally used to find a linear combination of features that characterizes or separates two or more classes of objects or events. The initial tweet collection was made specifically using hashtags or COVID-19-related keywords. Most of the sentiment analysis are done using VADER [11, 13–15] or TextBlob [8]. Some studies are done using custom-built neural network models [16]. A few works included transformers such as *simpletransformer* [17] while others deployed CNN [10, 18, 19] to detect text tweet sentiment.

Location and Topic-based Study

[9] conducted a country-wise measuring ratio of sentiments during COVID-19. [13] was based on finding whether people are against specific topics such as mask-wearing, lockdown, etc. Detecting positive, negative, and neutral sentiments are the most common forms of analysis. Some studies also took another step and did emotion analysis such as identifying whether a user post has anger, hope, optimism, etc [16, 17].

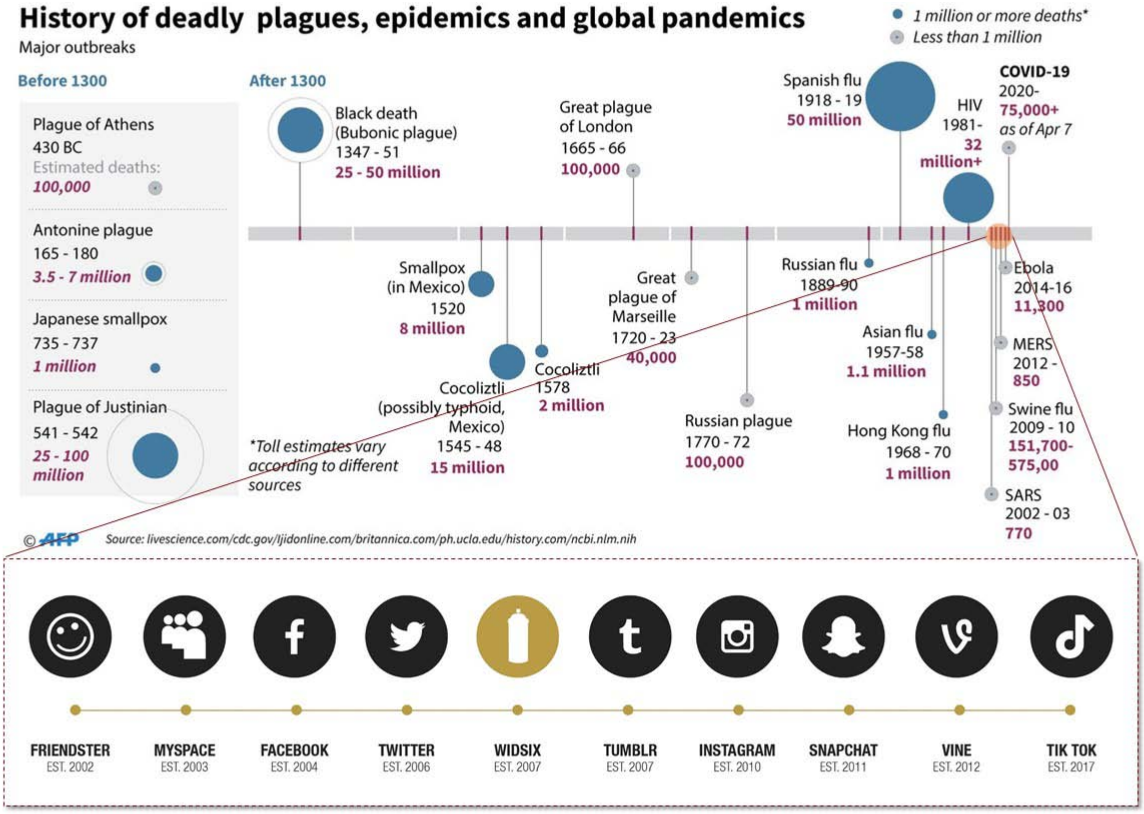


Figure 2.1: History of pandemic alongside the emergence of social media. Image Courtesy: #AFPGraphics [1] and WidSix [2]

The purpose of the studies was also different. Some researchers conducted studies entirely to track down sentiment on various topics. [14] collected hospitality and healthcare-related tweets during COVID-19. This work also analyzed tourism-related tweets at that time. Some research work focused on finding the COVID situation in a specific country such as Nepal [10] or India [16].

Knowledge Graph and Frameworks

Some work focused on building knowledge graph [20] based on COVID-related tweets. Some work focused on finding the authenticity of published news articles [13, 21]. Some work focused on building a comprehensive framework for finding all possible relevant posts in popular social media as well as news articles [15, 16, 19].

Image Sentiment Detection

Some research work included image analysis. Some works clustered collected images based on corresponding text [13]. Some had put image sentiment based on related text sentiment [21]. Some analyzed corresponding images to find relevant metadata such as the information about the user [15]. Some research designed a multi-modal search engine [21] for analysis. In that work, sentiments collected from metadata were assigned to related images based on user preferred weight. Another work focused on building a framework where image sentiment analysis was done using the popular Fast R-CNN [19]. Additionally, they have GIFs are handled using an optical character recognizer which separates texts from images for defining the polarity.

Change-point Detection

Change-point detection was done in a few research works as well. It was common in a particular type of study where the authors studied the tweets to build a knowledge graph [20], or tried to correlate the tweets with a concurrent event [15].

Analysis Duration

The duration of analysis varied widely over the relevant literature works as well. Some papers focused on March or April, right after the COVID-19 outbreak to see minor changes. They do not offer any pre-pandemic context view [8, 9]. Other papers [12] focused on the announcement date of COVID-19 (December 2019 to March 2020). We also saw papers focusing on the post-pandemic view analyzing either the novel cases [16] or sharing public perspectives towards topics such as tourism [14] or veracity assessment [13].

Dataset Crafting

In addition, a portion of the effort was devoted to the construction of a sentiment-annotated twitter dataset [17]. They collected tweets related to COVID in popular languages such as English, Arabic, and Spanish, among others, and assigned sentiment and emotion tags to each tweet using a transformer framework that had been pre-trained in the relevant language and fine-tuned for multi-label sentiment classification.

Summary

To summarize, existing research studies to date focused on analyzing the sentiments covering only text-based social media posts along with topic modeling using Latent Dirichlet Allocation (LDA) [9, 11, 12, 22]. These studies generally used out-of-the-box Python libraries such as VADER [11,13–15] or TextBlob [8,23] for text-based sentiment detection. Moreover, most of the studies relied on a single algorithm for sentiment detection, which often increases the chance of misclassification. Besides, there exist only a limited number of studies on image-based sentiment analysis in the literature [13, 15, 21]. Here, the image-based sentiment detection techniques are limited to only specific types of hashtags [13], specific geographic locations [24, 25], or narrow timelines [15].

Thus, to the best of our knowledge, an extended longitudinal study on the changes in trends of sentiments due to COVID-19 is still unexplored in the literature. This happens as none of the existing studies includes any data from the prior (pre-COVID) year for comparing that with data of the next (during-COVID) year. However, such a longitudinal study is required to depict a comprehensive view of the trends of sentiments to help psychologists as well as policy-makers to get a bigger picture of what is happening in the psychological world owing to the pandemic. Additionally, measuring the correlations between sentiments expressed in different ways (through text and images) over social media is yet another unexplored aspect that can facilitate investigating the comprehensive view.

Research Gaps

We identified the following limitations of other works.

1. Most papers focused on the year 2020 (and some till June 2021). None had the focus on 2019, which would give a context to the change in sentiment that happened due to COVID-19.

2. Most papers only worked with text tweets (and did topic modeling). A handful of papers worked with both text and image, but they strictly focused on a single topic (mask/restriction/politics). Our study can be a baseline for those who are looking for the big picture.

3. Most papers focused on the COVID-19 tweets, ours were on the overall timeline of 569 users. This gives a unique picture of how the online social life of the COVID-19 interested users changed instead of only the COVID-19 sentiments.

4. Most of the works used VADER or TextBlob for sentiment analysis. Some papers were entirely dedicated to developing a state-of-the-art sentiment analyzer, but they did not conduct the full-scale analysis as we did. We proposed a better heuristic for text sentiment classifier that combines both rule-based and neural network-based approaches.

These identified gaps especially lacking context is crucial for the policymakers and health worker in making any administrative decisions. They would want to know whether the rise of negative sentiment is actually due to COVID 19, or it is just a seasonal fluctuation. The analysis has to include image sentiment on top of context as is it possible that people may express emotion differently in images compared to text. There has to be a robust text sentiment classification technique as most of the social media content is still text-based. Most of the sentiment libraries are not fit to classify social media tweets. Finally, a complete framework has to be laid out to carry out this bimodal sentiment analysis to aid the researchers with similar interests.

Chapter 3

Methodology

To achieve the possible outcomes mentioned in Section 1.4, we planned a methodology for our study. To get an understanding of the pattern of sentiment change over cyberspace, we initially collected public thoughts and opinions regarding COVID-19 from social media. We selected Twitter which is a popular micro-blogging platform used worldwide for posting ideas and sharing experiences of daily life. The justification for choosing Twitter as a sentiment analysis data source is discussed in Section 1.2.4.

Data Collection

The subjects of our investigation are only those who have shown interest during the COVID-19 pandemic. For selecting these users, we collected tweets based on COVID-related hashtags that are commonly referred to in other research literature. The following 21 hashtag combinations were used for probing tweets.

> coronavirus; COVID-19; dashboard; comorbidity; lockdown; quarantine; vaccine; worldometer; who, corona; who, COVID; corona, dashboard; COVID, dashboard; data visualization, corona; data visualization, COVID; data visualization, dashboard, corona; data visualization, dashboard, COVID; hopkin, corona; hopkin, COVID; social, distance; worldometer, corona; worldometer, COVID

We wanted to select the users who are actively concerned about the infection rate, vaccine development progress, and also restrictions imposed due to COVID-19. Therefore we have included hashtags that are closely related to COVID-19 such as lockdown, worldometer, social distance, etc. For collecting the related tweets, an advanced tweet scraping Python library named Twint was used. In total, we collected 236,782 tweets posted from 1 January 2020 till 2 September 2020, 1:11 PM (GMT).

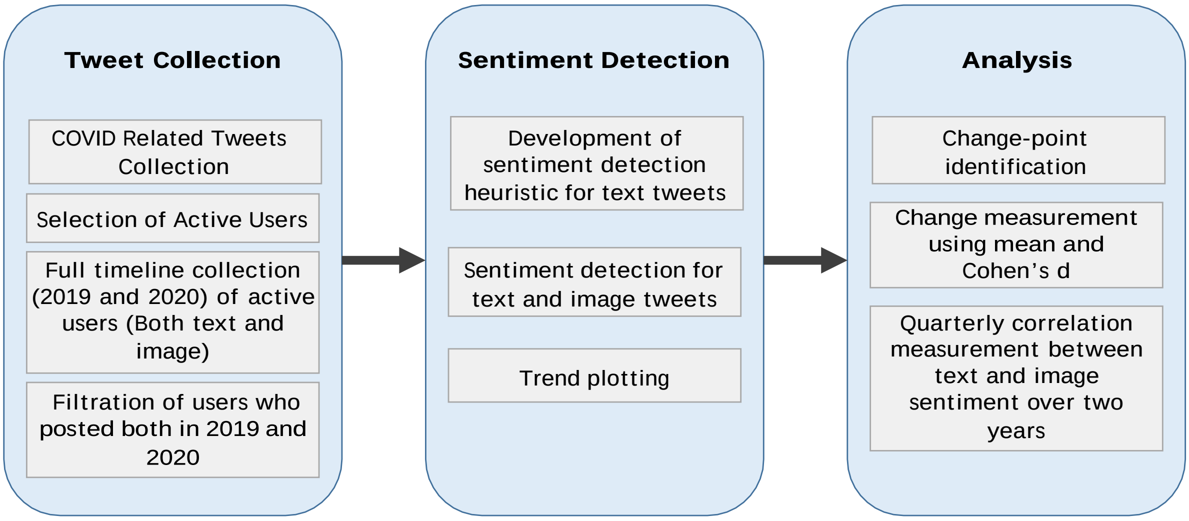


Figure 3.1: Steps of our methodology

From the collected tweets, the number of unique users we found was 149,210. These are the initial set of users who showed interest during the COVID-19 outbreak and shared tweets using the hashtags. It is computationally challenging to collect the full timeline of all of these users. Therefore it was needed to find significant users.

To find the most active users, we initially sorted them based on the number of tweets they made. However, after a quick inspection, it was found that the top posting users on the list are spammers and posted the same tweet in large volumes. It can be understood from this that most of these top-posting users in reality are automated robot agents who post the same tweet at a very high frequency. We noticed that although these types of users post in a very large volume, their activity is limited to only a handful of days.

To filter these users, we decided to sort the users based on the number of individual days they posted something on Twitter. It was found that the 1000^{th} user's tweets spread over only five days, which further proves our observation. Accordingly, considering the number of days over which we found users' tweets, We selected the top 720 users from the list (see Figure 3.2). We collected all of the public tweets of these selected users from 1 January 2019 to 24 December 2020. Over 724 days spanning this period, the selected users posted a total of 7,409,429 tweets. Their posts also contained 1,233,743 media links that include images and video thumbnails. During the collection of these images, it turned out that 1,198,715 links were valid and the rest were either deleted, made private, or the corresponding user became inactive altogether.

Since we want to detect the change in sentiment of the tweets of a user in 2020 (during COVID-19), it was important that the users who tweeted in 2020 also should be active in 2019. Therefore we had to further prune down the list of 720 users and selected only the users who tweeted both in 2019 and 2020. The number of such users is 569. They have collectively posted 5,353,462

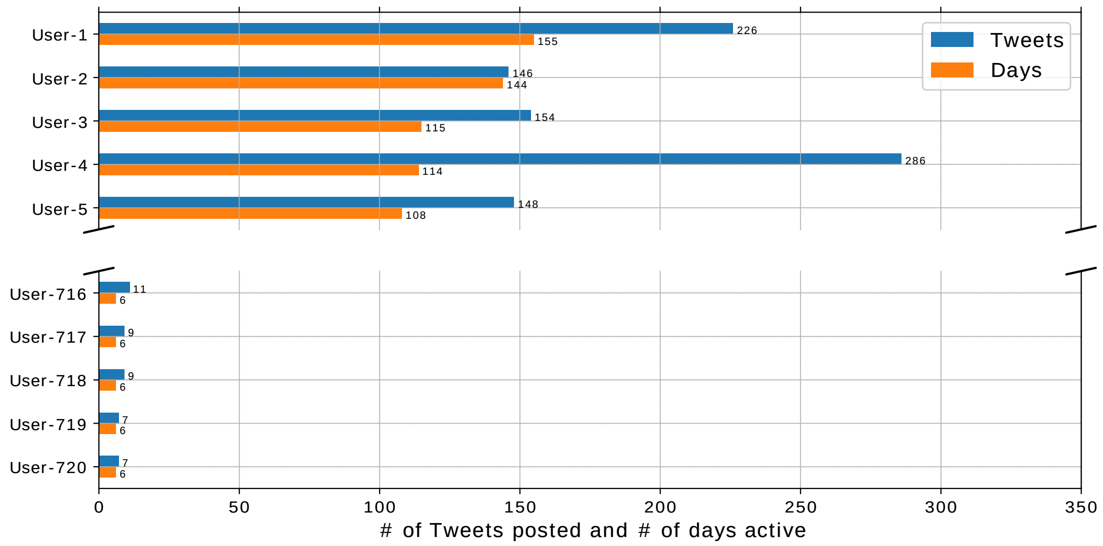


Figure 3.2: Listing users based on days active in Twitter

tweets and 1,029,444 images. The whole process is summarized in Figure 3.3. The number would reduce even more if the year span increases.

Sentiment Detection

After collecting the text and image tweets, we assigned one of the three sentiments (negative, neutral, or positive) to each of them. Only English tweets were considered for text tweets. For text sentiment classification, we developed a heuristic that combines the output of both classical rule-based models and modern transformer-based classifiers. For image sentiment classification, we used a cross-media learning model initially released in 2017. The detailed process of classification is discussed in the following sections.

Text Sentiment Classification

Text is the most popular form of idea-sharing mode in social media. Therefore, it is crucial that the sentiment of the shared text is correctly classified. However, traditional sentiment detection libraries do not perform very well in capturing the sentiment of the shared tweets due to different styles of expression on social media. We selected five sentiment classifiers (Table 3.1) for detecting text sentiment.

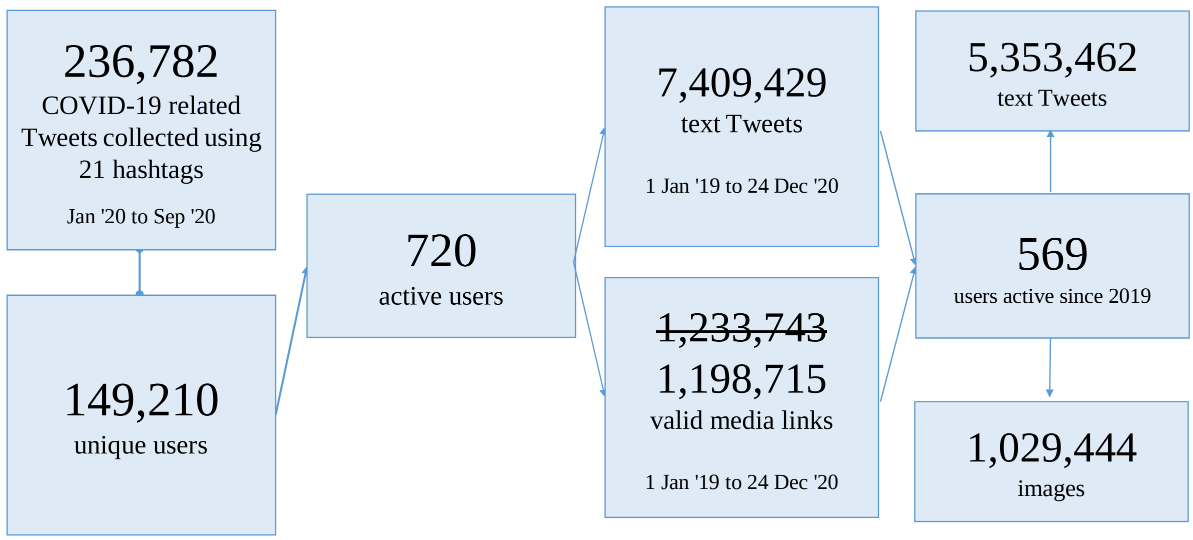


Figure 3.3: Active User Selection Process

Tweet Preprocessing

All text tweets were cleaned before sentiment classification using the *tweet-preprocessor* library available in Python. This tool has 266 GitHub stars and is specialized for cleaning tweet texts removing irrelevant text information such as URLs, emojis, hashtags, mentions, twitter reserved words (such as RT or FAV), smileys, and numbers - to name a few.

In order to apply the cleaning function, all the text tweets were put into a Pandas Dataframe. It is a 2-dimensional data structure for processing a large number of data in an optimized manner. We then applied the 'clean' function of the tweet-preprocessor library to the text tweet column. It removed the URLs, hashtags, mentions, and any other unwanted text. The results were recorded in another column and saved in JSON format for further use.

Library Selection

Most of the recent tweet sentiment-related research studies [15], [8] relied on VADER (Valence Aware Dictionary and sEntiment Reasoner, a lexicon and rule-based sentiment analysis tool) or TextBlob (a Python library for processing textual data) for text sentiment classification. However, we observed that these classifiers do not always perform up to the mark in the case of text sentiment tweets. To explore further, we primarily selected five popular sentiment detection libraries for text sentiment classification. Here, we wanted to include a wide variety of classifiers to explore how they perform against human perception especially for COVID-19-related tweets.

Table 3.1: Baseline text sentiment detection libraries

Ser	Library	Type	Remarks
1	VADER [Chandrasekaran et al, 2020]	Rule-based	Extremely popular among researchers
2	Afinn [Nielsen et al, 2011]	Rule-based	Specialized for microblogging
3	TwitterSentiment [Crepineau et al, 2018]	Transformer	Specialized for Twitter sentiment
4	HugginFace [Punith et al, 2021]	Transformer	State-of-the-art pre-trained models
5	TextBlob [Manguri et al, 2020]	Rule-based	Lightweight NLP tool

Accordingly, we considered rule-based (TextBlob, VADER, and Afinn) as well as transformer-based (HuggingFace and TwitterSentiment) sentiment classifiers. Among them, VADER and Afinn are specialized in social media sentiment analysis and TwitterSentiment is a fine-tuned Twitter sentiment classifier written in the PyTorch framework. All of these libraries have their own thresholds for classifying negative, neutral, and positive sentiments. Their working modalities also substantially differ.

VADER presents a lexicon and rule-based sentiment analysis tool that is tuned for classifying social media sentiments. Under the MIT License, it is completely open-source. VADER employs a sentiment lexicon, which is a collection of lexical features (e.g., words). The features are classified as positive or negative based on their semantic orientation. VADER not only displays the Positivity and Negativity scores, but also the degree to which a sentiment is favorable or negative. Here, the compound score is calculated by adding the valence ratings of each word in the lexicon, adjusting them according to the rules, and then normalizing them to a range of -1 (most extreme negative) to +1 (most extreme positive). If we want a single unidimensional measure of sentiment for a given text, this is probably the best metric to employ. Researchers who want to define uniform thresholds for identifying statements as positive (for example, $compoundscore >= 0.05$), neutral (for example, compound score is between -0.05 and 0.05), or negative (for example, $compoundscore <= -0.05$) will find it useful. For this classifier, in our study, the neutral range was set as $-0.05 < compound\ score < 0.75$.

AFINN is another tool that works on a list of words with valence ratings ranging from -5 (negative) to +5 (positive). The AFINN lexicon rates word with negative values indicating negative sentiment and positive values indicating positive sentiment. For this classifier, in this study, the neutral range was set as $-0.1 < sentiment\ score < 0.0$.

TwitterSentiment is a sentiment analysis software written in Python utilizing the PyTorch framework. The goal of this classifier is to create a sentiment analyzer tailored to the Twitter domain. Because of the limited amount of characters permitted in a tweet, the majority of tweets do not reflect standard English syntax and vocabulary. This necessitates extra caution in order to achieve better results, which is why this initiative exists. To create a text corpus, a prominent Twitter sentiment analysis dataset containing 1.4 million tagged tweets was preprocessed. The Keras tokenizer is used to transform each different word in the corpus into tokens. Then, using Glove Twitter (200 Dimension), which has been pre-trained on the Twitter corpus, text embedding is generated. Finally, the torchtext module loads and processes text embeddings into a matrix. To create the classifier model, a model utilizing Generative recurrent neural networks (GRU) was trained using the dataset. For TwitterSentiment, in our study, the input was classified as neutral sentiment if the sentiment score is 0.

Transformers (previously pytorch-transformers and pytorch-pretrained-BERT) library provides tens of thousands of pre-trained models for performing tasks over text, vision, and audio. In over 100 languages, these models may be used to perform tasks such as text classification, information extraction, question answering, summarization, translation, and text synthesis. In our study of text classification with this classifier, the neutral range was set as $-0.7 < sentiment\ score < 0.7$.

TextBlob is a text processing package for Python 2 and 3. It offers a basic API for doing standard natural language processing (NLP) activities such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, etc. By this classifier, a named tuple of the form *Sentiment (polarity, subjectivity)* is returned by the sentiment property. Here, the polarity score is a floating point number between -1.0 and 1.0. Subjectivity is also a floating point value between 0.0 and 1.0, with 0.0 being very objective and 1.0 being very subjective. For this classifier, in our study, the neutral range was set as $-0.05 < sentiment\ polarity < 0.05$.

Ground Truth for Text Sentiment

We selected 500 text tweets and five human observers classified their sentiment as either negative, positive, or neutral. The five raters were from CS backgrounds aged from 20 to 28 years. All raters were male, proficient in English, and had sufficient experience in using social media. Fleiss Kappa is a popular co-efficient for measuring inter-rater agreement [26, 27] in the case of three or more raters. The Fleiss Kappa value for the inter-rater agreement is 0.397, which suggests that the five raters have fair agreement [28]. The reason behind this level of agreement is the fact that the tweets during the COVID-19 period are different in nature compared to the tweets during the conventional non-pandemic period, and therefore, it is not very usual that the tweets during the COVID-19 period are always classified in the same manner by all.

The results of classification by the rates were aggregated through the notion of majority voting.

The aggregated results were considered in the process of determining the baseline accuracy for our text-based sentiment classification. To do so, the same 500 tweets were classified using the five libraries mentioned previously. Accordingly, the accuracy, as well as F-1 score, was calculated for all of the libraries with respect to the baseline classification achieved through the majority voting over the five raters. The results obtained in this way are shown in Section 4.1.

Heuristic Development

In order to combine the outputs of the libraries, we tried numerous combinations, and three of these heuristics are noteworthy (See Table 4.2, H-1 to H-3). The confusion matrix of H-1 revealed the libraries have a bias towards marking the tweets negatively. Furthermore, H-1 was less accurate owing to the poor accuracy of TwitterSentiment and TextBlob. Therefore, the other three libraries were chosen (which show less of such bias) and H-2 and H-3 are developed based on their output.

Upon close inspection, it was observed that HugginfaceTransformer can detect the neutral sentiment fairly well keeping the other two sentiments in balance. H-2 performed majority voting on the best three libraries (VADER, Affin, and Transformer), and in case of a tie, will check if the Transformer gave the verdict of neutral or not. If not, the outcome of VADER would be the final outcome. H-3 was exactly the same except Afinn is used in place of VADER at the last deciding step. This increased the accuracy to 68% outperforming all the other results. The pseudo-code of the three heuristics are described in Algorithm 1, Algorithm 2, Algorithm 3.

Algorithm 1 Majority voting on all five classifiers (Heuristic-1)

$X[0] \leftarrow get\_VADER\_decision(str)$
$X[1] \leftarrow get\_Afinn\_decision(str)$
$X[2] \leftarrow get\_TwitterSentimentLib\_decision(str)$
$X[3] \leftarrow get\_Transformer\_decision(str)$
$X[4] \leftarrow get\_TextBlob\_decision(str)$
if $X.count(\text{“}Positive\text{”}) \geq 3$ **then**
  return “Positive”
end if
if $X.count(\text{“}Neutral\text{”}) \geq 3$ **then**
  return “Neutral”
end if
if $X.count(\text{“}Negative\text{”}) \geq 3$ **then**
  return “Negative”
end if

Algorithm 2 Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-1 (Heuristic-2)

$X[0] \leftarrow get\_VADER\_decision(str)$
$X[1] \leftarrow get\_Afinn\_decision(str)$
$X[2] \leftarrow get\_Transformer\_decision(str)$
if $X.count(\text{"}Positive\text{"}) \geq 2$ **then**
  return "Positive"
end if
if $X.count(\text{"}Neutral\text{"}) \geq 2$ **then**
  return "Neutral"
end if
if $X.count(\text{"}Negative\text{"}) \geq 2$ **then**
  return "Negative"
end if
if $get\_Transformer\_decision(str) = \text{"}Neutral\text{"}$ **then**
  return "Neutral"
else
  return $get\_VADER\_decision(str)$
end if

Image Sentiment Classification

Image sentiment classification was not very common in terms of COVID-19 analysis and most of the attempts we studied deployed indirect approaches for sentiment detection as discussed in Section 2.1. We performed sentiment detection based on the implementation of a Cross-Media Learning model [29] dedicated toward Twitter image sentiment analysis. The implementation offered several pre-trained models among which the VGG-19 model was used. the VGG-19 had an 88.1% accuracy on the test dataset. The implementation provided a vector for each of the images containing confidence in the three sentiments. Majority voting was used to find the final sentiment of the image.

User Sentiment Group Migration

Before diving into granular trend analysis, we wanted to generate a high-level overview of how the nature of sentiment changed for each person. We wanted to find out whether those who shared tweets having mostly a particular sentiment in a year exhibit the same behavior in the next year or not. For this purpose, we have classified the users of a particular year (2019 or 2020) based on the following criteria.

Negative Users Users who have posted mostly negative tweets (text or image) throughout the year

Neutral Users Users who have posted mostly neutral tweets (text or image) throughout the year

Algorithm 3 Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-2 (Heuristic-3)

```
X[0] ← get_VADER_decision(str)
X[1] ← get_Afinn_decision(str)
X[2] ← get_Transformer_decision(str)
 if X.count("Positive") ≥ 2 then
   return "Positive"
end if
 if X.count("Neutral") ≥ 2 then
   return "Neutral"
end if
 if X.count("Negative") ≥ 2 then
   return "Negative"
end if
if get_Transformer_decision(str) = "Neutral" then
   return "Neutral"
else
   return get_Afinn_decision(str)
end if
```

Positive Users Users who have posted mostly positive tweets (text or image) throughout the year

In order to find how many people have changed their overall sentiment during COVID-19, we divided the tweet and image sentiments into yearly groups (2019 and 2020). If a user's tweets were mostly negative during year 2019, then he/she is marked as a 'Negative User' for that year. We have marked all individual users as either 'Negative User', 'Neutral User', or 'Positive User' for both the years and observed the changes in the count.

Plotting and Change-point Detection

After assigning sentiment to each of the text tweets and images, we divided them (referred to as data points) into six individual datasets (See Table 3.2) based on category (text or image) and sentiment type (negative, neutral, or positive). We further aggregated the datasets counting the number of data points on day-wise groups hence generating six time-series having 724 data points each. All counts of sentiments in a category (text or image) were normalized.

COVID-19 drew public attention from around December 2019 and was declared a global pandemic on 11 March 2020 [11]. However, the study from [21] suggests that there is usually a delay between the occurrence of an event and the response on social media. In order to exactly find out when the change occurred for each time series, we ran three different change-point detection algorithms, namely CUSUM (Cumulative Sum), Pelt (Pruned Exact Linear Time)

Table 3.2: Six time series generated from the posts of 569 users spanning over 724 days

Category	Sentiment	Time Series Name	Total Data Points	Ratio
Text	Positive	TS-1	1,122,934	21% of Text Tweets
	Neutral	TS-2	2,857,648	53% of Text Tweets
	Negative	TS-3	1,372,880	26% of Text Tweets
Image	Positive	TS-4	177,503	17% of Images
	Neutral	TS-5	439,492	43% of Images
	Negative	TS-6	412,449	40% of Images

Search, and DP (Dynamic Programming) Search. We also decomposed each series into trends and compared them against the rolling window mean and standard deviation to perform a visual inspection.

CUSUM (Cumulative Sum)

The CUSUM (or cumulative sum control chart) is a sequential analysis approach developed by E. S. Page of the University of Cambridge for statistical quality control. It's commonly used for change detection monitoring [30]. CUSUM was published in Biometrika in 1954, a few years after Wald's SPRT algorithm was published.

As its name implies, CUSUM involves the calculation of a cumulative sum (which is what makes it "sequential"). Samples from a process x_n are assigned weights ω_n, and summed as follows:

$$S_0 = 0 \tag{3.1}$$

$$S_{n+1} = \max(0, S_n + x_n - \omega_n) \tag{3.2}$$

The ω usually represents a likelihood function. Change-point is considered to be found if the value of S crosses a threshold value.

We used Facebook's Kats Library to implement CUSUM. The base detector employs a Gaussian distribution model for measuring the cumulative sum and assumes that there is only a single change-point. The wrapper library runs the base detector multiple times to detect multiple change-points [31].

PELT Search

Pruned Exact Linear Time (Pelt) Search is popular among researchers [32, 33] for change-point detection. It is an exact method for detecting change points and works by assigning a penalty to changes to a time series. We used *ruptures* for detecting the change-point using PELT search, which is a Python library that contains offline change-point detection algorithms.

Dynamic Programming (DP) Search

Dynamic Programming (DP) search identifies the best partition for which the sum of errors is minimum given a segment model. *Ruptures* was also used for applying this search method.

Series Decomposition

In order to aid the visual inspection, we also decomposed each time series to see the underlying trend. Decomposing a time series entails viewing it as a collection of level, trend, seasonality, and noise components. Decomposition is a useful abstract paradigm for thinking about time series in general, as well as for better comprehending challenges in time series analysis and forecasting. The trend component of a time series represents the series' long-term evolution. When data has a consistent increasing or declining tendency, it is called a trend. It is not necessary for the trend component to be linear.

Furthermore, mean and standard deviation were calculated over a 30-day rolling window period for each of the time series. A sharp spike on the standard deviation graph would denote changes in the trend. On the other hand, the plotted average would visualize the number of changes more clearly.

Effect Size of Sentiment Change

Once the change-points are identified, we calculated the change in mean and the Cohen's d for the pre- and post-change-point sentiments. The change in mean in our case provides a generalized idea of increase and decrease. Cohen's d, on the other hand, provides a contextualized effect size by introducing the standard deviation between two series. It is an appropriate effect size for the comparison of two series where the overall mean is different. It can be calculated from the following formula.

$$\text{Cohen's D} = \frac{\textit{Mean Difference}}{\textit{Standard Deviation}} \tag{3.3}$$

The interpretation of different values for Cohen's d was taken from [34].

Correlation between Text and Image Sentiment

We correlated the text time series (TS-1, TS-2, TS-3) with the image time series (TS-4, TS-5, TS-6) using both parametric (Pearson's correlation coefficient [35]) and non-parametric (Kendal's correlation coefficient and Spearman's correlation coefficient [36]) methods. The methods require comparing two series. In order to track the changes over time, we divided the time series into eight quadrants (Q1-2019, Q2-2019, Q3-2019, Q4-2019, Q1-2020, Q2-2020, Q3-2020, Q4-2020) where each quadrant consists of three consecutive months. The r coefficients were calculated for each quadrant of similar sentiments between text and image and plotted with a trend-line to show how the correlation changed over time. The interpretation of Pearson's r was taken from [37].

Settling Trend Identification

We wanted to calculate how much time it took for each of the time series to return to their original form after the COVID-19 outbreak. This would denote how much time people took to recover from the pandemic. In order to quantify the changes that took place for each sentiment time series, we have selected the COVID-affected segment of each time series with the non-COVID-affected segment from the previous year till its change point. Each non-COVID segment was compared with the corresponding time series after the change-point till 15 December 2020. The distance calculated for the six time-series (TS-1, TS-2, etc.) is 313, 233, 283, 293, 263, and 425 - excluding the first data point. All segments were smoothed using the LOWESS Smoothing algorithm in order to capture the trend.

LOWESS Smoothing Filter

LOWESS filter is a common smoothing method that uses a locally weighted regression function. It draws a smooth line through a time-plot or scatter-plot to make it easier to spot relationships between variables and predict trends.

This method employs a weighting function, which has the effect of decreasing the influence of an adjacent value on the smoothed value at a given point as the distance between them grows.

DTW Calculation

We have calculated DTW (Dynamic Time Warping) distance for each pair of segments for all time series to understand how much change has occurred due to COVID for each individual time series. Dynamic time warping (DTW) is a time series analysis algorithm that measures the similarity of two temporal sequences that may vary in speed. DTW has been applied to temporal sequences of video, audio, and graphics data — in fact, DTW can analyze any data that can be converted into a linear sequence.

Series Reconstruction

In order to visualize how the series would originally look without COVID-19 impact, we trained two popular forecasting models with data one month prior to the change-point of each time series. This helps us visualize the impact of COVID-19 on the regular sentiment trends. The models used for our analysis are briefly described below. Each of them has a separate training method and parameters.

Prophet

The Prophet is a library developed by Facebook for forecasting time series data based on an additive model where non-linear trends are addressed with the help of periodical seasonality. The parameters of this model were cross-validated to minimize the deviation. Yearly, weekly, and daily seasonality parameters were enabled for the model during training.

Orbit

Orbit is a Python package developed by Uber for Bayesian time series modeling and inference. The Kernel-based Time-varying Regression model was used to cross-check the output of Prophet. We have used pyro-svi as an estimator and used weekly seasonality along with 2021 as a randomization seed. The value of the N bootstrap value was 10^4 as suggested by the documentation.

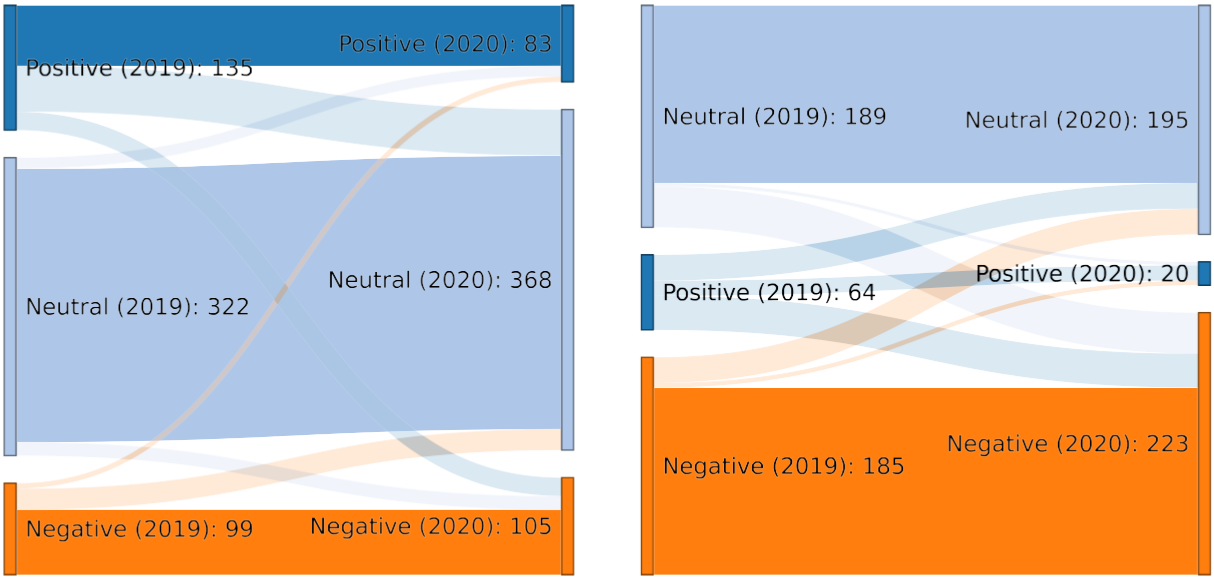
Effect of Smoothing Filter for Forcasting

A smoothing filter often increases forecast accuracy. Therefore, we applied ten different smoothing filters (Exponential Smoothing, Convolutional Smoothing, Spectral Smoothing with Fourier Transform, Polynomial Smoothing, Spline Smoothing, Gaussian Smoothing, Binner Smoothing, LOWESS, Seasonal Decompose Smoothing, and Kalman Smoothing) on the COVID-19 unaffected stable part of each time series (the first 300 days of 2019). The results are discussed in Section 4.7.3.

Chapter 4

Findings

It was a big challenge for us to identify the genuine users from the Twitter bots to filter out the human tweets relating to COVID-19. After shortlisting the 569 users, we plotted their tweet sentiment trends over the years 2019 and 2020 for both image and text data separately. Our first major finding in these trend lines is the different change points detected in the time series. Secondly, we measured the amount of deviation which took place after the change-point for each series. Finally, we measured the correlation coefficient r between text and image sentiment over the entire period.



(a) Text Tweets (b) Image Posts

Figure 4.1: Migration of users between various sentiments in 2020

Sentiment Library Evaluation and Heuristic Development

Five human observers were tasked with classifying the tone of 500 randomly selected tweets as either negative, positive, or neutral. The conclusion was drawn up based on the majority vote, and this was considered to be the benchmark for determining how accurately the text expressed its sentiment. The same 500 tweets were analyzed for their sentiment using the five libraries that were just discussed, and the accuracy and F-1 score for each library were computed in comparison to the accuracy of the baseline accuracy. The outcomes are displayed in Table 4.1 and Table 4.2 (L-1 to L-5). The accuracy reported for the classifiers/libraries in their source documentation differs from what we get in our study[1] model underneath and is fine-tuned with 150k English reviews along with reviews from another five languages. Pre-training the HuggingFace Transformer with our subjective data would increase the accuracy further, but 500 labeled data are merely sufficient for training such a large dataset. Also, labeling a large portion of the collected dataset accurately would require significant effort and time from the individual raters.. This happens as the tweets during COVID-19 time are different from the regular times. As a result, the pre-trained models and libraries used to classify sentiments of the tweets during the COVID-19 period resulted in less accuracy compared to that listed in the formal documentation of the classifiers/libraries.

User Sentiment Group Migration

Figure 4.1 summarizes the findings from this mode of analysis. After marking each user with a sentiment using the methodology discussed in Section 3.3, we observed if they stayed into the same sentiment in 2020, or changed the majority of the sentiment. Our finding is that the major migration happened into Neutral in terms of text-based tweets. We can see that 23% negative users and 37% positive users became neutral in 2020. Overall we see a 14% increase in neutral sentiment, 38% decrease in positive sentiment, and 6% increase in negative sentiment.

In terms of image posts, we see a different picture. We see that most users migrated into the negative zone (19% from neutral and 45% from positive). Overall we see a 29% increase in negative sentiment, 8% increase in neutral sentiment, and 64% decrease in positive sentiment.

[1]Pre-training the HuggingFace Transformer with our subjective data would increase the accuracy further, however, our 500 labeled data would not be sufficient for the necessary training. Moreover, labeling a large dataset (as needed for the training) would require substantial effort and time from the individual raters. The required effort and time would be ever higher in our case than general ones, as labeling pandemic-related tweets is more challenging than labeling tweets in normal time as experienced by our raters.

Table 4.1: Precision and Recall score of the five libraries and the three heuristics used

Library/Heuristic	Precision			Recall		
	Negative	Neutral	Positive	Negative	Neutral	Positive
L-1: VADER	0.73	0.57	0.73	0.76	0.79	0.26
L-2: Afinn	0.71	0.64	0.50	0.79	0.54	0.48
L-3: TwitterSentiment	0.55	0.50	0.25	0.76	0.01	0.41
L-4: Transformer	0.58	0.83	0.52	0.96	0.03	0.54
L-5: TextBlob	0.90	0.40	0.44	0.43	0.59	0.59
H-1: Majority voting on all five (L-1 to L-5)	0.70	0.60	0.56	0.81	0.52	0.51
H-2: Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-1	0.69	0.68	0.68	0.64	0.59	0.64
H-3: Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-2	0.70	0.63	0.67	0.83	0.61	0.46

Table 4.2: Evaluation of five popular libraries and a few heuristics based on human perspective. The best and second-best results are highlighted in green and yellow color.

Library/Heuristic	F1 Score			Accuracy
	Negative	Neutral	Positive	
L-1: VADER	0.74	0.66	0.39	**66%**
L-2: Afinn	0.75	0.59	0.49	64%
L-3: TwitterSentiment	0.63	0.01	0.31	44%
L-4: Transformer	0.72	0.06	0.53	57%
L-5: TextBlob	0.58	0.48	0.5	52%
H-1: Majority voting on all five (L-1 to L-5)	0.75	0.56	0.53	65%
H-2: Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-1	0.76	0.56	0.54	64%
H-3: Majority voting on L-1, L-2, and L-4. In case of a tie, decide neutral based on only L-4; otherwise, decide based on only L-2	0.76	0.62	0.54	**68%**

Table 4.3: Change-points detected by the three algorithms and the final change-point considered

Time Series	CUSUM	DP	PELT	Change-Point
TS-1	-	15 Feb 2020	15 Feb 2020	15 Feb 2020
TS-2	26 Apr 2020	20 Apr 2020	20 May 2020	5 May 2020
TS-3	14 Mar 2020	11 Mar 2020	21 Mar 2020	16 Mar 2020
TS-4	4 Mar 2020	6 Mar 2020	6 Mar 2020	6 Mar 2020
TS-5	5 Apr 2020	5 Apr 2020	5 Apr 2020	5 Apr 2020
TS-6	-	7 Nov 2019	13 Oct 2019	26 Oct 2019

Table 4.4: Effect size in sentiment trend after the change-points and the interpretation

Category	Sentiment	Change in mean	Cohen's d	Cohen's d Interpretation
Text	Negative	10.55%	1.78	Very Large Increment
	Neutral	-4.26%	-1.36	Very Large Decrement
	Positive	-3.30%	-0.59	Medium Decrement
Image	Negative	5.75%	0.75	Large Increment
	Neutral	8.29%	1.24	Very Large Increment
	Positive	-24.52%	-2.37	Huge Decrement

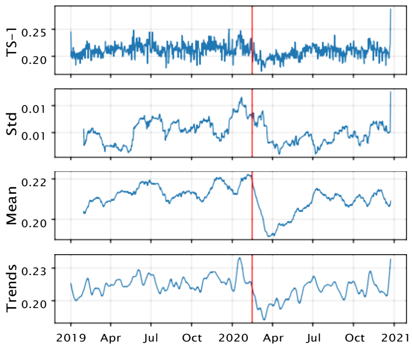
Change-point Identification

The detected change-points are listed in Table 4.3. The CUSUM algorithm was unable to detect any change-point for TS-1 and TS-6. It closely agrees in change-point with the other two in TS-3, TS-4, and TS-5. Since DP and PELT agree on TS-1, this date is considered as the change-point for this time series. The same is true for TS-4 and TS-5. In the case of TS-2 and TS-3, the average between the dates is taken as the change-point.

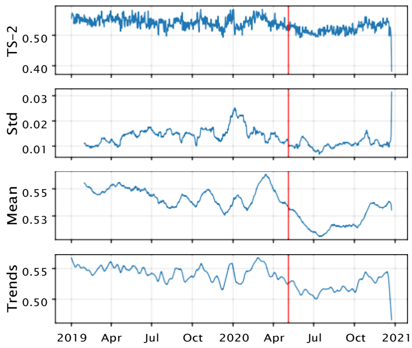
The six sentiment trends along with their corresponding rolling standard deviation, rolling mean, and trend are shown in Figure 4.2. Change points are also marked on each graph with a red vertical line.

Text Sentiments

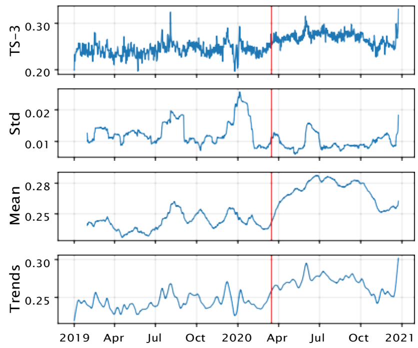
For the positive text sentiments, the change-point was found to be towards the mid of February. As depicted in the graph, there is a steep fall in the trend of the time series at the identified change point. It is also the point where the mean starts to sharply decline after reaching a peak in late February. The time series for the neutral text sentiments shows that the change-point was calculated to be in mid-May. Both the trend and the mean follow a steady decline during that point in time, while the standard deviation shows a natural fluctuation. In the case of the negative text sentiments, the change-point was found to be towards the end of March. At that point, the standard deviation is found to be rising from a peak low and both the trend and the mean continue to show an upward trend.



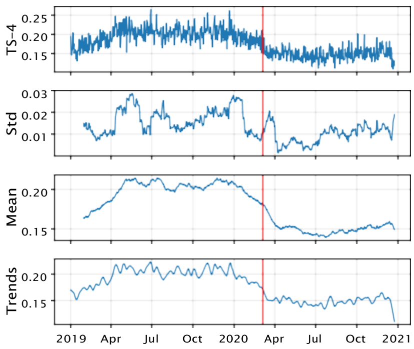
(a) Positive Text



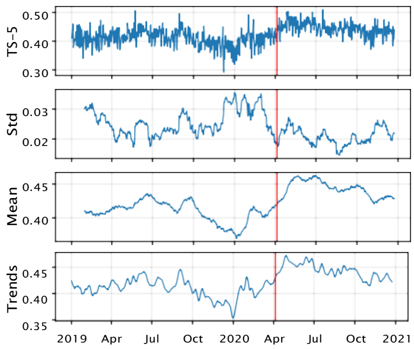
(b) Neutral Text



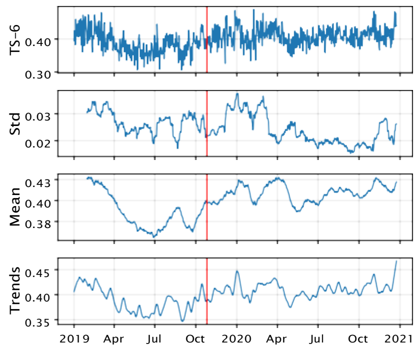
(c) Negative Text



(d) Positive Image



(e) Neutral Image



(f) Negative Image

Figure 4.2: Change-point (marked using red vertical line) for Text and Image Sentiments

Image Sentiments

Analyzing the positive image sentiments, the change-point was found to be in early March when the standard deviation started to rise from a peak low. The trend and mean for this time series continued to decline steadily during the change-point. In the case of the neutral image sentiments, the change-point was identified in early April. This was when the time series mean and trend both showed a steady upward trend whilst the standard deviation began to rise from a peak low. And finally, the change-point for the negative image sentiments was found to be towards the end of October or early November. The mean showed a steady upward trend at that point. The trend of the time series also showed an upward trend with continuous fluctuations. The standard deviation fluctuated naturally at the change-point.

Effect Size of Sentiment Change

The effect size for sentiments is listed in Table 4.4. Cohen's d provides a normalized effect size compared to the change in mean and offers insightful results. For example, we can see that the decrement in neutral text sentiment is much more significant than that of positive sentiment although the change in the mean is almost similar. It also seems that the negative text sentiment

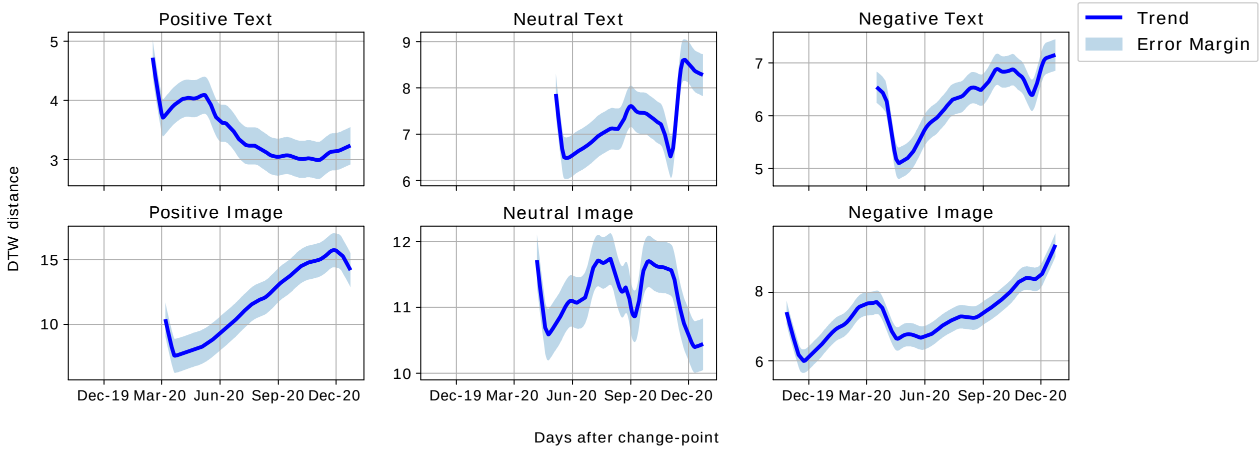


Figure 4.3: Settling trend after the change-point for six time series

accumulates the decreased values from neutral and positive text sentiments.

In terms of image sentiment change, undoubtedly the positive sentiment suffers an exceptional loss. The loss is contributed to both neutral and negative image sentiments, neutral bearing the larger portion.

Settling Trend Identification

Figure 4.3 depicts how the sentiments ‘settled‘ to their normal form after the COVID-19 outbreak and change-point. The fall of DTW in positive text sentiment suggests that the trend quickly returned to normal after June 2020. The DTW distance of both positive and neutral image sentiment remained fairly high whereas the other three (neutral text, negative text, and negative image) distances were average. Among these trends, the positive image showed a steadily increasing distance during the COVID-19 period.

Correlation between Text and Image Sentiment

To perform a bimodal analysis, we divided the two-year timeline into eight quarters. Each data point in the X-axis of Figure 4.4 represents a quarter of a year and the corresponding Y value shows the correlation between text and image sentiment for that quarter’s starting month along with the next two.

Looking at Pearson’s r, we see an increase in inverse correlation between the text and image sentiments as the pandemic progresses. The only exception is during the peak of COVID-19 when we see a strong positive relation between positive text and positive image sentiment. The overall shape of the graph remains similar to Spearman’s and Kendall’s r, but Kendall’s graph

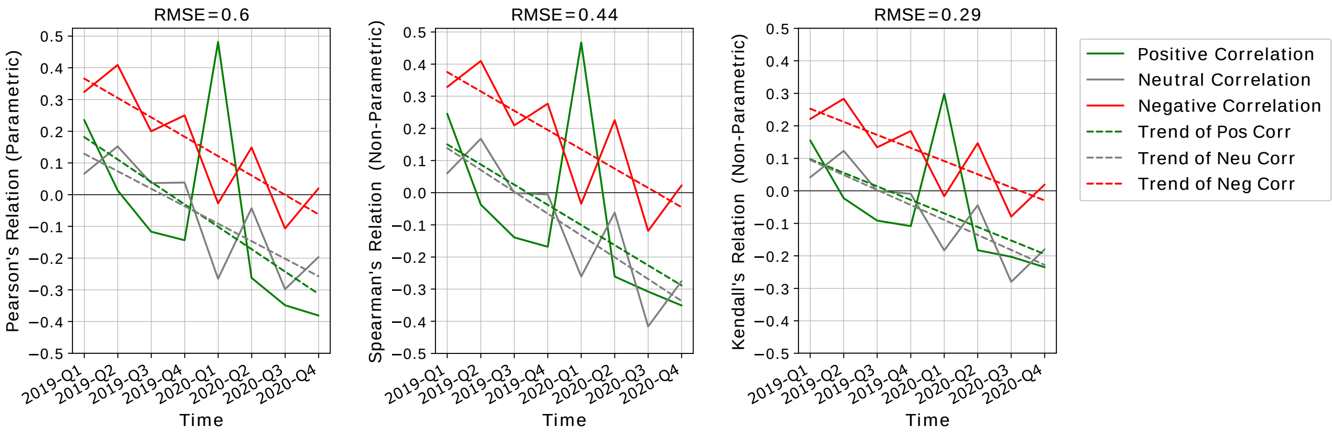


Figure 4.4: Three correlation coefficients (left: Pearson's relation, middle: Spearman's relation, and right: Kendall's relation) denote similarity patterns between text and image sentiments over eight quadrants (2019 to 2020). Each color represents the corresponding sentiment (green for positive, gray for neutral, and red for negative).

appears to be shorter in overall amplitude. Pearson's r and Spearman's r are almost identical except the trend-line for positive is slightly steeper than neutral in Pearson's r.

Series Reconstruction

We have reconstructed the six time series based on the steps mentioned in Section 3.8. The reconstructed series are shown in Figure 4.5. The observations are described below.

Text Sentiments

The construction graph for the positive text sentiments shows that both of the forecasting tools predict a steady increase in positive sentiment after the change point in late February. Orbit predicts a higher upward shift of the percentage of positive sentiment compared to Prophet. The actual trend, however, showed a sharp decline in the proportion of positive sentiment from late January to early March before rising back to its pre-covid trend again from March to July.

The actual trend then reached a plateau, maintaining a steady trend from July onwards. That being said, the percentage of positive text sentiment was significantly lower relative to the forecast by both Orbit and Prophet.

For the neutral text sentiment, interestingly, the predicted trends for Orbit and Prophet were strikingly different after the change point. The actual trend data showed a somewhat similar fluctuation to the forecast by Prophet, declining initially from about 55% in April to 50% in July and then rising again from July onwards to 55% in December 2020. However, the proportion of neutral text sentiment in the actual test data was much lower than the prediction by Prophet where the proportion of neutral sentiment did not fall below 52%.

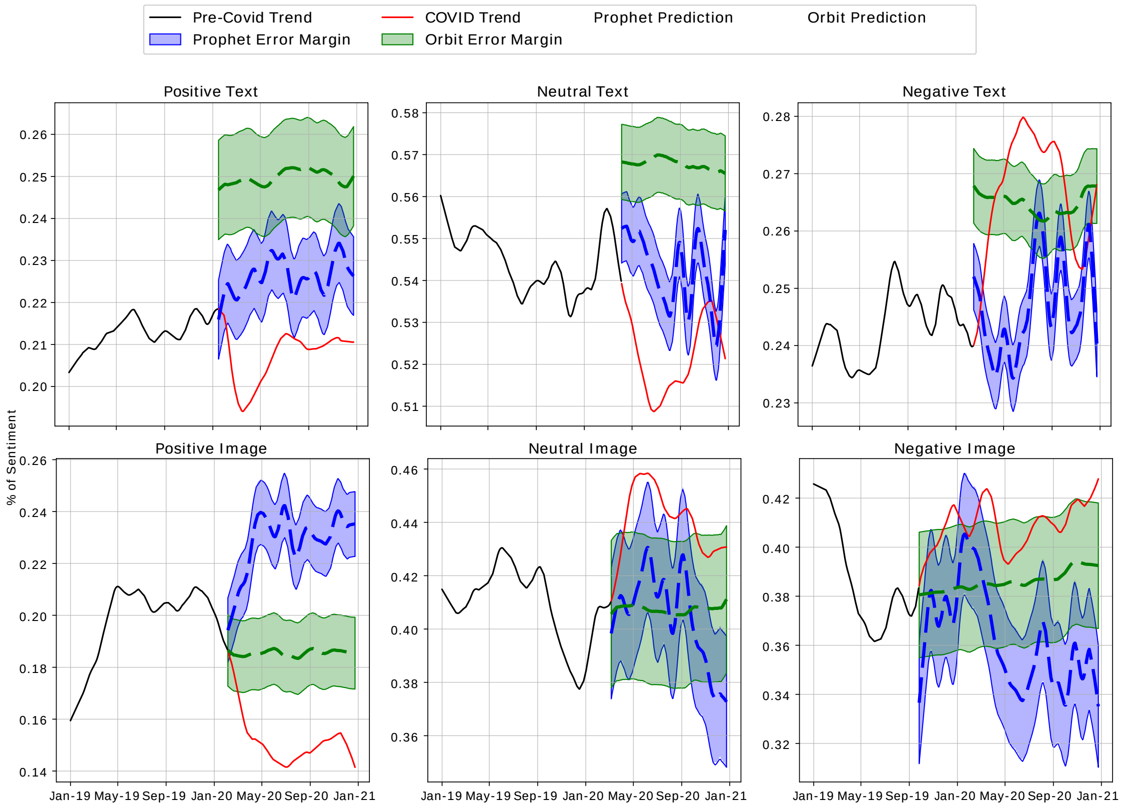


Figure 4.5: Series Reconstruction using Prophet and Orbit library

On the contrary, Orbit forecasted a steady trend at 50%. As seen from the actual trend data, it fluctuated between 50% to 56% and did not follow a steady constant trend as forecasted by Orbit.

The forecast for the negative text sentiment by Orbit is steady at 24% and by Prophet is a rising, fluctuating trend with peak highs of about 27% in early August and mid-September.

The actual trend shows a sharp increase in the percentage of negative text sentiment from about 24% at the changepoint in mid-February. It reached a peak high of 31% in late May before declining again to about 25% by the year-end.

Image Sentiments

.

Analyzing the positive image sentiment, we find that the actual trend data falls sharply from about 19% in the change-point in early February to a peak low of 12% in July before rising back to about 16% in mid-December 2020.

However, the trend data showed a steep decline from mid-December to about 11% at the year-end. This trend of positive image sentiment was in stark contrast to the prediction by both forecasting tools. Orbit showed a steady trend averaging 21% while Prophet showed a sharp rise in the percentage of positive image sentiment from about 19% in the change-point to about 25% in mid-April which stayed relatively steady from April onwards.

Table 4.5: Effect of various smoothing algorithms on series forecasting

TS	Series Description	Mean	Standard Deviation	Ratio	Preferred Smoothing Algorithm(s)
TS-4	Pos Img	0.195	0.025	7.8	None
TS-6	Neg Img	0.389	0.035	11.1	Seasonal Decompose, Gaussian, Polynomial, Convolutional, LOWESS, Spectral, Kalman
TS-3	Neg Text	0.242	0.015	16.1	Seasonal Decompose, Gaussian, Polynomial, Convolutional, LOWESS
TS-5	Neu Img	0.416	0.025	16.6	Seasonal Decompose, Gaussian, Polynomial
TS-1	Pos Text	0.212	0.011	19.3	None
TS-2	Neu Text	0.547	0.015	36.5	None

The forecast for the neutral image sentiment by Prophet showed a fluctuating trend after the change-point averaging about 40% from March to October before falling to about 37% by the year-end. As for Orbit, it forecasted a steady trend at about 37.5% from the change-point to the year-end. The actual trend data for neutral image sentiment showed an increase from 39% in the change-point in early March. The data peaked at about 48% in the second quarter of 2020 before falling slightly to about 44% at the end of the year.

In the case of the negative image sentiment, the prediction by Orbit was a steady slightly rising trend averaging about 33% from the change-point in October 2019 to the end of 2020. The forecast model by Prophet showed a highly fluctuating trend ranging from 32% to about 42.5% from the change-point onwards.

Inspecting the actual trend data for negative image sentiment we see a similar trend to the forecast by Prophet. However, contrary to the Prophet forecast, the actual trend showed a consistently higher percentage of negative sentiment relative to the Prophet model. The actual test data fluctuated in the range of 35% to 48% from the change point till the end of 2020.

Effect of Smoothing Filter

Smoothing appears to have a negligible effect on the improvement of prediction accuracy, according to the results. Only three of the ten smoothing filters improved prediction accuracy for three of the six time series (see Table 4.5). For this reason, we did not use an external smoothing filter for predicting. However, to project the trend of each type of sentiment, the final output series are all filtered using the LOWESS smoothing algorithm (see Section 3.7.1).

Chapter 5

Discussion

In the previous sections, we discussed how the quantification of sentiment change was done for both text and image. This can respond to crucial questions coming from the policy-makers while answering our research questions. In this chapter, we will attempt to answer the research questions (RQ) posed in Section 1.3. At the same time, we will compare our results with other similar works and correlate any similar events.

Quantification of Sentiment Change

To calculate the changes in sentiment, we first identified the changepoints for each of the six time series. The quantification of sentiment change was done for both text and image before and after these changepoints. This answers our **RQ-1** (*How have the sentiments changed over COVID-19 shared on social media?*).

Text Sentiment

In the case of text sentiment, mostly the neutral sentiments became negative. The change-point for positive text is much earlier (Figure 4.2 (a)), which is immediately followed by negative text (Figure 4.2 (c)) whereas the rise of neutral sentiment came much later after around 1.5 months (Figure 4.2 (b)). This indicates that the initial growth of negative sentiment came from positive text and the later (after May) portion came from neutral text.

Image Sentiment

The overall effect size noted in Table 4.4 indicates that the change in sentiment is not the same for text and image. Here we begin to answer our **RQ-2** (*Was the sentiment change the same for image and text?*). The story is quite different for image sentiments. We can see from Figure 4.2(f) that the change-point of the negative image trend was much earlier (during November 2019). The peaks in the standard deviation graph suggest changes that took place during 2020. We can see that the peaks of 2020 in January, April, and November are not significantly different from the early quarter of 2019, suggesting a low actual impact of COVID-19 in the negative images shared. Thus it can be concluded that the consistent low positive image sentiment (Figure 4.2(d)) is due to the rise of neutral image sentiments (Figure 4.2(e)) in 2020.

Effect Size of Change for Sentiments

The effect size calculated in Table 4.4 answers our **RQ-3** (*To what degree did the changes happen for all sentiments?*). The interpretations mentioned in the fifth column suggest that during COVID-19, people started sharing more negative text (10.55% increase) instead of neutral text (4.26% decrease) and they posted more neutral images (8.29% increase) than positive images (24.52% decrease) on average.

Correlation in Text and Images

The longitudinal study of image sentiments was another novelty of our work. Shedding a light on the correlation trend between text and image sentiments can act as a marker or benchmark for identifying subtle incidents similar to a pandemic outbreak. Since Spearman's r is more sensitive to error, the shorter amplitude of Kendall's r (Figure 4.4 (c)) suggests that there is a less overall non-linear correlation between the text and image sentiment. Therefore, the parametric correlation (Pearson's r) begs the most attention for analysis.

The quarterly correlation trend for Pearson's r is described in Table 5.1. This table also has answers to the second part of **RQ-2** (*Was the sentiment change the same for image and text? If not, how are they correlated?*). We can see a notable change in correlation in the post-COVID cyber world. The smaller inverse correlation followed by a strong correlation between positive text and image can be an indicator of a crisis outbreak. This can be seen from Figure 4.2(a) and (d) that both positive sentiment trend for text and image takes a dip after January till March. The overall inverse correlation between text and image in neutral sentiment can also be an indicator of post-pandemic cyberspace.

Settling Trend and Graph Reconstruction

From Figure 4.3 we see an increase in deviation for almost all the trends except positive text sentiment. This graph provides the answer to our final research question, **RQ-4** (*What amount of time did it take for the sentiment change to settle down?*). We also see signification disagreement between the reconstructed graph from both Prophet and Orbit for all sentiments (See Figure 4.5).

Table 5.1: Interpretation of Pearson's r correlation between text and image sentiments over the quarters

Text and Image Sentiment Correlation	Q1-2019	Q2-2019	Q3-2019	Q4-2019	Q1-2020	Q2-2020	Q3-2020	Q4-2020
Negative	Medium Positive	Medium Positive	Medium Positive	Medium Positive	Small Negative	Small Positive	Small Negative	Small Positive
Neutral	Small Positive	Medium Positive	Small Positive	Small Positive	Medium Negative	Small Negative	Medium Negative	Medium Negative
Positive	Medium Positive	Small Positive	Small Negative	Small Negative	Large Positive	Medium Negative	Medium Negative	Medium Negative

Threats to Validity

Since many global events took place during the years 2019 and 2020, it could also be possible that the observed sentiment fluctuation is not entirely due to COVID-19. To mitigate this issue, before collecting tweets for analysis in our study, we carefully selected only those users who were solely interested in COVID-19 through manual inspections. This eliminates potential alternative explanations for the sentiment trend observed in this study by putting control over extraneous variables.

In our study, we shortlisted 720 users who actively engaged in posting COVID-19-related tweets. We then discovered that many of them were not active in 2019, and so the user count became 569. The number would reduce even more if the year span increases. Accordingly, to retain the number of users who were active during both pre- and post-COVID-19 times, we did not extend our period beyond 2019-2020. Here, it is worth mentioning that other recent studies on trend analysis consider around a year [14, 20]. Accordingly, we have our consideration of the two years (or 724 days) for the trend analysis in our study. Nonetheless, we do acknowledge that extending the period further could result in new findings.

Besides, to avoid narrowing our results, we considered images, as well as text content, shared on social media. Here, the COVID-19-related hashtags used in collecting related tweets were also gathered from noteworthy research studies to avoid selection bias. Consideration of content shared over other platforms other than social media or consideration of other hashtags could result in new findings.

Furthermore, we used a complete range (from -1.0 to +1.0) of sentiment for sentiment classification. Here, to distinguish neutral sentiment, we followed the guidelines provided in the documentation of the corresponding sentiment classifier libraries. Besides, to avoid any bias, we manually inspected the tweets and did not filter the tweets of any particular user. Thus, incorporating any other classification of the neutral sentiment or any automated filtering of the tweets could result in different findings.

Implications of Our Study

Initially, in this study, we identified gaps in the existing literature (Section 1.3). These gaps specially present a lacking in the context that is crucial for the policymakers and health workers in making any administrative decision. It is worth mentioning that sentiment analysis over social media interactions has already been focused on by the policy makers [38] and health workers [39]. Here, augmenting the notion of trend analysis could assist them in critically analyzing on the road to taking effective decisions. For example, policymakers and health workers could want to know whether the rise of negative sentiment is actually due to COVID-19, or it is just a seasonal fluctuation. To answer such a question, an analysis needs to cover both text and images in sentiment classification, as people may express their emotions differently through images compared to what they express through text. Our study, for the first time in the literature to the best of our knowledge, contributes to this context.

There also has to be a robust text sentiment classification technique, as most of the social media content is still text-based. However, as found in this study, most of the sentiment libraries are not fit enough to classify social media tweets, especially the tweets that are related to the pandemic. This work also contributes to this context by presenting a new and robust text sentiment classification technique. Besides, even though different frameworks for analyzing sentiments exist in the literature [5], a new complete framework has to be devised to carry out bimodal sentiment analysis to aid researchers with similar interests. In this regard, this study presents the first footstep.

Nonetheless, sentiment analysis over social media has already exhibited a multitude of applications. Examples include analyzing polarity of people towards a particular brand [38, 40] or product [40], opinion mining [41, 42], predicting stock movement [43], aiding political predictions [38, 44, 45], analyzing legal matters [38], e-learning [46], helping in disaster relief management and determining crowds' reaction [5], etc. In all these cases, augmenting with trend analysis through a bimodal investigation opens up new horizons to be explored. In such explorations, this study can contribute substantially.

Comparison with Other Similar Research Studies

Comparing the outcomes of this work with other similar research studies further testifies our findings and observations. For example, a surge in negative text sentiment is discussed in [12], which presents similar findings compared to a part of our observation. However, our findings portray more comprehensive observations that include other two surges during April and June, which completes the full picture. Another research work [13] concludes that the Facebook images shared regarding COVID-19 mostly cluster around the neutral sentiment. Even though this work focuses on a different platform, it reports a finding comparable to our work. Thus, although our study was only based on Twitter users, our outcomes could also be relevant to other platforms.

Another research study [8] explores COVID-19-related text tweets over the second week of April and found a higher percentage of neutral sentiment (45% - 53%). This finding coincides with our observations at that time (50%-55%). Furthermore, the drop in positive text sentiment mentioned in that study is also visible in our study (Figure 4.2 (a)) around April 2020. Another study on tourism-related tweets [14] from April to December 2020 mentions a higher percentage of positive text sentiment. This finding can be a potential explanation behind the steady rising trend of positive sentiment found in our study (Figure 4.2 (a)) after April 2020. The research study in [16], on the other hand, suggests that optimistic, sarcastic, and joking tweets prevail over the monthly tweets (May to July) with a much lower ratio of negative text sentiments. This finding could be consistent with our study considering that the optimistic tweets are perhaps mostly classified as positive by our classifier.

Our conclusions are slightly different from the findings of the research study in [11]. This study states that the proportions of neutral and negative tweets remained fairly high (approximately 35% to 45%) in the first week of March 2020 (before the WHO announced that COVID-19 was a pandemic). The observation was similar in our case for the neutral sentiment, however, not for the negative sentiment. As per our finding, the extent of neutral sentiment was high throughout that period (around 55% at that week). The negative sentiment, however, was low at that time (around 25%) and it rose steadily after the announcement.

Finally, we compare our sentiment trends with that pertinent to other global events. For example, in the case of Russia-Ukraine conflict, the study in [47] suggests that the extent of negative sentiment increases in tweets once the conflict started. Another study was carried out in [48] on stock returns in the fragile market of Zimbabwe between February 2019 and June 2020. It suggested that the sentiments related to stock price have shown a strong coherence with twitter sentiments. The study also reported a significant drop in positive sentiment from July 2019 to October 2019 as well as in March 2020, which coincides with our findings.

Chapter 6

Future Work

Our investigation strictly focused on measuring the positive, negative, and neutral sentiment trends and provided evidence of specific bimodal correlation patterns during the COVID-19 outbreak. We tried to explain the major changes found in the trend with existing literature work, but many of the smaller changes in sentiment can also be explained if topic modeling is done in these time segments for text tweets. Incorporating other metadata such as geographic location, user background, or contemporary event data can also provide a broader context to the change in sentiment trends. We have used Twitter as our data source, which can be a limiting factor as well because this social media is not very popular in all countries. Considering Facebook or LinkedIn in the investigation scope would increase the diversity of this work to a great extent.

Image Captioning

We initially performed image captioning using the Inception V3 model to gather the context of the image, but the generated captions were fairly incorrect and thus practically unusable. Clustering similar images or generating metadata from images as described in [15] can certainly improve the qualitative analysis done in this study. Further, the causal analyses behind different trends in sentiments expressed through the images would also reveal more insights in the future.

Topic Modeling

Much related work has applied Natural Language Processing techniques such as topic modeling using LDA (Latent Dirichlet allocation modeling), named-entity detection, summarization, etc. on the collected tweets. This exploration can suggest what people are talking about and reveal more insights into why the sentiment trends changed the way they did. Other NLP techniques such as N-gram modeling, Named Entity Recognition, summarization, etc. can also provide valuable insights from the tweets.

Geographical Analysis

It is worth mentioning that sometimes the tweets are geo-tagged. Therefore, it could be possible to extract geolocations from the collected tweets, if available. Even if the location data is not present, the author's location can be inferred from the time zone provided with the tweets. Grouping tweets based on the source locations can help identify significant patterns in the sentiment trend. Such a spatial analysis of sentiments in tweets is left as future work.

Chapter 7

Conclusion

In our study, we perform a comprehensive exploration to analyze human sentiments over cyberspace during a pandemic with the necessary context from the previous year. The human interactions with COVID-19-related data and visualizations revealed through this study will help to provide insights regarding mass response during a pandemic on a large scale, as it encompasses the overall sentiments of humans during and after the pandemic. Throughout our work, we also propose an improved heuristic to determine sentiments over text tweets. Additionally, we present a novel bimodal analysis over human sentiments expressed through text and images. We identify the changes that took place over text and images in the COVID-19 scenario and measure the deviation in the sentiments based on the coefficient. We hope that insights found from our analysis will lead to a better understanding of pandemic-like critical situations and will define a way of research during national as well as international crises.

References

[1] A. N. Agency, "History of deadly plagues, epidemics and global pandemicsafpgraphics pic.twitter.com/xn8gb3fna8," Nov 2021.

[2] "The evolution of social media." https://widsix.com/written-word/ the-evolution-of-social-media/, 2022. last accessed on August 25, 2022.

[3] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Computer Science*, vol. 161, pp. 707–714, 2019. The Fifth Information Systems International Conference, 23-24 July 2019, Surabaya, Indonesia.

[4] H. Gao, G. Barbier, and R. Goolsby, "Harnessing the crowdsourcing power of social media for disaster relief," *IEEE Intelligent Systems*, vol. 26, no. 3, pp. 10–14, 2011.

[5] G. Beigi, X. Hu, R. Maciejewski, and H. Liu, "An overview of sentiment analysis in social media and its applications in disaster relief," *Sentiment analysis and ontology engineering*, pp. 313–340, 2016.

[6] K. Ahmed, N. E. Tazi, and A. H. Hossny, "Sentiment analysis over social networks: An overview," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2174–2179, 2015.

[7] MitulMakadia and MitulMakadia, "Applications of sentiment analysis," Oct 2021.

[8] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," *Kurdistan Journal of Applied Research*, pp. 54–65, 2020.

[9] A. D. Dubey, "Twitter Sentiment Analysis During COVID-19 Outbreak," *Available at SSRN 3572023*, 2020.

[10] C. Sitaula, A. Basnet, A. Mainali, and T. Shahi, "Deep learning-based methods for sentiment analysis on nepali covid-19-related tweets," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[11] R. Chandrasekaran, V. Mehta, T. Valkunde, E. Moustakas, *et al.*, “Topics, Trends, and Sentiments of Tweets about the COVID-19 Pandemic: Temporal Infoveillance Study,” *Journal of Medical Internet Research*, vol. 22, no. 10, p. e22624, 2020.

[12] S. Boon-Itt, Y. Skunkan, *et al.*, “Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study,” *JMIR Public Health and Surveillance*, vol. 6, no. 4, p. e21978, 2020.

[13] T. Olaleye, P. Ugege, A. Ademoroti, T. Olomola, O. Ilugbo, and O. Shofoluwe, “Veracity Assessment of Multimedia Facebook Posts for Infodemic Symptom Detection using Bi-modal Unsupervised Machine Learning Approach,” *International Journal for Research in Applied Science Engineering Technology (IJRASET)*, vol. 9, pp. 2234–2241, Dec 2021.

[14] R. Mishra, S. Urolagin, J. Jothi, A. Neogi, and N. Nawaz, “Deep Learning-based Sentiment Analysis and Topic Modeling on Tourism during COVID-19 Pandemic,” *Frontiers in Computer Science*, vol. 3, 2021.

[15] N. Yeung, J. Lai, and J. Luo, “Face off: Polarized Public Opinions on Personal Face Mask Usage During the COVID-19 Pandemic,” in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4802–4810, IEEE, 2020.

[16] R. Chandra and A. Krishna, “Covid-19 sentiment analysis via deep learning during the rise of novel cases,” *Plos one*, vol. 16, no. 8, p. e0255615, 2021.

[17] Q. Yang, H. Alamro, S. Albaradei, A. Salhi, X. Lv, C. Ma, M. Alshehri, I. Jaber, F. Tifratene, W. Wang, *et al.*, “Senwave: monitoring the global sentiments under the covid-19 pandemic,” *arXiv preprint arXiv:2006.10842*, 2020.

[18] T. T. Mengistie and D. Kumar, “Deep learning based sentiment analysis on covid-19 public reviews,” in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 444–449, IEEE, 2021.

[19] J. S. Murthy, A. C. Shekar, D. Bhattacharya, R. Namratha, and D. Sripriya, “A novel framework for multimodal twitter sentiment analysis using feature learning,” in *International Conference on Advances in Computing and Data Sciences*, pp. 252–261, Springer, 2021.

[20] D. Flocco, B. Palmer-Toy, R. Wang, H. Zhu, R. Sonthalia, J. Lin, A. L. Bertozzi, and P. J. Brantingham, “An Analysis of COVID-19 Knowledge Graph Construction and Applications,” in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 2631–2640, IEEE, 2021.

[21] K. L. O’Halloran, G. Pal, and M. Jin, “Multimodal Approach to Analysing Big Social and News Media Data,” *Discourse, Context & Media*, vol. 40, p. 100467, 2021.

[22] H. Yin, X. Song, S. Yang, and J. Li, "Sentiment Analysis and Topic Modeling for COVID-19 Vaccine Discussions," *World Wide Web*, pp. 1–17, 2022.

[23] G. Chandrasekaran and J. Hemanth, "Deep Learning and TextBlob-based Sentiment Analysis for Coronavirus (COVID-19) Using Twitter Data," *International Journal on Artificial Intelligence Tools*, vol. 31, no. 01, p. 2250011, 2022.

[24] Z. B. Nezhad and M. A. Deihimi, "Twitter Sentiment Analysis from Iran about COVID-19 Vaccine," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 16, no. 1, p. 102367, 2022.

[25] S. H. Sahir, R. S. A. Ramadhana, M. F. R. Marpaung, S. R. Munthe, and R. Watrianthos, "Online Learning Sentiment Analysis During the COVID-19 Indonesia Pandemic Using Twitter Data," in *IOP Conference Series: Materials Science and Engineering*, vol. 1156, p. 012011, IOP Publishing, 2021.

[26] M. Sievert, K. Mantsopoulos, S. K. Mueller, M. Eckstein, R. Rupp, M. Aubreville, F. Stelzle, N. Oetter, A. Maier, H. Iro, *et al.*, "Systematic interpretation of confocal laser endomicroscopy: larynx and pharynx confocal imaging score," *Acta Otorhinolaryngologica Italica*, vol. 42, no. 1, p. 26, 2022.

[27] S. Calder, G. Schamberg, C. Varghese, S. Waite, G. Sebaratnam, J. S. Woodhead, P. Du, C. N. Andrews, G. O'Grady, and A. A. Gharibans, "An automated artifact detection and rejection system for body surface gastric mapping," *Neurogastroenterology & Motility*, p. e14421, 2022.

[28] "Fleiss' kappa." https://www.wikiwand.com/en/Fleiss%27_kappa. last accessed on August 25, 2022.

[29] L. Vadicamo, F. Carrara, A. Cimino, S. Cresci, F. Dell'Orletta, F. Falchi, and M. Tesconi, "Cross-Media Learning for Image Sentiment Analysis in the Wild," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 308–317, Oct 2017.

[30] O. A. Grigg, V. Farewell, and D. Spiegelhalter, "Use of risk-adjusted cusum and rsprtcharts for monitoring in medical contexts," *Statistical methods in medical research*, vol. 12, no. 2, pp. 147–170, 2003.

[31] S. C. Xiaodong Jiang, Sudeep Srivastava, "Kats toolkit," 8 2021.

[32] G. D. Wambui, G. A. Waititu, and A. Wanjoya, "The power of the pruned exact linear time (pelt) test in multiple changepoint detection," *American Journal of Theoretical and Applied Statistics*, vol. 4, no. 6, p. 581, 2015.

[33] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

[34] S. S. Sawilowsky, "New effect size rules of thumb," *Journal of modern applied statistical methods*, vol. 8, no. 2, p. 26, 2009.

[35] M. B. Abdullah, "On a robust correlation coefficient," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 39, no. 4, pp. 455–460, 1990.

[36] A. Rebekić, Z. Lončarić, S. Petrović, and S. Marić, "Pearson's or spearman's correlation coefficient-which one to use?," *Poljoprivreda*, vol. 21, no. 2, pp. 47–54, 2015.

[37] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[38] D. Alessia, F. Ferri, P. Grifoni, and T. Guzzo, "Approaches, tools and applications for sentiment analysis implementation," *International Journal of Computer Applications*, vol. 125, no. 3, 2015.

[39] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Computer Science*, vol. 161, pp. 707–714, 2019.

[40] M. Rathan, V. R. Hulipalled, P. Murugeshwari, and H. Sushmitha, "Every post matters: a survey on applications of sentiment analysis in social media," in *2017 International Conference on Smart Technologies For Smart Nation (SmartTechCon)*, pp. 709–714, IEEE, 2017.

[41] A. Mehmood, A. S. Palli, and M. Khan, "A study of sentiment and trend analysis techniques for social media content," *International Journal of Modern Education and Computer Science*, vol. 6, no. 12, p. 47, 2014.

[42] W. Zhang, M. Xu, and Q. Jiang, "Opinion mining and sentiment analysis in social media: Challenges and applications," in *International Conference on HCI in Business, Government, and Organizations*, pp. 536–548, Springer, 2018.

[43] B. Li, K. C. Chan, C. Ou, and S. Ruifeng, "Discovering public sentiment in social media for predicting stock movement of publicly listed companies," *Information Systems*, vol. 69, pp. 81–92, 2017.

[44] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to italy and france," *New media & society*, vol. 16, no. 2, pp. 340–358, 2014.

[45] A. Bakliwal, J. Foster, J. van der Puil, R. O'Brien, L. Tounsi, and M. Hughes, "Sentiment analysis of political tweets: Towards an accurate classifier," Association for Computational Linguistics, 2013.

[46] C. Troussas, A. Krouska, and M. Virvou, "Trends on sentiment analysis over social networks: pre-processing ramifications, stand-alone classifiers and ensemble averaging," in *Machine Learning Paradigms*, pp. 161–186, Springer, 2019.

[47] R. Ibar-Alonso, R. Quiroga-Garc´ıa, and M. Arenas-Parra, "Opinion mining of green energy sentiment: A russia-ukraine conflict analysis," *Mathematics*, vol. 10, no. 14, p. 2532, 2022.

[48] K. Nyakurukwa and Y. Seetharam, "The wisdom of the twitter crowd in the stock market: Evidence from a fragile state," *African Review of Economics and Finance*, vol. 14, no. 1, pp. 203–228, 2022.

Generated using Postgraduate Thesis LaTeX Template, Version 1.03. Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Sunday 18th September, 2022 at 5:42pm.