

## Article

# Investigating and Analyzing Self-Reporting of Long COVID on Twitter: Findings from Sentiment Analysis

Nirmalya Thakur 

Department of Computer Science, Emory University, Atlanta, GA 30322, USA; nirmalya.thakur@emory.edu

**Abstract:** This paper presents multiple novel findings from a comprehensive analysis of a dataset comprising 1,244,051 Tweets about Long COVID, posted on Twitter between 25 May 2020 and 31 January 2023. First, the analysis shows that the average number of Tweets per month wherein individuals self-reported Long COVID on Twitter was considerably high in 2022 as compared to the average number of Tweets per month in 2021. Second, findings from sentiment analysis using VADER show that the percentages of Tweets with positive, negative, and neutral sentiments were 43.1%, 42.7%, and 14.2%, respectively. To add to this, most of the Tweets with a positive sentiment, as well as most of the Tweets with a negative sentiment, were not highly polarized. Third, the result of tokenization indicates that the tweeting patterns (in terms of the number of tokens used) were similar for the positive and negative Tweets. Analysis of these results also shows that there was no direct relationship between the number of tokens used and the intensity of the sentiment expressed in these Tweets. Finally, a granular analysis of the sentiments showed that the emotion of sadness was expressed in most of these Tweets. It was followed by the emotions of fear, neutral, surprise, anger, joy, and disgust, respectively.

**Keywords:** COVID-19; Long COVID; social media; Twitter; big data; data analysis; natural language processing; data science; sentiment analysis



**Citation:** Thakur, N. Investigating and Analyzing Self-Reporting of Long COVID on Twitter: Findings from Sentiment Analysis. *Appl. Syst. Innov.* **2023**, *6*, 92. <https://doi.org/10.3390/asi6050092>

Academic Editor: Dorota Temple

Received: 6 September 2023

Revised: 8 October 2023

Accepted: 10 October 2023

Published: 12 October 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The pandemic of coronavirus disease 2019 (COVID-19) posed an immense menace to public health on a global scale. COVID-19 stems from the infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was initially uncovered and identified in individuals who had been exposed at a seafood market in Wuhan City, situated in the Hubei Province of China in December 2019 [1]. Analogous to the discoveries linked to SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV), it is believed that SARS-CoV-2 has the capacity to leap across species barriers, thus instigating primary infections in humans; currently, its primary mode of transmission predominantly occurs through human-to-human contact and respiratory droplets. Although the mortality rate attributed to COVID-19 is lower when juxtaposed with the rates observed for SARS and MERS, the resultant pandemic linked to COVID-19 has been markedly more severe and devastating [2–4]. As of September 21, 2023, there have been a total of 770,778,396 cases and 6,958,499 deaths due to COVID-19 [5], and many people all over the world are suffering from Long COVID.

### 1.1. Overview of the SARS-CoV-2 Virus and Its Effect on Humans

The SARS-CoV-2 virus particles measure between 60 and 140 nanometers in diameter and boast a positive-sense, single-stranded RNA genome spanning a length of 29,891 base pairs [6]. Upon scrutinizing the genome sequences, it became evident to researchers in this field that SARS-CoV-2 shares a striking 79.5% sequence resemblance with its predecessor, SARS-CoV. Moreover, it exhibits an astonishing 93.1% identity when compared to the genetic makeup of the RaTG12 virus, which was originally isolated from a bat species

known as *Rhinolophus affinis*, dwelling in China's Yunnan Province [7,8]. An exhaustive investigation into the SARS-CoV-2 genome, alongside its predecessor SARS-CoV, revealed the presence of nearly thirty open reading frames (ORFs) and two unique insertions [9]. Upon scrutinizing the genomes of SARS-CoV and bat coronaviruses, it became apparent to researchers in this field that certain regions, such as ORF6, ORF8, and the S gene, exhibit relatively low levels of sequence similarity across coronaviruses as a whole [10,11]. Prior works in this field have determined that the interaction between the SARS-CoV-2 S protein and its cell surface receptor, angiotensin-converting enzyme 2 (ACE2), sets in motion the process of viral entry into type II pneumocytes located within the human lung. This vital interaction marks the commencement of the virus's infiltration into the respiratory system. Consequently, the S protein assumes a pivotal and indispensable role in not only the initial transmission but also the persistent and continual infection caused by SARS-CoV-2. In essence, the S protein stands as a linchpin, orchestrating the virus's ability to penetrate the human body and perpetuate its presence [12].

Although the infection of different organs has been reported in different cases, the SARS-CoV-2 virus usually affects the respiratory systems in patients. Initial research on infections from the SARS-CoV-2 virus in Wuhan, China, has shed light on a variety of symptoms, such as fever, dry cough, difficulty breathing, headache, dizziness, exhaustion, nausea, and diarrhea, that a patient experiences in the first few days of the infection. It is crucial to note that not everyone experiences the same symptoms and that both the extent and severity of these symptoms may vary greatly from individual to individual [13]. After a virus undergoes genetic modifications, it is referred to as having experienced mutations. Because of these genetic alterations, various genetic forms of a virus evolve, which are known as variants. Strains are variants that exhibit differences in their visible traits [14]. The genetic sequences of SARS-CoV-2 were made available to people worldwide on January 10, 2020, by GISAID. Since then, GISAID has made more than 5 million genetic sequences of SARS-CoV-2, obtained from 194 nations and territories, accessible to researchers from different disciplines via their database [15,16].

### 1.2. Concept of "Long COVID"

The phrase "Long COVID" first appeared on social media to characterize the symptoms that persisted after the initial SARS-CoV-2 infection [17,18]. Regardless of one's viral state, "Long COVID" is used as a broad name for a variety of symptoms that last for a long time after contracting the SARS-CoV-2 virus. The term "post-COVID syndrome" is also often used to describe Long COVID [19]. These symptoms may present in a variety of ways, reoccurring often or seldom [20]. These symptoms might include the recurrence of one or more acute COVID symptoms or the appearance of completely distinct symptoms. It is noteworthy that a considerable number of people with post-COVID syndrome test negative for PCR, showing microbiological recuperation. Post-COVID syndrome, put simply, is the interval of time between microbial recuperation and clinical recuperation [20,21]. Patients with Long COVID often exhibit biochemical and radiological improvements. Depending on how long these symptoms last, Long COVID may be divided into two categories: "post-acute COVID", where symptoms last between 3 weeks and 12 weeks, and "chronic COVID", where symptoms last for longer than 12 weeks. Therefore, in a SARS-CoV-2-infected individual, the persistence of one or more symptoms, whether constantly or sporadically, whether they are distinct from or evoke one of acute COVID symptoms past the intended time frame of clinical recuperation is referred to as Long COVID [20–22].

A study conducted in Italy revealed that a significant proportion, specifically 87%, of individuals who had recuperated and were discharged from hospitals exhibited the persistence of at least one lingering symptom even after a span of 60 days [23]. Within this group, it was observed that 32% endured one or two lingering symptoms, while a substantial 55% reported the presence of three or more. Notably, these lingering symptoms did not include fever or signs indicative of acute illness. Instead, the most commonly reported issues included fatigue (53.1%), a decline in overall quality of life (44.1%), difficulty breath-

ing (43.4%), joint pain (27.3%), and chest discomfort (21.7%). Additionally, individuals mentioned experiencing symptoms like cough, skin rashes, palpitations, headaches, diarrhea, and a sensation akin to “pins and needles”. These persistent symptoms significantly hindered the ability of these individuals to perform activities of daily living (ADLs) [24]. Being able to perform ADLs on a routine basis is a crucial component of independent living and also impacts one’s quality of life [25,26]. So, the persistence of such symptoms had a significant impact on the elderly population worldwide, who, in addition to being the diversity group that was most affected by COVID-19, were also among the diversity groups that were most affected by Long COVID [27]. Moreover, the elderly population, as well as patients from other age groups grappling with these enduring symptoms, also reported mental health challenges, including anxiety, depression, and post-traumatic stress disorder. Another study noted that individuals who had been discharged from the hospital after battling COVID-19 continued to struggle with breathlessness and overwhelming fatigue even three months post-discharge [28]. Notably, the prevalence of these residual symptoms varies depending on the context of treatment. Approximately 35% of individuals treated for COVID-19 on an outpatient basis reported lingering symptoms, while the percentage significantly rose to around 87% among cohorts of hospitalized patients [23,29]. According to several recent studies, the frequently observed symptoms of Long COVID include fatigue, headaches, attention-related issues, hair loss, and difficulty breathing [30–33].

Recent research works in this field have indicated that rehabilitation could be effective in addressing certain instances of Long COVID. In the realm of rehabilitation, patients are advised to engage in gentle aerobic exercises adjusted to their personal capabilities. The intensity of these exercises is gradually heightened within manageable limits, typically over a span of four to six weeks. Rehabilitation also encompasses respiratory exercises designed to regulate slow, deep breaths, with the primary aim of bolstering the efficiency of respiratory muscles, particularly the muscles of the diaphragm. Additionally, supplementary behavioral adjustments and psychological support may contribute to enhancing the well-being and mental health of individuals experiencing Long COVID [34–37]. To date, only one randomized-controlled trial (RCT) involving 72 elderly COVID-19 survivors has illustrated that a six-week rehabilitation program resulted in improved breathing, ability to perform exercises, and reduced anxiety [38]. Currently, no pharmaceutical medication has demonstrated the ability to alleviate or mitigate the symptoms associated with Long COVID in an RCT or similar. However, works in this area have reported that paracetamol and non-steroidal anti-inflammatory drugs (NSAIDs) can be utilized to manage specific symptoms of Long COVID, such as fever [21,31]. In particular, mounting evidence indicates that Long COVID bears resemblances to ME/CFS and POTS. Numerous instances of POTS diagnoses following SARS-CoV-2 infection have been reported [39–43]. In a survey involving 1146 individuals with symptoms of Long COVID, 13.5% and 10.3% of them received diagnoses of POTS and ME/CFS, respectively [44].

### *1.3. Relevance of Mining and Analysis of Social Media Data during Virus Outbreaks*

Throughout the annals of history, humanity has grappled with numerous outbreaks of infectious diseases, each inflicting a toll on lives and global economies. Within this context, social media platforms have emerged as invaluable sources of information, capable of offering deep insights into the characteristics and status of such outbreaks. Leveraging the power of text mining, health-related data can be gleaned from platforms like Twitter [45–47]. The comprehensive information contained within Twitter data grants researchers access to different forms of web behavior on the internet [48,49] and user-generated content [50,51], facilitating early response strategies and informed decision-making. The realm of social media mining assumes a pivotal role in monitoring diseases and gauging public awareness of health concerns, thereby enabling proactive disease forecasting [52]. It is worth emphasizing that text analyses of Twitter data have become a focal point in the realm of medical informatics research [53].

The utilization of social media data to conduct syndromic surveillance, with a keen focus on public health-related matters through web-based content, holds paramount importance [54,55]. A significant rationale underpinning this approach is the realization that during an outbreak, social media platforms serve as vital conduits for real-time public sentiment, offering a window into the prevailing concerns and anxieties through user comments and outreach. Of these social media platforms, Twitter stands out as a prominent communication channel during disease outbreaks [56]. Its vast pool of information not only serves to heighten public awareness but also acts as a beacon, illuminating the locations and contexts of outbreaks. This wealth of real-time data from Twitter proves invaluable in shedding light on the multifaceted aspects of a wide range of topics and matters of interest to the scientific community from different disciplines, such as infectious disease outbreaks [57–61], cryptocurrency and stock markets [62,63], public health concerns [64–67], societal problems [68–72], emerging technologies [73,74], human behavior analysis [75–78], and humanitarian issues [79–83], as can be seen from several prior works in these fields, which focused on sentiment analysis and other forms of content analysis of Tweets. Following the COVID-19 epidemic, a growing corpus of studies have used Twitter data to analyze public reactions during this global health emergency [84,85]. However, it is important to note that there is still very little research on Long COVID. Additionally, there has not been any work in this area that has emphasized the study of Tweets wherein individuals self-reported Long COVID. This study aims to address these research gaps by presenting multiple novel findings from a comprehensive analysis of a dataset consisting of 1,244,051 Tweets about Long COVID, posted on Twitter between 25 May 2020 and 31 January 2023.

The remainder of the paper is presented as follows. Section 2 presents a comprehensive review of recent works in this field. A detailed description of the methodology used for this research work is presented in Section 3. In Section 4, the results of this study are described, and the salient contributions of this paper are highlighted. This section also includes a discussion on how the findings of this study help to interpret global public opinion towards Long COVID, and how the contributions of this study are expected to be helpful for public health agencies. Section 4 is followed by the conclusion section, wherein the plans for future work are also stated.

## 2. Literature Review

An overview of recent studies in this area, with a particular emphasis on the study of social media discussions regarding COVID-19 and Long COVID, with Twitter serving as the key platform of interest, is outlined in this section. Shamrat et al. [86] used the k-kNN approach to classify COVID-19-related Tweets into three separate classes—positive, negative, and neutral. Sorting through pertinent Tweets was the first stage in their investigation. After the Tweets had first been filtered, the kNN approach was used to conduct a detailed investigation of the opinions expressed in these Tweets. Sontayasara et al. [87] analyzed Tweets in which people expressed a desire to go to Bangkok during the COVID-19 outbreak. The authors applied an SVM-based classifier and were able to categorize the Tweets as positive, negative, and neutral. The trending themes about COVID-19, as expressed on Twitter, were identified and investigated by Asgari-Chenaghlu et al. [88]. Their work involved deducing the overarching concerns expressed in those Tweets. Amen et al.'s approach [89] involved using a directed acyclic network model to identify abnormalities associated with events related to COVID-19 as communicated on Twitter.

The focus of Lyu et al.'s study [90] was the development of a method specifically designed to examine Tweets related to COVID-19 that had a strong connection to the recommendations made by the Centers for Disease Control and Prevention (CDC) regarding COVID-19. Their main objective was to identify the range of public opinions, including worries, interests, and demands, related to the guidelines of the CDC for COVID-19. Al-Ramahi et al.'s research [91] included filtering and examining Tweets about COVID-19 from 1 January 2020 to 27 October 2020. In their research, the authors specifically analyzed



Tweets in which people expressed varied opinions on whether wearing masks effectively slowed the spread of COVID-19. Jain et al. [92] devised a technique to assign an influence score to individuals who published COVID-19-related Tweets on Twitter. Their research also involved identifying prominent users on Twitter who actively participated in the COVID-19 dialogue on Twitter and contributed towards impacting the global discourse and sentiments. Madani et al. [93] used multiple concepts of machine learning and data analytics and developed a misinformation detection classifier that examined Tweets about COVID-19 using the random forest approach. The accuracy of this classifier was observed to be 79%.

Shokoohyar et al. [94] developed a comprehensive system designed to delve into the expansive realm of Twitter discourse, wherein individuals voiced their opinions concerning the COVID-19-induced lockdown measures in the United States. Their aim was to gain insights into the multifaceted array of sentiments, perspectives, and reactions related to the pandemic response. Chehal et al. [95] developed a framework to analyze the collective mindset of the Indian population, as reflected in their Tweets during the two nationwide lockdowns enforced by the Indian government in response to the COVID-19 crisis. Glowacki et al. [96] developed a systematic approach geared towards the identification and examination of COVID-19-related Tweets, wherein the Twitter community engaged in dialogues surrounding issues of addiction. This initiative provided a unique window into the evolving discourse on addiction amidst the backdrop of a global pandemic. Selman et al.'s explorative study [97] centered on studying Tweets wherein individuals shared heart-wrenching accounts of their loved ones succumbing to COVID-19. Their particular focus honed in on those instances wherein patients faced their final moments in isolation. This research illuminated the profoundly emotional and isolating aspects of the pandemic's impact. Koh et al. [98] endeavored to study Tweets containing specific keywords, spotlighting conversations among Twitter users grappling with profound feelings of loneliness during the COVID-19 era. Their analysis, encompassing a corpus of 4492 Tweets, provided valuable insights into the emotional toll of the pandemic. Mackey et al. [99] zeroed in on Tweets wherein individuals voluntarily reported their symptoms, experiences with testing sites, and recovery status in relation to COVID-19. Their work served as a means for understanding the public's firsthand experiences and interactions with healthcare systems. In [100], Leung et al. aimed to analyze the intricate nuances of anxiety and panic-buying behaviors during the COVID-19 pandemic, with a specific emphasis on the unprecedented rush to acquire toilet paper. Their methodology involved the analysis of 4081 Tweets. This unique analysis shed light on the peculiar buying trends and emotional undercurrents that manifested during the pandemic.

The broad focus of the research conducted by Pokharel [101] was to perform sentiment analyses of Tweets about the COVID-19 pandemic. Specifically, their dataset was developed to comprise Tweets from Twitter users who had shared their location as "Nepal" and had actively engaged in tweeting about COVID-19 between 21 May 2020 and 31 May 2020. The methodology involved using Tweepy and TextBlob to analyze the sentiments of these Tweets. The findings of this study revealed a multifaceted tapestry of emotions coursing through the Nepali Twitter community during that period. While the prevailing sentiment was positivity and hope, there were discernible instances wherein emotions of fear, sadness, and even disgust surfaced. Vijay et al. [102] conducted an in-depth examination of the sentiments conveyed in Tweets related to COVID-19 from November 2019 to May 2020 in India. This study involved the categorization of all Tweets into three distinct categories, namely "Positive", "Negative", and "Neutral". The findings of this investigation showed that, initially, the majority of individuals seemed inclined towards posting negative Tweets, perhaps reflecting the uncertainties and anxieties that characterized the early stages of the pandemic. However, as time progressed, there emerged a noticeable shift in the sentiment landscape, with a growing trend towards the expression of positive and neutral Tweets.

The study conducted by Shofiya et al. [103] involved analyzing Tweets posted by individuals from Canada which contained keywords related to social distancing. The

research employed a two-pronged approach involving the utilization of the SentiStrength tool in conjunction with the Support Vector Machine (SVM) classifier to gauge the sentiments of these Tweets. A dataset comprising 629 Tweets was analyzed in this study. The findings showed that 40% of the analyzed Tweets expressed a neutral emotion, 35% of the Tweets expressed a negative emotion, and 25% of the Tweets expressed a positive emotion. The primary objective of the investigation conducted by Sahir et al. [104] centered on the examination of public sentiment surrounding online learning during the outbreak of the COVID-19 pandemic, specifically in October 2020. This methodology involved document-based text mining and the implementation of the Naïve Bayes classifier to detect the sentiments. The results showed that 25% of the Tweets expressed a positive sentiment, 74% of the Tweets expressed a negative sentiment, and only 1% of the Tweets expressed a neutral sentiment. In a similar work [105], a dataset of Tweets about online learning during COVID-19 was developed.

The study conducted by Pristiyono et al. [106] aimed to delve into the perspectives and viewpoints held by the Indonesian population concerning the COVID-19 vaccine during January 2021. The methodology involved the utilization of RapidMiner for data collection. The results of the sentiment analysis showed that the percentages of negative Tweets, positive Tweets, and neutral Tweets were 56%, 39%, and 1%, respectively. In a similar study [107], the researchers analyzed a corpus of Tweets pertaining to COVID-19 from Ireland. This dataset spanned an entire year, commencing on 1 January 2020, and concluding on 31 December 2020. The primary objective of this study was to perform sentiment analysis. The findings showed that a significant portion of the sentiments were marked by robust criticism, with particular emphasis placed on Ireland's COVID tracker app. The critique was predominantly centered on concerns related to privacy and data security. However, the study also showed that a noteworthy segment of sentiments also expressed positivity. These positive sentiments were directed towards the collective efforts and resilience towards COVID-19. Awoyemi et al. [108] used the Twarc2 tool to extract a corpus of 62,232 Tweets about Long COVID posted between 25 March 2022 and 1 April 2022. The methodology involved the application of different natural language processing-based approaches, including the Latent Dirichlet Allocation (LDA) algorithm. The outcome of this analysis highlighted the top three prevailing sentiments in the Twitter discourse. Notably, trust emerged as a prominent sentiment, constituting 11.68% of the emotional spectrum, underlining a notable level of faith and reliance within the online community. Fear also held a significant presence, encompassing 11.26% of the emotional landscape, reflecting concerns and apprehensions embedded in the discussions. Alongside fear, sadness played a substantial role, constituting 9.76% of the emotional spectrum, indicative of the emotional weight carried by certain aspects of the discourse. The primary aim of the work of Pitroda et al. [109] was to delve into the collective societal responses and attitudes surrounding the phenomenon of Long COVID, as expressed on Twitter. A dataset comprising 98,386 Tweets posted between 11 December 2021 and 20 December 2021 was used for this study. The methodology for performing sentiment analysis utilized the AFFIN lexicon model. The findings showed that 44% of the Tweets were negative, 34% of the Tweets were positive, and 23% of the Tweets were neutral. To summarize, there has been a significant amount of research and development in this field since the beginning of COVID-19. However, these research papers have some major drawbacks.

- Insufficient attention towards the phenomenon of Long COVID: These studies have covered various topics pertaining to COVID-19 such as traveling [87], current trends [88], worries of the general public [88], evaluation of events [89], opinions on mask-wearing [91], inquiries into influencer behaviors [92], detecting and tracking misinformation [93], studies of addiction trends [96], identifying loneliness [98], and evaluations of impulse purchases [100]. In the last couple of years, scholars from different disciplines have also conducted thorough investigations by analyzing pertinent Tweets in order to delve into the diverse inquiries of the global population in the context of COVID-19. However, such works [87–89,91–93,96,98,100] did not investigate Tweets pertaining to Long COVID.

- Limitations in the existing Long COVID studies: Although a few studies (e.g., [108,109]) have examined Long COVID-related Tweets, a significant constraint of these studies is the restricted temporal scope of the analyzed Tweets. For example, the research conducted in [108] focused its investigation on a specific timeframe from 25 March 2022 to 1 April 2022. Similarly, the investigation in [109] studied Tweets about Long COVID published between 11 December 2021 and 20 December 2021. These time periods constitute just a fraction of the total span for which Long COVID has had a lasting impact on the global population.
- The emotions of the general public pertaining to Long COVID have not been investigated in previous sentiment analysis-based research works ([86,87,94,101–104,106,107]) that involved opinion mining of COVID-19-related Tweets.
- Studying the self-reporting of healthcare conditions on Twitter has garnered attention from scholars across many disciplines, as can be seen from multiple studies wherein Tweets related to the self-reporting of mental health problems [110], autism [111], Alzheimer's [112], depression [113], breast cancer [114], swine flu [115], flu [116], chronic stress [117], post-traumatic stress disorder [118], and dental issues [119] were analyzed. In light of the emergence of the COVID-19 pandemic, scholarly investigations in this domain, such as [99], have been focused on developing approaches to examine Tweets wherein individuals voluntarily disclosed symptoms related to COVID-19. However, previous studies have not specifically examined Tweets pertaining to the self-reporting of Long COVID.

This study attempts to address these challenges by conducting an analysis of a database of Tweets wherein individuals self-reported Long COVID. The step-by-step methodology that was followed for this work is described in Section 3.

### 3. Methodology

This section is divided into two parts. In Section 3.1, a theoretical overview of sentiment analysis, the specific sentiment analysis approach that was used for this work and its salient features, and a technical overview of RapidMiner, the data science platform used for this work, are presented. In Section 3.2, the step-by-step methodology that was followed for this work is presented and discussed.

#### 3.1. Theoretical Overview of Sentiment Analysis and Technical Overview of RapidMiner

Sentiment Analysis, often referred to as Opinion Mining, represents the computational exploration of individuals' sentiments, viewpoints, and emotional expressions regarding a given subject. This subject can encompass various entities, such as individuals, events, or topics [120]. The terms Sentiment Analysis (SA) and Opinion Mining (OM) are commonly used interchangeably, denoting a shared essence. Nevertheless, some scholars have posited subtle distinctions between OM and SA [121,122]. Opinion Mining, in its essence, involves the extraction and examination of people's opinions pertaining to a specific entity. In contrast, Sentiment Analysis seeks to identify the sentiment embedded within a text and subsequently analyze it. Consequently, SA endeavors to unearth opinions, discern the sentiments they convey, and classify these sentiments based on their polarity. This classification process can be envisioned as a three-tiered hierarchy comprising document-level, sentence-level, and aspect-level SA. At the document level, the primary objective is to categorize an entire opinion document as either expressing a positive or negative sentiment. Here, the document functions as the fundamental information unit typically focused on a single overarching topic or subject. In sentence-level SA, the aim is to classify the sentiment within each individual sentence. The initial step involves distinguishing between subjective and objective sentences. Subsequently, for subjective sentences, sentence-level SA ascertains whether they convey positive or negative opinions [123]. It is worth noting that Wilson et al. [124] highlighted that sentiment expressions may not always possess a subjective nature. However, the distinction between document and sentence-level classifications is not fundamentally profound since sentences can be regarded as concise documents [125].

While document and sentence-level classifications offer valuable insights, they often fall short of providing the granular details necessary for evaluating opinions on various facets of the entity. To obtain this comprehensive understanding, aspect-level SA comes into play. This level of analysis endeavors to categorize sentiments with respect to specific aspects or attributes associated with entities. The initial step involves the identification of these entities and their respective facets. Importantly, opinion-holders can articulate diverse sentiments concerning distinct aspects of the same entity. In essence, SA or OM encompasses a multifaceted process that spans various levels of analysis, from overarching documents to individual sentences and, ultimately, the nuanced evaluation of specific aspects related to entities. This comprehensive approach to sentiment analysis is invaluable in unveiling the intricate tapestry of opinions and emotions expressed in text data, enabling a deeper understanding of public sentiment in various contexts [126].

The examination of emotions can encompass a range of techniques, such as human annotation, Linguistic Inquiry and Word Count (LIWC), Affective Norms for English Words (ANEW), the General Inquirer (GI), SentiWordNet, and machine learning-based approaches like Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM). However, the methodology used in this study included the utilization of VADER, an acronym for Valence Aware Dictionary for Sentiment Reasoning [127]. The selection of VADER as the sentiment analysis methodology is influenced by many considerations. First of all, VADER demonstrates outstanding efficiency, surpassing manual annotation with respect to both precision and efficacy. Furthermore, previous research has indicated that VADER effectively addresses the limitations encountered by other sentiment analysis approaches. The following presents an overview of some of the unique features of VADER, as well as drawbacks in alternate sentiment analysis approaches, which are not present in VADER:

- (a) VADER differentiates itself from LIWC by exhibiting enhanced sensitivity towards sentiment patterns that are often seen in the analysis of texts from social media.
- (b) The General Inquirer has a limitation in its incorporation of sentiment-relevant linguistic components frequently observed in conversations on social media.
- (c) The ANEW lexicon exhibits a reduced degree of reactivity regarding the linguistic components often linked to emotion in social media posts.
- (d) The SentiWordNet lexicon exhibits a significant level of noise, as a noteworthy fraction of its synsets lack clear opposite polarity.
- (e) The Naïve Bayes classifier is predicated on the premise of feature independence, which might be considered a simplistic premise. VADER's more nuanced strategy effectively addresses this limitation.
- (f) The Maximum Entropy approach integrates the concept of information entropy by providing feature weightings without making the assumption of conditional independence between features.
- (g) Both machine learning classifiers and validated sentiment lexicons face the same obstacle of requiring a significant quantity of data for training. Furthermore, the efficacy of machine learning algorithms is contingent upon the training set's ability to correctly capture a diverse array of properties.

The VADER technique is characterized by the implementation of a succinct rule-based architecture, enabling the creation of a specialized sentiment analysis algorithm tailored to the vocabulary often seen on social media sites. The framework exhibits remarkable adaptability in its ability to adjust to several contexts without requiring domain-specific learning information. Instead, it utilizes a versatile sentiment lexicon determined by valence, which has undergone rigorous evaluation by human experts to establish its credibility as a consistent benchmark. The VADER approach is renowned for its exceptional efficacy due to its ability to perform the real-time evaluation of continuous information streams. In addition, the system is capable of maintaining its computational efficiency, as evidenced by its time complexity of  $O(N)$ . Furthermore, it is important to highlight that VADER is easily accessible to everyone without any prerequisites for subscription or financial transactions.



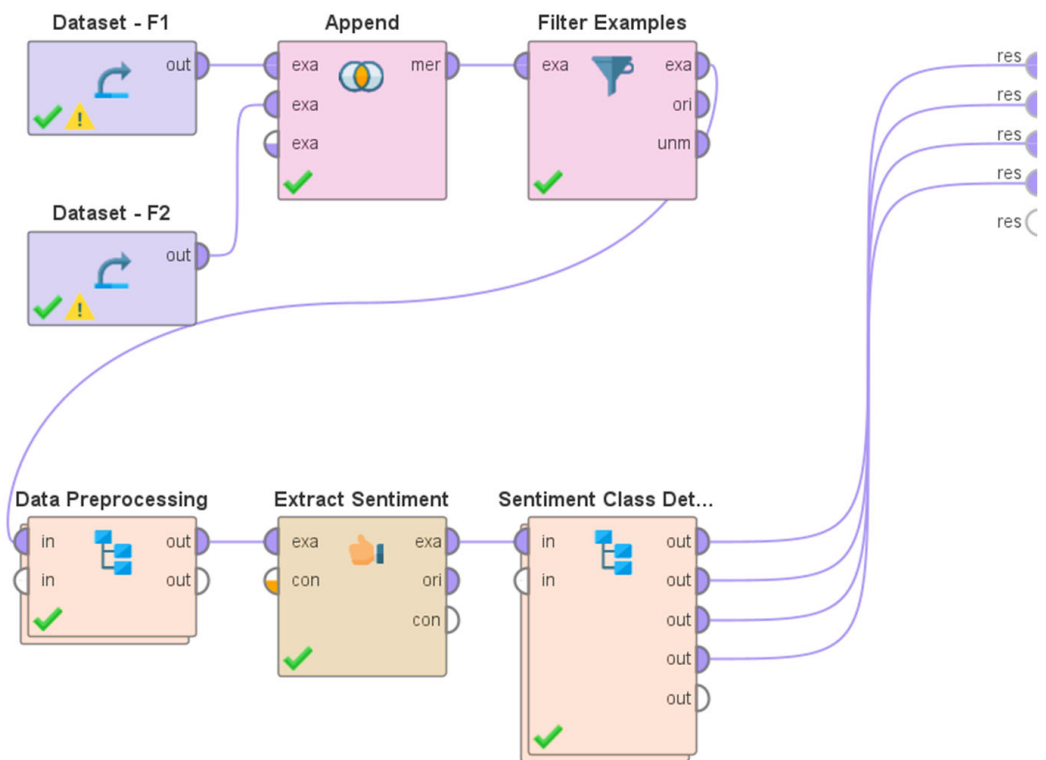
Finally, VADER possesses the ability to evaluate the intensity of sentiment expressed in texts, in addition to categorizing it into positive, negative, or neutral polarities.

In order to set up the system design for sentiment analysis in this study, RapidMiner was utilized [128]. RapidMiner, initially referred to as Yet Another Learning Environment (YALE), is a highly adaptable data science software framework that facilitates the development, execution, and utilization of a broad spectrum of models and algorithms from different disciplines. RapidMiner Studio adopts an open-core paradigm with a comprehensive Graphical User Interface (GUI) designed to allow users to create a variety of applications and processes. Additionally, it facilitates features to support the design and implementation of algorithms. In RapidMiner, designated actions or functions are referred to as “operators”. These operators may be organized in a logical sequence, either sequentially in a hierarchy or in some combination of both, to form a “process” aimed at accomplishing a certain purpose or task. RapidMiner facilitates the design and implementation of these “processes” by providing a wide range of already assembled “operators” that may be readily used. Moreover, there exists a distinct category of “operators” that may be used to alter the fundamental attributes of other “operators”, introducing an additional level of adaptability and personalization to the platform.

### 3.2. Overview of the System Architecture and Design

To develop a “process” in RapidMiner for addressing this research problem, at first, a relevant dataset had to be identified. The dataset used for this work is a dataset of 1,244,051 Tweets about Long COVID, posted on Twitter between 25 May 2020 and 31 January 2023 [129]. This dataset contains a diverse range of topics represented in the Tweets about Long COVID, present in two different files. To develop the “process” (shown in Figure 1), these files were imported into RapidMiner and merged using the Append “operator”. As the focus of this study is to analyze Tweets about Long COVID wherein individuals self-reported Long COVID, a text processing-based approach was applied to identify such Tweets. By reviewing similar works in this field that studied Tweets involving self-reporting of mental health problems [110], autism [111], dementia [112], depression [113], breast cancer [114], swine flu [115], flu [116], chronic stress [117], post-traumatic stress disorder [118], and dental issues [119], a Bag of Words was developed that contained specific phrases related to the self-reporting of Long COVID. These phrases were—“I have long COVID”, “I am experiencing long COVID”, “I have been experiencing long COVID”, “I am suffering from long COVID”, “I am going through long covid”, “I have been feeling symptoms of long covid”, “I have had long covid”, “I have been diagnosed with long covid”, “I had long covid”, “I am feeling symptoms of long covid”, “I am experiencing long covid”, “I felt symptoms of long covid”, “I experienced long covid”, and “I am diagnosed with long covid”. A data filter was used to apply the Bag of Words methodology to the Tweets, resulting in a set of 7348 Tweets in which individuals self-reported Long COVID. Subsequently, all these Tweets underwent data preprocessing. The procedure for data preprocessing included the following sequential steps:

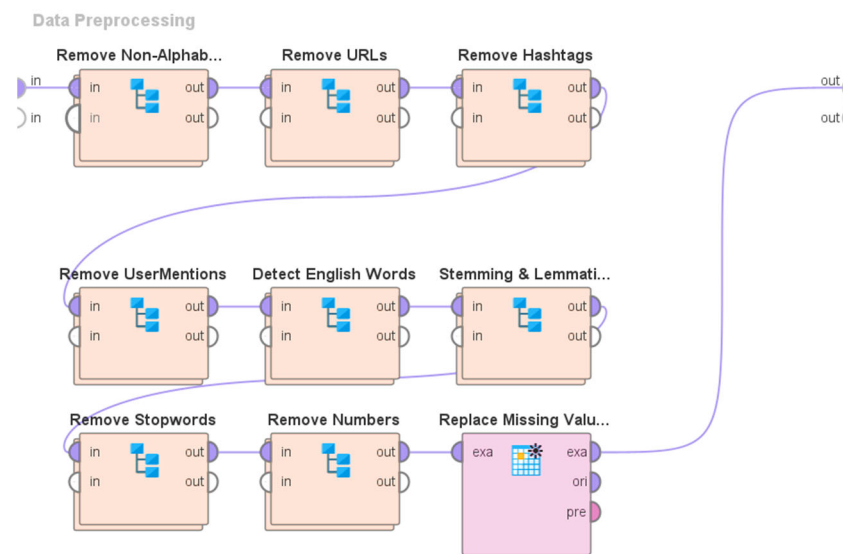
- (a) The elimination of non-alphabetic characters.
- (b) The elimination of URLs.
- (c) The elimination of hashtags.
- (d) The elimination of user mentions.
- (e) The identification of English words using the process of tokenization.
- (f) Stemming and Lemmatization.
- (g) The elimination of stop words.
- (h) The elimination of numerical values.
- (i) Addressing missing values.



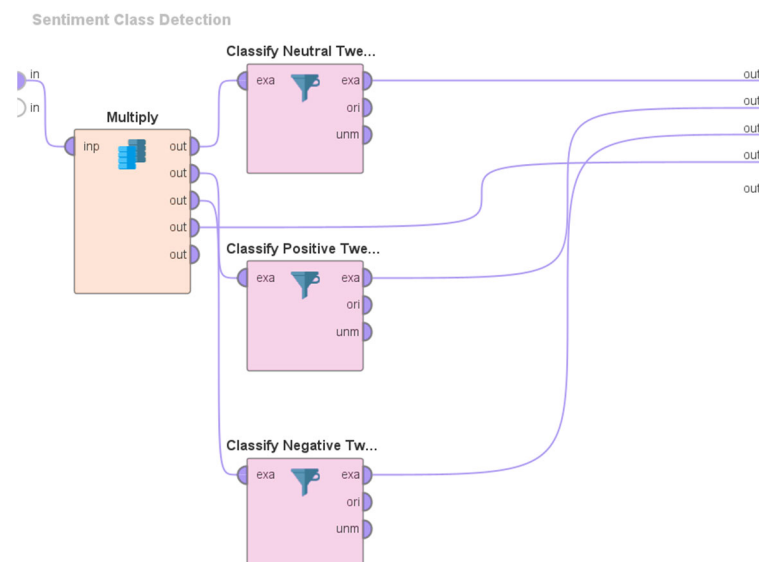
**Figure 1.** The main “process” that was developed in RapidMiner to perform the sentiment analysis of Tweets (posted on Twitter between 25 May 2020 and 31 January 2023) wherein individuals self-reported Long COVID.

A number of functionalities were built in RapidMiner for the purpose of performing data preprocessing. Subsequently, the Extract Sentiment operator within the RapidMiner platform was tailored to utilize the VADER sentiment analysis methodology for the purpose of analyzing this corpus of 7348 Tweets. The VADER approach calculated a compound sentiment score ranging from +4 to −4 for all these Tweets.

A methodology following a set of specific rules was devised to deduce the number of positive, negative, and neutral Tweets. The operational mechanism of this rule-based approach included the analysis of the compound sentiment score assigned to every single Tweet. If the compound sentiment score was greater than zero, the Tweet was categorized as positive; conversely, if the compound sentiment score was less than zero, the Tweet was categorized as negative. In the event that the compound sentiment score was exactly 0, the Tweet was categorized as neutral. This technique consisted of various functions that operated sequentially. These functions were implemented as “operators” grouped as a “sub-process” within the primary “process”, which was designed and implemented in RapidMiner for the purpose of performing sentiment analysis. Figure 1 depicts this primary “process”. In Figure 1, “Dataset-F1” and “Dataset-F2” are used to refer to the two files in the dataset [129] that were imported into RapidMiner for system design and development. This “process”, shown in Figure 1, comprises two sub-processes for data preprocessing and the rule-based identification of sentiment classes. The specific “operators” that comprised these sub-processes are shown in Figures 2 and 3, respectively. The results of running this “process” on all the Tweets in the dataset are presented and discussed in Section 4.



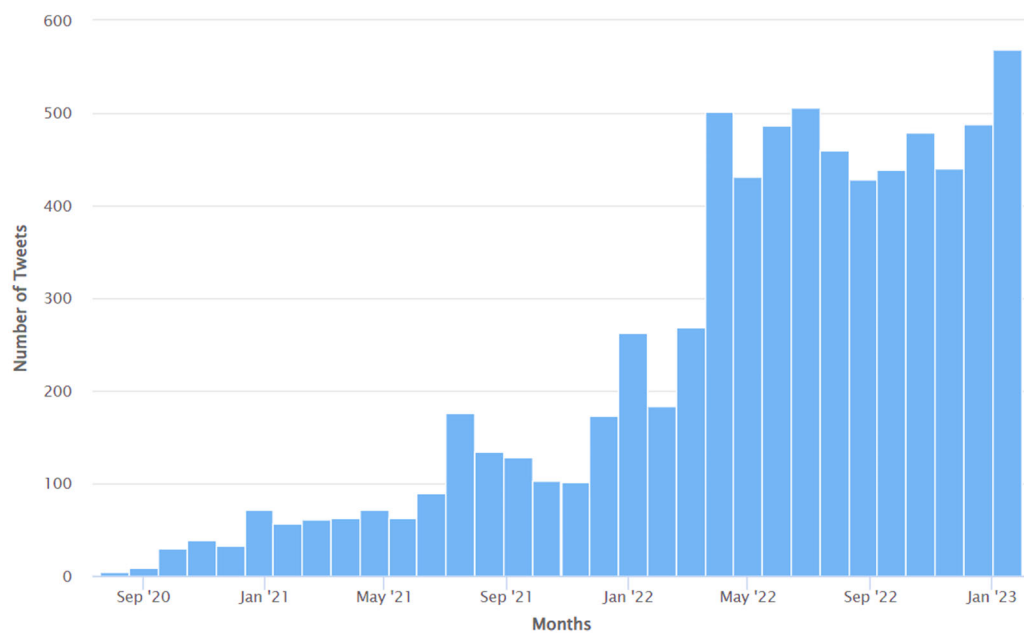
**Figure 2.** Representation of the “operators” that comprised the data preprocessing “sub-process” shown in Figure 1.



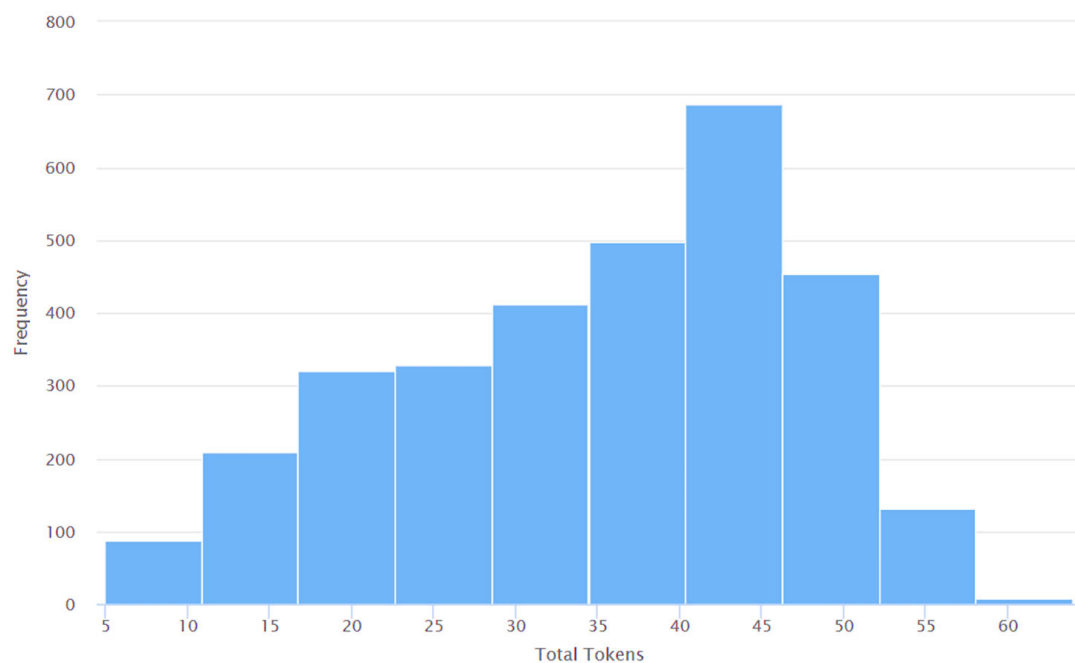
**Figure 3.** Representation of the “operators” that comprised the sentiment class detection “sub-process” shown in Figure 1.

#### 4. Results and Discussions

This section presents the results of performing sentiment analyses of Tweets (posted on Twitter between 25 May 2020 and 31 January 2023) wherein individuals self-reported Long COVID. As described in Section 3, a text-processing-based approach that involved the development of a Bag of Words was utilized to develop a corpus of Tweets, each of which involved the self-reporting of Long COVID. Figure 4 shows a monthly variation of these Tweets during this time range, i.e., between 25 May 2020 and 31 January 2023, using a histogram. To analyze this monthly variation of Tweets, a total of 32 bins were used. As can be seen from this Figure, the average number of Tweets per month wherein individuals self-reported Long COVID on Twitter was considerably high in 2022 as compared to the average number of Tweets per month in 2021. As outlined in Section 3, the process of tokenization was conducted before applying VADER. Figure 5 presents a histogram-based illustration that showcases the tweeting patterns of Tweets with a negative sentiment, specifically focusing on the number of tokens utilized in these Tweets.



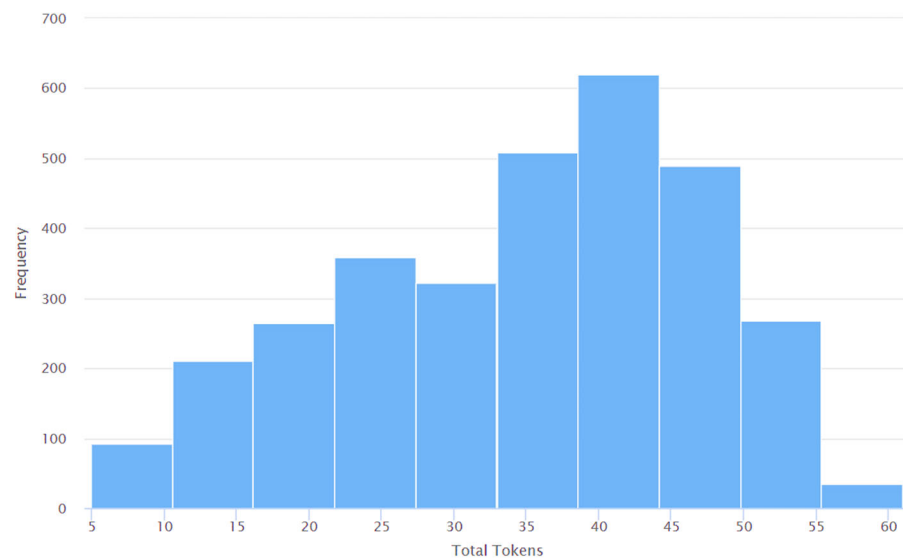
**Figure 4.** Representation of the monthly variation in Tweets (posted on Twitter between 25 May 2020 and 31 January 2023) wherein individuals self-reported Long COVID.



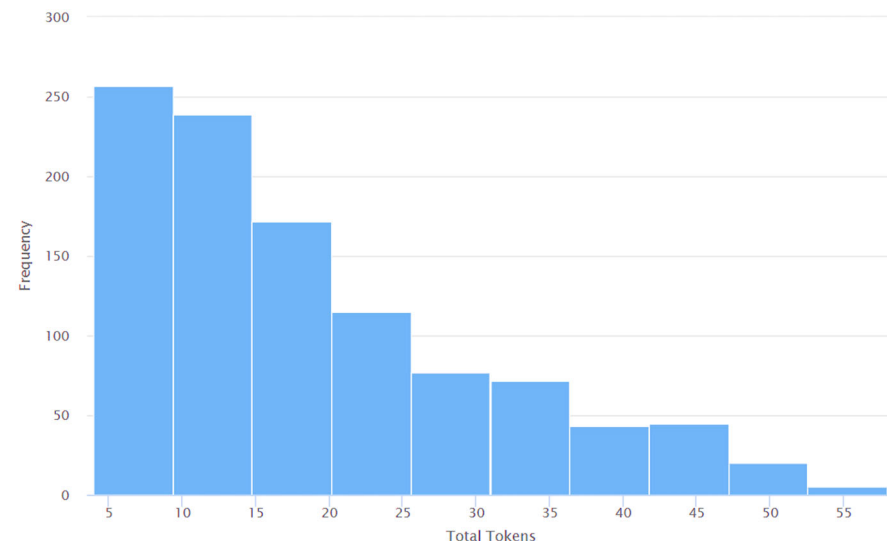
**Figure 5.** Representation of the variation in token usage in Tweets with a negative sentiment.

Figures 6 and 7 present illustrations of the tweeting trends, specifically focusing on the number of tokens utilized for positive and neutral Tweets, respectively. Figures 5 and 6 indicate that there was an identical pattern of utilization of tokens in Tweets expressing both positive and negative sentiments. Based on the analysis of Figure 7, it can be inferred that a considerable proportion of Tweets with a neutral emotion contained a small number of tokens. Additionally, this figure also illustrates that a significant proportion of Tweets with a neutral emotion featured a different tweeting pattern, characterized by different numbers of tokens utilized, in comparison to positive and negative Tweets.





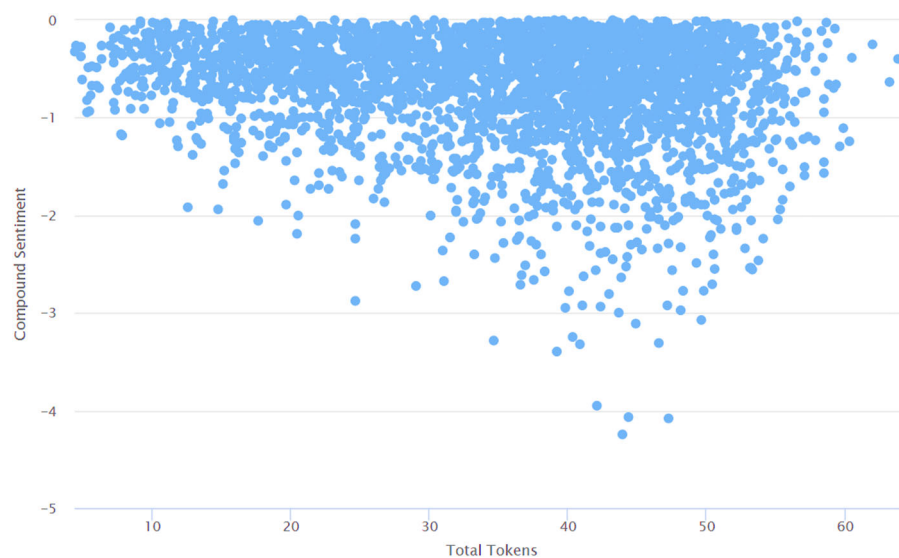
**Figure 6.** Representation of the variation in token usage in Tweets with a positive sentiment.



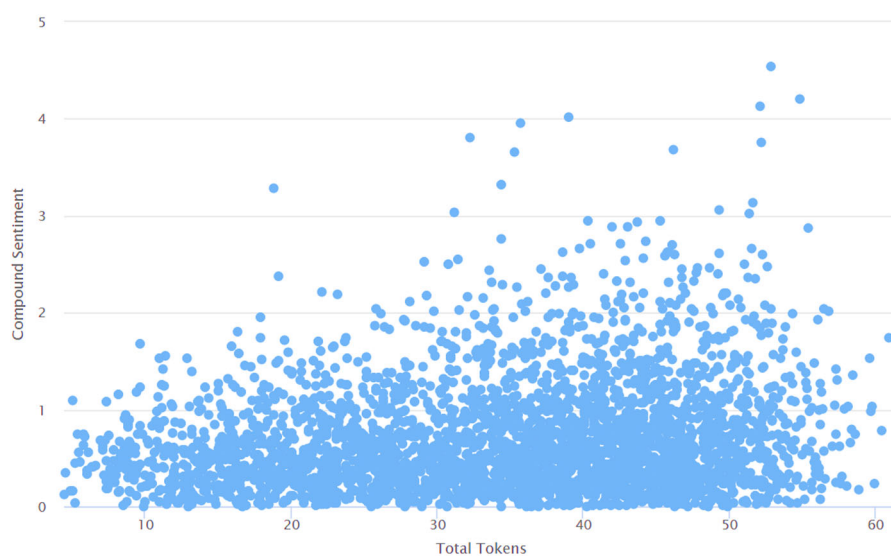
**Figure 7.** Representation of the variation in token usage in Tweets with a neutral sentiment.

Figure 8 illustrates the variability of negative emotion in Tweets, in terms of the intensity of the sentiment and the number of tokens utilized. In Figure 8, the X-axis denotes the number of tokens utilized in the Tweets, while the Y-axis denotes the compound sentiment score assigned to these Tweets. Figures 9 and 10 provide similar illustrations of the relationship between the total number of tokens and the compound sentiment scores for positive and neutral Tweets, respectively.

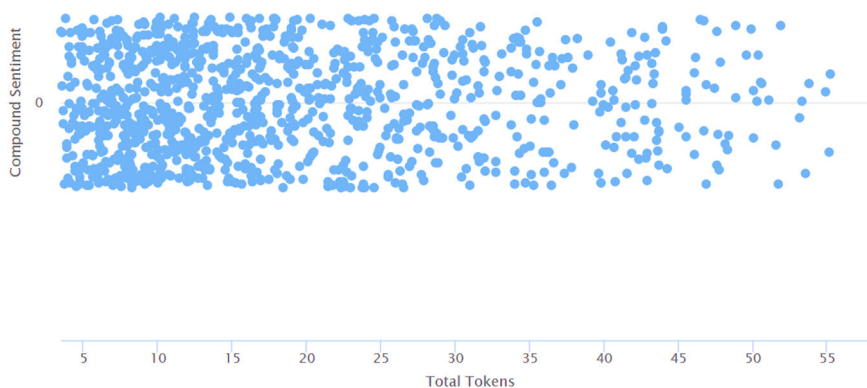
Figures 8 and 9 illustrate that the tweeting patterns of the global population in terms of the utilization of tokens in positive and negative Tweets were diverse. Furthermore, from these figures, it can also be concluded that no correlation existed between the number of tokens utilized in Tweets and the intensity of the expressed sentiment (ranging from +4 to 0 for positive Tweets and 0 to −4 for negative Tweets) in Tweets wherein individual self-reported Long COVID. The results of the RapidMiner “process” depicted in Figure 1 are illustrated in Figure 11. As can be seen from Figure 11, the percentages of Tweets (correct to one decimal place) with positive, negative, and neutral sentiments were 43.1%, 42.7%, and 14.2%, respectively.



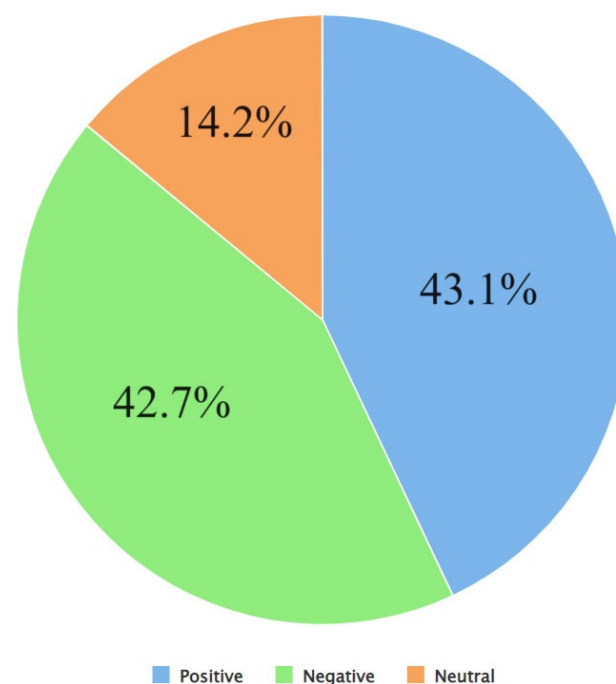
**Figure 8.** Representation of the variation in token usage and the intensity of the expressed sentiment (from the compound sentiment score) in Tweets with a negative sentiment.



**Figure 9.** Representation of the variation in token usage and the intensity of the expressed sentiment (from the compound sentiment score) in Tweets with a positive sentiment.

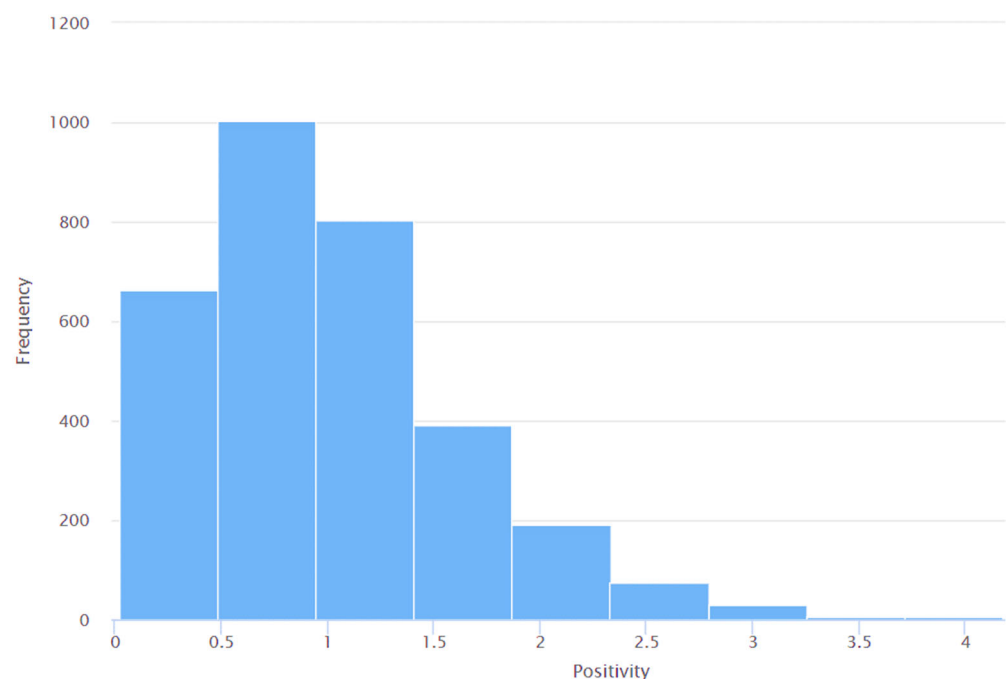


**Figure 10.** Representation of the variation in token usage and the intensity of the expressed sentiment (from the compound sentiment score) in Tweets with a neutral sentiment.

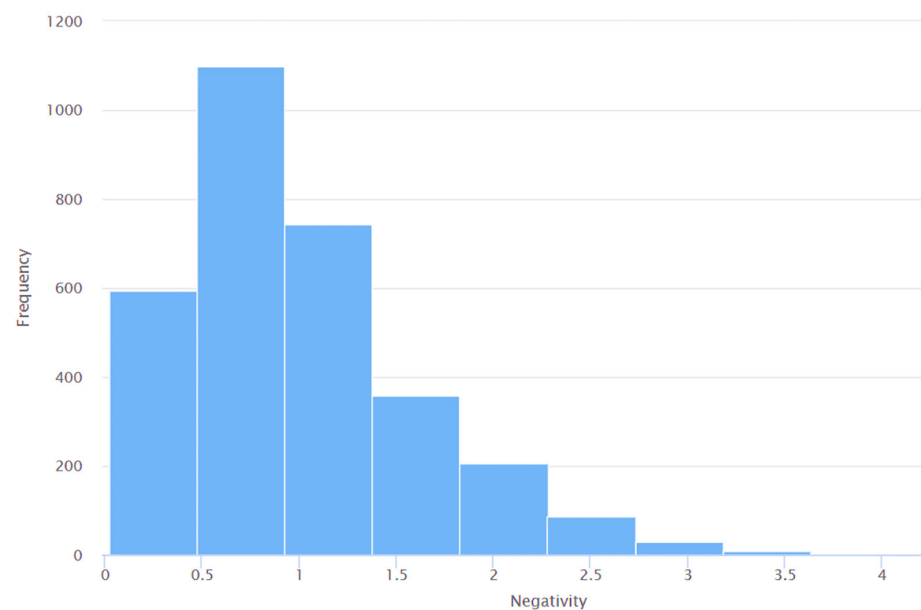


**Figure 11.** Representation of the results of the RapidMiner “process” shown in Figure 1.

Figure 12 is a histogram that presents the distribution of Tweets categorized by their levels of positivity, as determined by the RapidMiner “process” depicted in Figure 1. Similarly, Figure 13 displays a histogram that illustrates the number of Tweets categorized by varied levels of negativity. Figures 12 and 13 reveal a consistent trend in the variability of positive and negative sentiments within the analyzed Tweets. This analysis shows that a majority of Tweets wherein a positive sentiment was expressed, as well as a majority of Tweets wherein a negative sentiment was expressed, were not highly polarized. In other words, most of these Tweets did not convey a high positivity (+4 on the VADER scale) or a high negativity (−4 on the VADER scale).

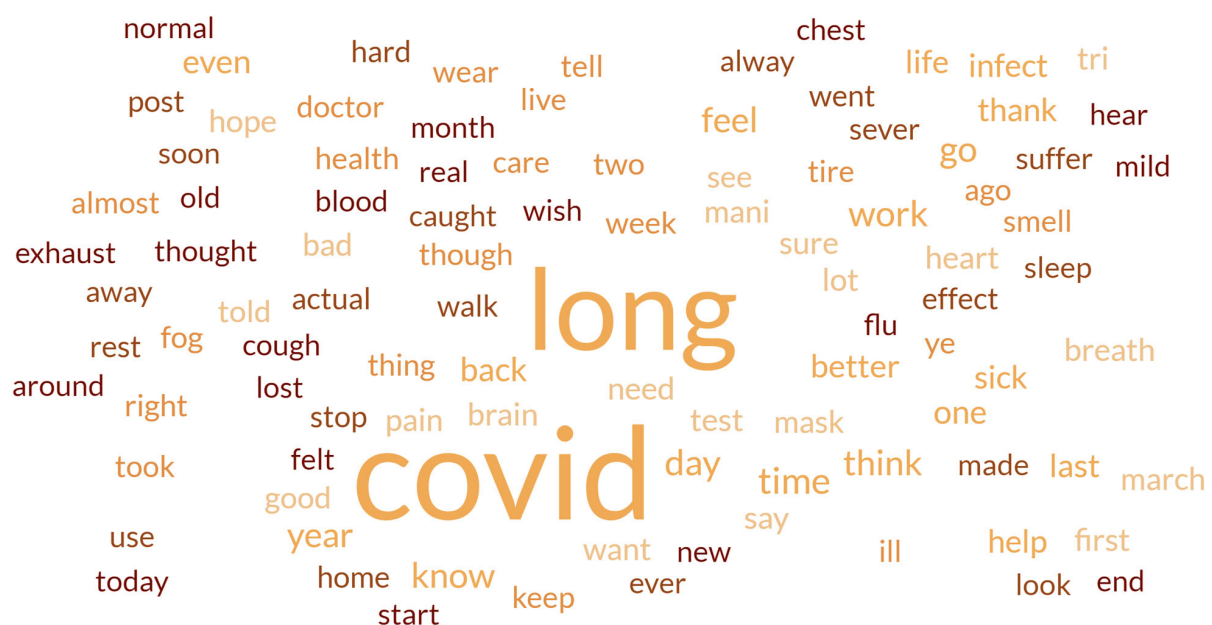


**Figure 12.** Representation of the number of positive Tweets of different intensities wherein individuals self-reported Long COVID.



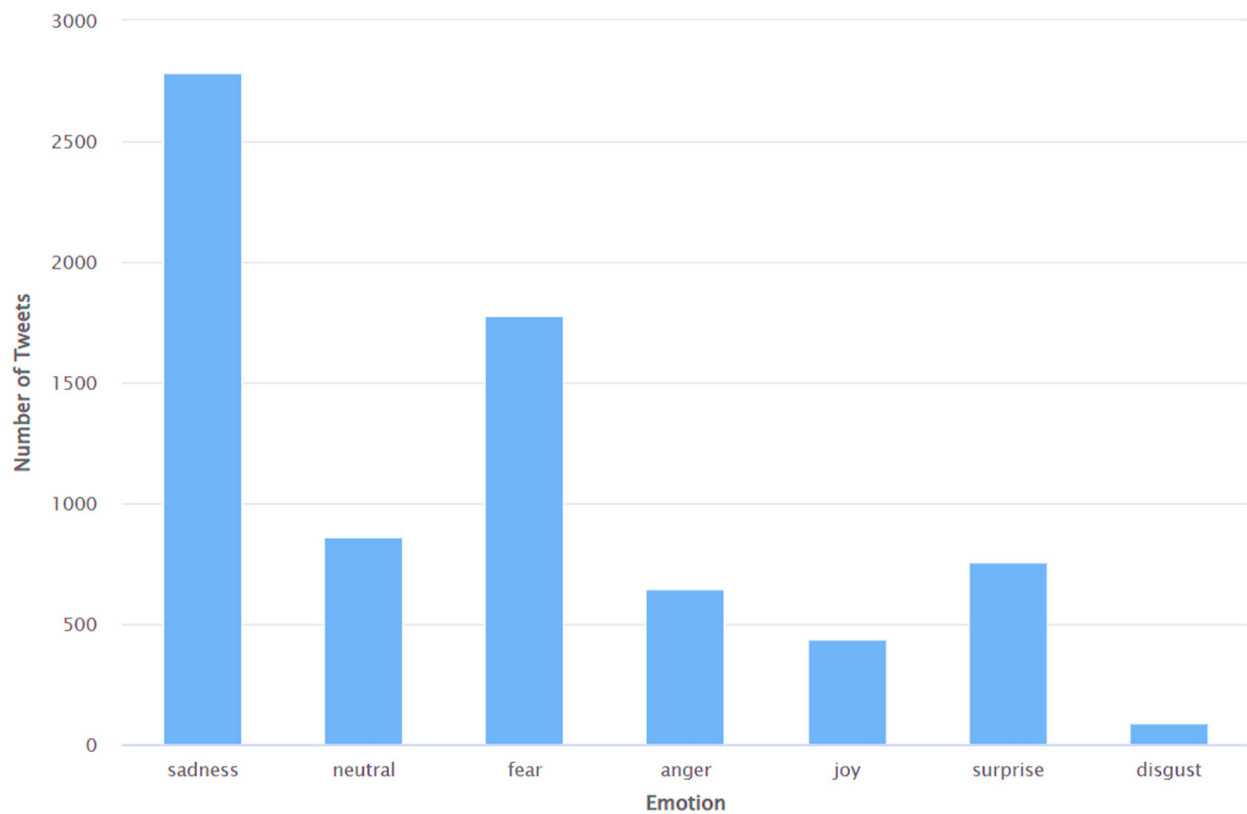
**Figure 13.** Representation of the number of negative Tweets of different intensities wherein individuals self-reported Long COVID.

Figure 14 uses a word cloud to illustrate the top 100 most frequent words in Tweets wherein people self-reported Long COVID. The word cloud analysis reveals that some of the frequently used words were “exhaust”, “breath”, “chest”, “cough”, “smell”, and “sleep”. This suggests that a significant proportion of individuals conveyed their symptoms at the time of self-reporting Long COVID on Twitter. Thereafter, a comprehensive analysis of these Tweets was also performed to identify the number of Tweets in different fine-grain sentiment classes to better understand the distribution of specific emotions expressed by Twitter users when posting positive or negative Tweets wherein they self-reported Long COVID. These fine-grain sentiment classes were sadness, neutral, fear, anger, joy, surprise, and disgust. The result of this analysis is shown in Figure 15. As can be seen from Figure 15, the emotion of sadness was expressed in most of these Tweets. It was followed by the emotions of fear, neutral, surprise, anger, joy, and disgust, respectively.



**Figure 14.** A word cloud-based representation of the top 100 words that featured in these Tweets wherein individuals self-reported Long COVID.





**Figure 15.** Representation of the results of analyses of Tweets to categorize the same into fine-grained sentiment classes.

The following discussion is provided to explain how the research conducted in this paper endeavors to overcome the drawbacks of previous studies in this domain, as described in Section 2.

- As explained in Section 2, a wide range of research challenges pertaining to COVID-19 have been explored and investigated via the analysis of relevant Tweets in scholarly works over the past few years. These include traveling [87], current trends [88], worries of the general public [88], the evaluation of events [89], opinions on mask-wearing [91], inquiries into influencer behaviors [92], detecting and tracking misinformation [93], studies of addiction trends [96], identifying loneliness [98], and the evaluation of impulse purchases [100]. Despite the extensive exploration of many research questions within this particular domain, the existing body of literature [87–89,91–93,96,98,100] has not specifically investigated the Twitter discourse pertaining to Long COVID. The research described in this study addresses this limitation found in previous studies in this field [87–89,91–93,96,98,100] by focusing on the analysis of Tweets related to Long COVID.
- Despite the existence of some studies, such as those conducted by Awoyemi et al. [108] and Pitroda et al. [109], which investigated Tweets pertaining to Long COVID, a significant limitation of these studies is the restricted temporal scope of the analyzed Tweets. In [108], the examined Tweets were published between 25 March 2022 and 1 April 2022. Similarly, the Tweets analyzed in [109] were published between 11 December 2021 and 20 December 2021. These durations constitute a small portion of the complete timeframe over which Long COVID has had its impact on the global population. This study addresses this limitation by conducting an analysis of Tweets pertaining to Long COVID, published between 25 May 2020 and 31 January 2023.
- The application of sentiment analysis to Tweets has proven valuable in discerning the range of views and opinions expressed by the global population on Twitter about vari-

ous subjects of discussion during previous instances of virus outbreaks. Consequently, there has been a notable surge in the volume of literature pertaining to sentiment analysis since the onset of the COVID-19 pandemic. Despite the existence of many studies (e.g., [86,87,94,101–104,106,107]) that have performed sentiment analyses of Tweets pertaining to COVID-19, none of these studies have specifically examined the sentiments expressed in Tweets related to Long COVID. This study addresses this limitation using the Valence Aware Dictionary for Sentiment Reasoning (VADER) methodology to perform sentiment analyses of Tweets about Long COVID.

- The examination of conversation patterns of people on Twitter who self-report health-related issues has received significant attention from researchers across a wide range of disciplines. This can be observed through the increasing number of studies that have focused on analyzing such Tweets in the context of mental health [110], autism [111], dementia [112], depression [113], breast cancer [114], swine flu [115], influenza [116], chronic stress [117], post-traumatic stress disorder [118], and dental issues [119]. Since the beginning of the COVID-19 pandemic, scholars in this field, as exemplified by the work of researchers in [99], have redirected their attention toward devising approaches for the acquisition and analysis of Tweets in which individuals willingly disclose how they contracted COVID-19, including self-reported instances of COVID-19. However, previous studies in this area of research did not specifically examine Tweets wherein people self-reported Long COVID. This study addresses this limitation by examining Tweets in which Twitter users self-reported Long COVID.

This research demonstrates that the discourse related to Long COVID on Twitter has garnered worldwide interest. The results of this study also indicate that the global population has been proactive in seeking and disseminating information about Long COVID on Twitter. The opinions conveyed by the global population within these online discussions are diverse in terms of the types of opinions as well as the intensity of the opinions expressed. Furthermore, the results suggest variations in the tweeting patterns of the global population while expressing their emotions in this context. As discussed in Section 1, there is presently a lack of pharmacological interventions that have shown efficacy in alleviating or ameliorating the symptoms associated with Long COVID in a randomized-controlled trial (RCT). At the same time, rehabilitation has exhibited effectiveness in managing specific cases of Long COVID. The findings obtained in this study have the potential to provide a foundation for the prompt investigation of the underlying Tweets (positive, negative, or both) by public health agencies. Such an investigation is expected to provide an understanding of the primary worries, needs, or interests of the global population in relation to the self-reporting of Long COVID. The results of this investigation might also assist public health agencies in comprehending the views expressed by the global population about current treatment options for Long COVID, irrespective of whether they have previously received those treatments or are considering doing so. Scholars from several disciplines have conducted studies to understand and interpret the paradigms of transmission of news from news sources to social media sites [130,131]. Therefore, public health agencies could utilize the framework outlined in this study to monitor and examine Twitter conversations regarding Long COVID during days characterized by notable news events, such as the availability of new treatments or medications for Long COVID, or the publication of studies regarding severe adverse reactions to newly suggested treatments or medications for Long COVID. Monitoring and examining the discourse on Twitter about Long COVID during such days is expected to assist with the immediate evaluation of the views of the target populations of those treatments or medications. An investigation of such views on Twitter might also be useful in aiding public health agencies in establishing rapid responses as well as applicable regulations regarding such treatments or medications. This paper also shows that a substantial volume of Tweets comprising views, opinions, concerns, and attitudes pertaining to Long COVID across a diverse range of topics have been published on Twitter within a short period of time. Public health agencies may consider this finding for applicable policymaking to analyze the historical and current discussions surrounding

Long COVID on Twitter, and to evaluate whether the ongoing public discourse on Twitter effectively addresses the information disparity between the public's requirements and the information provided by healthcare and medical sectors. Moreover, conducting such an inquiry might also play a crucial role in identifying misinformation or the dissemination of any conspiracy theories linked to Long COVID on Twitter.

## 5. Conclusions

In the last decade and a half, the world has had to cope with a number of epidemics of infectious diseases, all of which have resulted in a devastating effect on public health and the economy on a global scale. During the course of these epidemics, social media platforms evolved as useful outlets for information exchange, community development, and addressing loneliness, as well as for understanding the intricacies and dynamics of these diseases. Analyzing data from social media platforms for syndromic surveillance, with a specific focus on public health, has been of keen interest to scholars from a variety of disciplines. The persistence of symptoms of COVID-19 for a number of weeks or even months after the first infection with SARS-CoV-2 is referred to as a "Long COVID". Since the first case of COVID-19, there has been an enormous increase in the number of discussions on social media sites like Twitter that are related to Long COVID. Several recent research works in this area have focused on performing sentiment analyses of Tweets about COVID-19 in order to reveal the range of opinions and emotions of the global population. However, the majority of these studies have not focused on Long COVID, and the few studies that focused on Long COVID analyzed Tweets that were posted on Twitter over a brief time frame. In addition to this, none of the previous research carried out on this topic has focused on investigating Tweets in which people self-reported Long COVID.

This study aims to address these research challenges in this field by presenting multiple novel findings derived from a comprehensive analysis of a dataset consisting of 1,244,051 Tweets pertaining to Long COVID, published on Twitter over the period spanning from 25 May 2020 to 31 January 2023. First, the findings demonstrate that the average number of Tweets per month wherein individuals self-report Long COVID on Twitter was considerably high in 2022, as compared to the average number of such Tweets per month in 2021. Second, the findings obtained from the sentiment analysis conducted using the VADER approach show that 43.1% of the Tweets expressed a positive sentiment, 42.7% expressed a negative sentiment, and 14.2% expressed a neutral sentiment. The results obtained from the sentiment analysis also indicate that the majority of Tweets expressing a positive sentiment, as well as those expressing a negative sentiment, were not highly polarized. In other words, a majority of these Tweets did not convey either a strong positive sentiment (+4 on the VADER scale) or a strong negative sentiment (−4 on the VADER scale). Third, the findings obtained from tokenization reveal that there was an identical tweeting pattern in terms of the utilization of tokens in Tweets expressing positive and negative sentiments. However, Tweets with a neutral sentiment exhibited a different tweeting pattern when compared to Tweets expressing positive and negative sentiments. The examination of the findings from tokenization also reveal that the tweeting patterns of the global population in this context were diverse, and there was no correlation between the number of tokens utilized in the Tweets and the degree of sentiment expressed in the Tweets. Finally, an in-depth sentiment analysis revealed that the predominant emotion conveyed in the majority of these Tweets was sadness. It was followed by the emotions of fear, neutral, surprise, anger, joy, and disgust, respectively.

As stated in this paper, no pharmaceutical medication has shown efficacy in alleviating or ameliorating the symptoms often linked with Long COVID in a randomized-controlled trial (RCT) or similar. However, both rehabilitation and non-steroidal anti-inflammatory medicines (NSAIDs) have shown efficacy in managing certain symptoms associated with Long COVID. As ongoing research in this field progresses, it is anticipated that pharmaceutical medicines, specialized rehabilitation programs, drug therapy, and several other treatment modalities for Long COVID will be developed soon. Such progress in the field of

medicine and healthcare with a specific focus on Long COVID is likely to drive discourse on Twitter pertaining to Long COVID, specifically regarding the available and emerging forms of treatment. So, future research will include the topic modeling of Tweets in which people self-report Long COVID, with the aim of comprehending the distinct themes that are prevalent in such conversations on Twitter. Furthermore, sentiment analyses of the Tweets in each of these distinct themes can be performed to infer the global population's viewpoint about advancements pertaining to the treatment and management of Long COVID, as well as to understand their views towards other topics associated with Long COVID.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data analyzed in this study are publicly available at <https://www.kaggle.com/datasets/matt0922/twitter-long-covid-2023> (accessed on 15 August 2023).

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [CrossRef]
2. Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H.-R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R.A.M.; et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976. [CrossRef]
3. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1953–1966. [CrossRef] [PubMed]
4. Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814–1820. [CrossRef] [PubMed]
5. WHO Coronavirus (COVID-19) Dashboard. Available online: <https://covid19.who.int/> (accessed on 21 September 2023).
6. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273. [CrossRef] [PubMed]
7. Chan, J.F.-W.; Yuan, S.; Kok, K.-H.; To, K.K.-W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.-Y.; Poon, R.W.-S.; et al. A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating Person-to-Person Transmission: A Study of a Family Cluster. *Lancet* **2020**, *395*, 514–523. [CrossRef] [PubMed]
8. Zhou, Y.; Zeng, Y.; Tong, Y.; Chen, C. Ophthalmologic Evidence against the Interpersonal Transmission of 2019 Novel Coronavirus through Conjunctiva. *bioRxiv* **2020**. [CrossRef]
9. Cui, H.; Gao, Z.; Liu, M.; Lu, S.; Mo, S.; Mkandawire, W.; Narykov, O.; Srinivasan, S.; Korkin, D. Structural Genomics and Interactomics of 2019 Wuhan Novel Coronavirus, 2019-nCoV, Indicate Evolutionary Conserved Functional Regions of Viral Proteins. *bioRxiv* **2020**. [CrossRef]
10. Chen, L.; Liu, W.; Zhang, Q.; Xu, K.; Ye, G.; Wu, W.; Sun, Z.; Liu, F.; Wu, K.; Zhong, B.; et al. RNA Based MNGS Approach Identifies a Novel Human Coronavirus from Two Individual Pneumonia Cases in 2019 Wuhan Outbreak. *Emerg. Microbes Infect.* **2020**, *9*, 313–319. [CrossRef] [PubMed]
11. Ceraolo, C.; Giorgi, F.M. Genomic Variance of the 2019-nCoV Coronavirus. *J. Med. Virol.* **2020**, *92*, 522–528. [CrossRef] [PubMed]
12. Gallagher, T.M.; Buchmeier, M.J. Coronavirus Spike Proteins in Viral Entry and Pathogenesis. *Virology* **2001**, *279*, 371–374. [CrossRef]
13. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [CrossRef]
14. Luring, A.S.; Hodcroft, E.B. Genetic Variants of SARS-CoV-2—What Do They Mean? *JAMA* **2021**, *325*, 529. [CrossRef]
15. Khare, S.; GISAIID Global Data Science Initiative (GISAIID); Gurry, C.; Freitas, L.; Schultz, M.B.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; TC Lee, R.; et al. GISAIID's Role in Pandemic Response. *China CDC Wkly.* **2021**, *3*, 1049–1051. [CrossRef] [PubMed]
16. GISAIID—Gisaid.org. Available online: <https://www.gisaid.org/> (accessed on 3 September 2023).
17. Perego, E. The #LongCovid #COVID19. Available online: <https://twitter.com/elisaperego78/status/1263172084055838721?s=20> (accessed on 3 September 2023).
18. Crook, H.; Raza, S.; Nowell, J.; Young, M.; Edison, P. Long Covid—Mechanisms, Risk Factors, and Management. *BMJ* **2021**, *374*, n1648. [CrossRef] [PubMed]
19. Nabavi, N. Long Covid: How to Define It and How to Manage It. *BMJ* **2020**, *370*, m3489. [CrossRef]
20. Garg, P.; Arora, U.; Kumar, A.; Wig, N. The “Post-COVID” Syndrome: How Deep Is the Damage? *J. Med. Virol.* **2021**, *93*, 673–674. [CrossRef]



21. Greenhalgh, T.; Knight, M.; A’Court, C.; Buxton, M.; Husain, L. Management of Post-Acute COVID-19 in Primary Care. *BMJ* **2020**, *370*, m3026. [\[CrossRef\]](#)
22. Raveendran, A.V. Long COVID-19: Challenges in the Diagnosis and Proposed Diagnostic Criteria. *Diabetes Metab. Syndrome* **2021**, *15*, 145–146. [\[CrossRef\]](#)
23. Van Elslande, J.; Vermeersch, P.; Vandervoort, K.; Wawina-Bokalanga, T.; Vanmechelen, B.; Wollants, E.; Laenen, L.; André, E.; Van Ranst, M.; Lagrou, K.; et al. Symptomatic Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Reinfection by a Phylogenetically Distinct Strain. *Clin. Infect. Dis.* **2021**, *73*, 354–356. [\[CrossRef\]](#)
24. Thakur, N.; Han, C.Y. Pervasive Activity Logging for Indoor Localization in Smart Homes. In Proceedings of the 2021 4th International Conference on Data Science and Information Technology, Shanghai China, 23–25 July 2021; ACM: New York, NY, USA, 2021.
25. Thakur, N.; Han, C.Y. An Approach for Detection of Walking Related Falls during Activities of Daily Living. In Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 12–14 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 280–283.
26. Thakur, N.; Han, C.Y. A Framework for Prediction of Cramps during Activities of Daily Living in Elderly. In Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 12–14 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 284–287.
27. Fung, K.W.; Baye, F.; Baik, S.H.; Zheng, Z.; McDonald, C.J. Prevalence and Characteristics of Long COVID in Elderly Patients: An Observational Cohort Study of over 2 Million Adults in the US. *PLoS Med.* **2023**, *20*, e1004194. [\[CrossRef\]](#)
28. Falahi, S.; Kenarkoohi, A. COVID-19 Reinfection: Prolonged Shedding or True Reinfection? *New Microbes New Infect.* **2020**, *38*, 100812. [\[CrossRef\]](#)
29. Carfi, A.; Bernabei, R.; Landi, F. For the Gemelli Against COVID-19 Post-Acute Care Study Group Persistent Symptoms in Patients after Acute COVID-19. *JAMA* **2020**, *324*, 603. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Arnold, D.T.; Hamilton, F.W.; Milne, A.; Morley, A.J.; Viner, J.; Attwood, M.; Noel, A.; Gunning, S.; Hatrick, J.; Hamilton, S.; et al. Patient Outcomes after Hospitalisation with COVID-19 and Implications for Follow-up: Results from a Prospective UK Cohort. *Thorax* **2021**, *76*, 399–401. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Tenforde, M.W.; Kim, S.S.; Lindsell, C.J.; Billig Rose, E.; Shapiro, N.I.; Files, D.C.; Gibbs, K.W.; Erickson, H.L.; Steingrub, J.S.; Smithline, H.A.; et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health among Outpatients with COVID-19 in a Multistate Health Care Systems Network—United States, March–June 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2020**, *69*, 993–998. [\[CrossRef\]](#)
32. Lopez-Leon, S.; Wegman-Ostrosky, T.; Perelman, C.; Sepulveda, R.; Rebolledo, P.A.; Cuapio, A.; Villapol, S. More than 50 Long-Term Effects of COVID-19: A Systematic Review and Meta-Analysis. *Sci. Rep.* **2021**, *11*, 16144. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Fernández-de-las-Peñas, C. Long COVID: Current Definition. *Infection* **2022**, *50*, 285–286. [\[CrossRef\]](#)
34. Cutler, D.M. The Costs of Long COVID. *JAMA Health Forum* **2022**, *3*, e221809. [\[CrossRef\]](#)
35. Altmann, D.M.; Whettlock, E.M.; Liu, S.; Arachchillage, D.J.; Boyton, R.J. The Immunology of Long COVID. *Nat. Rev. Immunol.* **2023**, *23*, 618–634. [\[CrossRef\]](#)
36. Greenhalgh, T.; Knight, M. Long COVID: A Primer for Family Physicians. *Am. Fam. Physician* **2020**, *102*, 716–717. [\[PubMed\]](#)
37. Siddiq, M.A.B. Pulmonary Rehabilitation in COVID-19 Patients: A Scoping Review of Current Practice and Its Application during the Pandemic. *Turk. J. Phys. Med. Rehabil.* **2020**, *66*, 480–494. [\[CrossRef\]](#)
38. Liu, K.; Zhang, W.; Yang, Y.; Zhang, J.; Li, Y.; Chen, Y. Respiratory Rehabilitation in Elderly Patients with COVID-19: A Randomized Controlled Study. *Complement. Ther. Clin. Pract.* **2020**, *39*, 101166. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Yong, S.J. Long COVID or Post-COVID-19 Syndrome: Putative Pathophysiology, Risk Factors, and Treatments. *Infect. Dis.* **2021**, *53*, 737–754. [\[CrossRef\]](#)
40. Blitshteyn, S.; Whitelaw, S. Postural Orthostatic Tachycardia Syndrome (POTS) and Other Autonomic Disorders after COVID-19 Infection: A Case Series of 20 Patients. *Immunol. Res.* **2021**, *69*, 205–211. [\[CrossRef\]](#)
41. Johansson, M.; Ståhlberg, M.; Runold, M.; Nygren-Bonnier, M.; Nilsson, J.; Olshansky, B.; Bruchfeld, J.; Fedorowski, A. Long-Haul Post-COVID-19 Symptoms Presenting as a Variant of Postural Orthostatic Tachycardia Syndrome. *JACC Case Rep.* **2021**, *3*, 573–580. [\[CrossRef\]](#)
42. Kanjwal, K.; Jamal, S.; Kichloo, A.; Grubb, B. New-Onset Postural Orthostatic Tachycardia Syndrome Following Coronavirus Disease 2019 Infection. *J. Innov. Card. Rhythm Manag.* **2020**, *11*, 4302–4304. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Miglis, M.G.; Prieto, T.; Shaik, R.; Muppidi, S.; Sinn, D.-I.; Jaradeh, S. A Case Report of Postural Tachycardia Syndrome after COVID-19. *Clin. Auton. Res.* **2020**, *30*, 449–451. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Davis, H.E.; Assaf, G.S.; McCorkell, L.; Wei, H.; Low, R.J.; Re’em, Y.; Redfield, S.; Austin, J.P.; Akrami, A. Characterizing Long COVID in an International Cohort: 7 Months of Symptoms and Their Impact. *EClinicalMedicine* **2021**, *38*, 101019. [\[CrossRef\]](#)
45. Thakur, N. Social Media Mining and Analysis: A Brief Review of Recent Challenges. *Information* **2023**, *14*, 484. [\[CrossRef\]](#)
46. Injadat, M.; Salo, F.; Nassif, A.B. Data Mining Techniques in Social Media: A Survey. *Neurocomputing* **2016**, *214*, 654–670. [\[CrossRef\]](#)
47. Zubiaga, A. Mining Social Media for Newsgathering: A Review. *Online Soc. Netw. Media* **2019**, *13*, 100049. [\[CrossRef\]](#)
48. Thakur, N.; Han, C.Y. Country-Specific Interests towards Fall Detection from 2004–2021: An Open Access Dataset and Research Questions. *Data* **2021**, *6*, 92. [\[CrossRef\]](#)

49. Thakur, N.; Han, C.Y. Google Trends to Investigate the Degree of Global Interest Related to Indoor Location Detection. In *Human Interaction, Emerging Technologies and Future Systems V*; Springer International Publishing: Cham, Switzerland, 2022; pp. 580–588. ISBN 9783030855390.
50. Mayrhofer, M.; Matthes, J.; Einwiller, S.; Naderer, B. User Generated Content Presenting Brands on Social Media Increases Young Adults' Purchase Intention. *Int. J. Advert.* **2020**, *39*, 166–186. [[CrossRef](#)]
51. Roma, P.; Aloini, D. How Does Brand-Related User-Generated Content Differ across Social Media? Evidence Reloaded. *J. Bus. Res.* **2019**, *96*, 322–339. [[CrossRef](#)]
52. Charles-Smith, L.E.; Reynolds, T.L.; Cameron, M.A.; Conway, M.; Lau, E.H.Y.; Olsen, J.M.; Pavlin, J.A.; Shigematsu, M.; Streichert, L.C.; Suda, K.J.; et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. *PLoS ONE* **2015**, *10*, e0139701. [[CrossRef](#)] [[PubMed](#)]
53. Chew, C.; Eysenbach, G. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. *PLoS ONE* **2010**, *5*, e14118. [[CrossRef](#)] [[PubMed](#)]
54. Li, S.; Wang, Y.; Xue, J.; Zhao, N.; Zhu, T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2032. [[CrossRef](#)]
55. Thakur, N. Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox. *Big Data Cogn. Comput.* **2023**, *7*, 116. [[CrossRef](#)]
56. Bartlett, C.; Wurtz, R. Twitter and Public Health. *J. Public Health Manag. Pract.* **2015**, *21*, 375–383. [[CrossRef](#)]
57. Tomaszewski, T.; Morales, A.; Lourentzou, I.; Caskey, R.; Liu, B.; Schwartz, A.; Chin, J. Identifying False Human Papillomavirus (HPV) Vaccine Information and Corresponding Risk Perceptions from Twitter: Advanced Predictive Models. *J. Med. Internet Res.* **2021**, *23*, e30451. [[CrossRef](#)]
58. Lee, S.Y.; Khang, Y.-H.; Lim, H.-K. Impact of the 2015 Middle East Respiratory Syndrome Outbreak on Emergency Care Utilization and Mortality in South Korea. *Yonsei. Med. J.* **2019**, *60*, 796. [[CrossRef](#)]
59. Radzikowski, J.; Stefanidis, A.; Jacobsen, K.H.; Croitoru, A.; Crooks, A.; Delamater, P.L. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. *JMIR Public Health Surveill.* **2016**, *2*, e1. [[CrossRef](#)]
60. Fu, K.-W.; Liang, H.; Saroha, N.; Tse, Z.T.H.; Ip, P.; Fung, I.C.-H. How People React to Zika Virus Outbreaks on Twitter? A Computational Content Analysis. *Am. J. Infect. Control* **2016**, *44*, 1700–1702. [[CrossRef](#)]
61. Thakur, N. MonkeyPox2022Tweets: A Large-Scale Twitter Dataset on the 2022 Monkeypox Outbreak, Findings from Analysis of Tweets, and Open Research Questions. *Infect. Dis. Rep.* **2022**, *14*, 855–883. [[CrossRef](#)] [[PubMed](#)]
62. Kraaijeveld, O.; De Smedt, J. The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices. *J. Int. Financ. Mark. Inst. Money* **2020**, *65*, 101188. [[CrossRef](#)]
63. Aharon, D.Y.; Demir, E.; Lau, C.K.M.; Zaremba, A. Twitter-Based Uncertainty and Cryptocurrency Returns. *Res. Int. Bus. Fin.* **2022**, *59*, 101546. [[CrossRef](#)]
64. Alkouz, B.; Al Aghbari, Z.; Al-Garadi, M.A.; Sarker, A. Deepluenza: Deep Learning for Influenza Detection from Twitter. *Expert Syst. Appl.* **2022**, *198*, 116845. [[CrossRef](#)]
65. Oyeyemi, S.O.; Gabarron, E.; Wynn, R. Ebola, Twitter, and Misinformation: A Dangerous Combination? *BMJ* **2014**, *349*, g6178. [[CrossRef](#)]
66. Thakur, N.; Hall, I.; Han, C.Y. A Comprehensive Study to Analyze Trends in Web Search Interests Related to Fall Detection before and after COVID-19. In Proceedings of the 2022 5th International Conference on Computer Science and Software Engineering (CSSE 2022), Guilin, China, 21–23 October 2022; ACM: New York, NY, USA, 2022.
67. Thakur, N.; Duggal, Y.N.; Liu, Z. Analyzing Public Reactions, Perceptions, and Attitudes during the MPox Outbreak: Findings from Topic Modeling of Tweets. *Computers* **2023**, *12*, 191. [[CrossRef](#)]
68. Mouronte-López, M.L.; Ceres, J.S.; Columbrans, A.M. Analysing the Sentiments about the Education System Trough Twitter. *Educ. Inf. Technol.* **2023**, *28*, 10965–10994. [[CrossRef](#)]
69. Li, X.; Hasan, S.; Culotta, A. Identifying Hurricane Evacuation Intent on Twitter. *Proc. Int. AAAI Conf. Web Soc. Media* **2022**, *16*, 618–627. [[CrossRef](#)]
70. Lawelai, H.; Sadat, A.; Suherman, A. Democracy and Freedom of Opinion in Social Media: Sentiment Analysis on Twitter. *PRJ* **2022**, *10*, 40–48. [[CrossRef](#)]
71. Thakur, N.; Han, C.Y. A Framework for Facilitating Human-Human Interactions to Mitigate Loneliness in Elderly. In *Human Interaction, Emerging Technologies and Future Applications III*; Springer International Publishing: Cham, Switzerland, 2021; pp. 322–327. ISBN 9783030553067.
72. Thakur, N.; Han, C.Y. A Human-Human Interaction-Driven Framework to Address Societal Issues. In *Human Interaction, Emerging Technologies and Future Systems V*; Springer International Publishing: Cham, Switzerland, 2022; pp. 563–571. ISBN 9783030855390.
73. Grover, P.; Kar, A.K.; Janssen, M.; Ilavarasan, P.V. Perceived Usefulness, Ease of Use and User Acceptance of Blockchain Technology for Digital Transactions—Insights from User-Generated Content on Twitter. *Enterp. Inf. Syst.* **2019**, *13*, 771–800. [[CrossRef](#)]
74. Mnif, E.; Mouakhar, K.; Jarbou, A. Blockchain Technology Awareness on Social Media: Insights from Twitter Analytics. *J. High Technol. Manag. Res.* **2021**, *32*, 100416. [[CrossRef](#)]
75. Lu, X.; Brelsford, C. Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami. *Sci. Rep.* **2014**, *4*, 6773. [[CrossRef](#)]

76. Thakur, N.; Han, C.Y. An Intelligent Ubiquitous Activity Aware Framework for Smart Home. In *Human Interaction, Emerging Technologies and Future Applications III*; Springer International Publishing: Cham, Switzerland, 2021; pp. 296–302. ISBN 9783030553067.
77. Buccafurri, F.; Lax, G.; Nicolazzo, S.; Nocera, A. Comparing Twitter and Facebook User Behavior: Privacy and Other Aspects. *Comput. Human Behav.* **2015**, *52*, 87–95. [\[CrossRef\]](#)
78. Thakur, N.; Han, C.Y. Indoor Localization for Personalized Ambient Assisted Living of Multiple Users in Multi-Floor Smart Environments. *Big Data Cogn. Comput.* **2021**, *5*, 42. [\[CrossRef\]](#)
79. Golder, S.; Stevens, R.; O'Connor, K.; James, R.; Gonzalez-Hernandez, G. Methods to Establish Race or Ethnicity of Twitter Users: Scoping Review. *J. Med. Internet Res.* **2022**, *24*, e35788. [\[CrossRef\]](#)
80. Chang, R.-C.; Rao, A.; Zhong, Q.; Wojcieszak, M.; Lerman, K. #RoeOverturned: Twitter Dataset on the Abortion Rights Controversy. *Proc. Int. AAAI Conf. Web Soc. Media* **2023**, *17*, 997–1005. [\[CrossRef\]](#)
81. Bhatia, K.V. Hindu Nationalism Online: Twitter as Discourse and Interface. *Religions* **2022**, *13*, 739. [\[CrossRef\]](#)
82. Peña-Fernández, S.; Larrondo-Ureta, A.; Morales-i-Gras, J. Feminism, gender identity and polarization in TikTok and Twitter. *Comunicar* **2023**, *31*, 49–60. [\[CrossRef\]](#)
83. Goetz, S.J.; Heaton, C.; Imran, M.; Pan, Y.; Tian, Z.; Schmidt, C.; Qazi, U.; Ofli, F.; Mitra, P. Food Insufficiency and Twitter Emotions during a Pandemic. *Appl. Econ. Perspect. Policy* **2023**, *45*, 1189–1210. [\[CrossRef\]](#)
84. Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; et al. A Conceptual Model for the Coronavirus Disease 2019 (COVID-19) Outbreak in Wuhan, China with Individual Reaction and Governmental Action. *Int. J. Infect. Dis.* **2020**, *93*, 211–216. [\[CrossRef\]](#)
85. Öcal, A.; Cvetković, V.M.; Baytiyeh, H.; Tedim, F.M.S.; Zečević, M. Public Reactions to the Disaster COVID-19: A Comparative Study in Italy, Lebanon, Portugal, and Serbia. *Geomat. Nat. Hazards Risk* **2020**, *11*, 1864–1885. [\[CrossRef\]](#)
86. Mehedi Shamrat, F.M.J.; Chakraborty, S.; Imran, M.M.; Muna, J.N.; Billah, M.M.; Das, P.; Rahman, M.O. Sentiment Analysis on Twitter Tweets about COVID-19 Vaccines Using NLP and Supervised KNN Classification Algorithm. *Indones. J. Electr. Eng. Comput. Sci.* **2021**, *23*, 463. [\[CrossRef\]](#)
87. Sontayasara, T.; Jariyapongpaiboon, S.; Promjun, A.; Seelpipat, N.; Saengtabtim, K.; Tang, J.; Leelawat, N. Twitter Sentiment Analysis of Bangkok Tourism during COVID-19 Pandemic Using Support Vector Machine Algorithm. *J. Disaster Res.* **2021**, *16*, 24–30. [\[CrossRef\]](#)
88. Asgari-Chenaghlu, M.; Nikzad-Khaskhaki, N.; Minaee, S. Covid-Transformer: Detecting COVID-19 Trending Topics on Twitter Using Universal Sentence Encoder. *arXiv* **2020**, arXiv:2009.03947.
89. Amen, B.; Faiz, S.; Do, T.-T. Big Data Directed Acyclic Graph Model for Real-Time COVID-19 Twitter Stream Detection. *Pattern Recognit.* **2022**, *123*, 108404. [\[CrossRef\]](#)
90. Lyu, J.C.; Luli, G.K. Understanding the Public Discussion about the Centers for Disease Control and Prevention during the COVID-19 Pandemic Using Twitter Data: Text Mining Analysis Study. *J. Med. Internet Res.* **2021**, *23*, e25108. [\[CrossRef\]](#)
91. Al-Ramahi, M.; Elnoshokaty, A.; El-Gayar, O.; Nasrallah, T.; Wahbeh, A. Public Discourse against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data. *JMIR Public Health Surveill.* **2021**, *7*, e26780. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Jain, S.; Sinha, A. Identification of Influential Users on Twitter: A Novel Weighted Correlated Influence Measure for COVID-19. *Chaos Solitons Fractals* **2020**, *139*, 110037. [\[CrossRef\]](#)
93. Madani, Y.; Erritali, M.; Bouikhalene, B. Using Artificial Intelligence Techniques for Detecting Covid-19 Epidemic Fake News in Moroccan Tweets. *Results Phys.* **2021**, *25*, 104266. [\[CrossRef\]](#)
94. Shokoohyar, S.; Rikhtehgar Berenji, H.; Dang, J. Exploring the Heated Debate over Reopening for Economy or Continuing Lockdown for Public Health Safety Concerns about COVID-19 in Twitter. *Int. J. Bus. Syst. Res.* **2021**, *15*, 650. [\[CrossRef\]](#)
95. Chehal, D.; Gupta, P.; Gulati, P. COVID-19 Pandemic Lockdown: An Emotional Health Perspective of Indians on Twitter. *Int. J. Soc. Psychiatry* **2021**, *67*, 64–72. [\[CrossRef\]](#)
96. Glowacki, E.M.; Wilcox, G.B.; Glowacki, J.B. Identifying #addiction Concerns on Twitter during the Covid-19 Pandemic: A Text Mining Analysis. *Subst. Abuse.* **2021**, *42*, 39–46. [\[CrossRef\]](#)
97. Selman, L.E.; Chamberlain, C.; Sowden, R.; Chao, D.; Selman, D.; Taubert, M.; Braude, P. Sadness, Despair and Anger When a Patient Dies Alone from COVID-19: A Thematic Content Analysis of Twitter Data from Bereaved Family Members and Friends. *Palliat. Med.* **2021**, *35*, 1267–1276. [\[CrossRef\]](#)
98. Koh, J.X.; Liew, T.M. How Loneliness Is Talked about in Social Media during COVID-19 Pandemic: Text Mining of 4,492 Twitter Feeds. *J. Psychiatr. Res.* **2022**, *145*, 317–324. [\[CrossRef\]](#)
99. Mackey, T.; Purushothaman, V.; Li, J.; Shah, N.; Nali, M.; Bardier, C.; Liang, B.; Cai, M.; Cuomo, R. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated with COVID-19 on Twitter: Retrospective Big Data Intelligence Study. *JMIR Public Health Surveill.* **2020**, *6*, e19509. [\[CrossRef\]](#)
100. Leung, J.; Chung, J.Y.C.; Tisdale, C.; Chiu, V.; Lim, C.C.W.; Chan, G. Anxiety and Panic Buying Behaviour during COVID-19 Pandemic—A Qualitative Analysis of Toilet Paper Hoarding Contents on Twitter. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1127. [\[CrossRef\]](#)
101. Pokharel, B.P. Twitter Sentiment Analysis during Covid-19 Outbreak in Nepal. *SSRN Electron. J.* **2020**. [\[CrossRef\]](#)
102. Vijay, T.; Chawla, A.; Dhanka, B.; Karmakar, P. Sentiment Analysis on COVID-19 Twitter Data. In Proceedings of the 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 1–3 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7.



103. Shofiya, C.; Abidi, S. Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5993. [CrossRef]
104. Sahir, S.H.; Ayu Ramadhana, R.S.; Romadhon Marpaung, M.F.; Munthe, S.R.; Watrianthos, R. Online Learning Sentiment Analysis during the Covid-19 Indonesia Pandemic Using Twitter Data. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1156*, 012011. [CrossRef]
105. Thakur, N. A Large-Scale Dataset of Twitter Chatter about Online Learning during the Current COVID-19 Omicron Wave. *Data* **2022**, *7*, 109. [CrossRef]
106. Pristiyono; Ritonga, M.; Ihsan, M.A.A.; Anjar, A.; Rambe, F.H. Sentiment Analysis of COVID-19 Vaccine in Indonesia Using Naïve Bayes Algorithm. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1088*, 012045. [CrossRef]
107. Lohar, P.; Xie, G.; Bendeache, M.; Brennan, R.; Celeste, E.; Trestian, R.; Tal, I. Irish Attitudes toward COVID Tracker App & Privacy: Sentiment Analysis on Twitter and Survey Data. In Proceedings of the 16th International Conference on Availability, Reliability and Security, Vienna, Austria, 17–21 August 2021; ACM: New York, NY, USA, 2021.
108. Awoyemi, T.; Ebili, U.; Olusanya, A.; Ogunniyi, K.E.; Adejumo, A.V. Twitter Sentiment Analysis of Long COVID Syndrome. *Cureus* **2022**, *14*, e25901. [CrossRef] [PubMed]
109. Pitroda, H. Long Covid Sentiment Analysis of Twitter Posts to Understand Public Concerns. In Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 25–26 March 2022; IEEE: Piscataway, NJ, USA, 2022; Volume 1, pp. 140–148.
110. Coppersmith, G.; Dredze, M.; Harman, C.; Hollingshead, K. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses. Available online: <https://aclanthology.org/W15-1201.pdf> (accessed on 4 September 2023).
111. Hswen, Y.; Gopaluni, A.; Brownstein, J.S.; Hawkins, J.B. Using Twitter to Detect Psychological Characteristics of Self-Identified Persons with Autism Spectrum Disorder: A Feasibility Study. *JMIR MHealth UHealth* **2019**, *7*, e12264. [CrossRef] [PubMed]
112. Talbot, C.; O'Dwyer, S.; Clare, L.; Heaton, J.; Anderson, J. Identifying People with Dementia on Twitter. *Dementia* **2020**, *19*, 965–974. [CrossRef] [PubMed]
113. Almouzini, S.; Khemakhem, M.; Alageel, A. Detecting Arabic Depressed Users from Twitter Data. *Procedia Comput. Sci.* **2019**, *163*, 257–265. [CrossRef]
114. Clark, E.M.; James, T.; Jones, C.A.; Alapati, A.; Ukandu, P.; Danforth, C.M.; Dodds, P.S. A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions across Twitter. *arXiv* **2018**, arXiv:1805.09959. [CrossRef]
115. Szomszor, M.; Kostkova, P.; de Quincey, E. #swineflu: Twitter Predicts Swine Flu Outbreak in 2009. In *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 18–26. ISBN 9783642236341.
116. Alshammari, S.M.; Nielsen, R.D. Less Is More: With a 280-Character Limit, Twitter Provides a Valuable Source for Detecting Self-Reported Flu Cases. In Proceedings of the 2018 International Conference on Computing and Big Data, Tokyo, Japan, 25–28 November 2018; ACM: New York, NY, USA, 2018.
117. Yang, Y.-C.; Xie, A.; Kim, S.; Hair, J.; Al-Garadi, M.; Sarker, A. Automatic Detection of Twitter Users Who Express Chronic Stress Experiences via Supervised Machine Learning and Natural Language Processing. *Comput. Inform. Nurs.* **2022**; Publish Ahead of Print. [CrossRef]
118. Coppersmith, G.; Harman, C.; Dredze, M. Measuring Post Traumatic Stress Disorder in Twitter. *Proc. Int. AAAI Conf. Web Soc. Media* **2014**, *8*, 579–582. [CrossRef]
119. Al-Khalifa, K.S.; Bakhurji, E.; Halawany, H.S.; Alabdurubalnabi, E.M.; Nasser, W.W.; Shetty, A.C.; Sadaf, S. Pattern of Dental Needs and Advice on Twitter during the COVID-19 Pandemic in Saudi Arabia. *BMC Oral Health* **2021**, *21*, 456. [CrossRef]
120. Tsytsarau, M.; Palpanas, T. Survey on Mining Subjective Data on the Web. *Data Min. Knowl. Discov.* **2012**, *24*, 478–514. [CrossRef]
121. Saberi, B.; Saad, S. Sentiment Analysis or Opinion Mining: A Review. Available online: <https://core.ac.uk/download/pdf/296919524.pdf> (accessed on 4 September 2023).
122. Liu, B. *Sentiment Analysis and Opinion Mining*; Springer Nature: Cham, Switzerland, 2022; ISBN 9783031021459.
123. Medhat, W.; Hassan, A.; Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [CrossRef]
124. Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Available online: <https://aclanthology.org/H05-1044.pdf> (accessed on 4 September 2023).
125. Do, H.H.; Prasad, P.W.C.; Maag, A.; Alsadoon, A. Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Syst. Appl.* **2019**, *118*, 272–299. [CrossRef]
126. Nazir, A.; Rao, Y.; Wu, L.; Sun, L. Issues and Challenges of Aspect-Based Sentiment Analysis: A Comprehensive Survey. *IEEE Trans. Affect. Comput.* **2022**, *13*, 845–863. [CrossRef]
127. Hutto, C.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. AAAI Conf. Web Soc. Media* **2014**, *8*, 216–225. [CrossRef]
128. Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T. YALE: Rapid Prototyping for Complex Data Mining Tasks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; ACM: New York, NY, USA, 2006.
129. Fung, M. Twitter-Long COVID 2023. Available online: <https://www.kaggle.com/datasets/matt0922/twitter-long-covid-2023> (accessed on 15 August 2023).



130. Orellana-Rodriguez, C.; Keane, M.T. Attention to News and Its Dissemination on Twitter: A Survey. *Comput. Sci. Rev.* **2018**, *29*, 74–94. [[CrossRef](#)]
131. Bruns, A.; Burgess, J. Researching News Discussion on Twitter: New Methodologies. *J. Stud.* **2012**, *13*, 801–814. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.