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

Nirmalya Thakur [*]

Communication

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

Nirmalya Thakur *

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

* Correspondence: authors: nirmalya.thakur@emory.edu

Abstract: Since the outbreak of COVID-19, social media platforms, such as Twitter, have experienced a tremendous increase in conversations related to Long COVID. The term "Long COVID" describes the persistence of symptoms of COVID-19 for several weeks or even months following the initial infection. Recent works in this field have focused on sentiment analysis of Tweets related to COVID-19 to unveil the multifaceted spectrum of emotions, viewpoints, and perspectives held by the Twitter community. However, most of these works did not focus on Long COVID, and the few works that focused on Long COVID have limitations. Furthermore, no prior work in this field has investigated Tweets where individuals self-reported experiencing Long COVID on Twitter. The work presented in this paper aims to address these research challenges by presenting multiple novel findings from a comprehensive analysis of a dataset comprising 1,244,051 Tweets about Long COVID, posted on Twitter between May 25, 2020, and January 31, 2023. First, the analysis shows that the average number of Tweets per month where individuals self-reported Long COVID on Twitter, has been considerably high in 2022 as compared to the average number of Tweets per month in 2021. Second, findings of sentiment analysis using VADER show that the percentage of Tweets with positive, negative, and neutral sentiment were 43.12%, 42.65%, and 14.22%, respectively. Third, the analysis of sentiments associated with these Tweets also shows 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

1. Introduction

The global pandemic of coronavirus disease 2019 (COVID-19) has posed an immense and substantial menace to public health on a worldwide 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 strongly 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. 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]. At the time of writing of this paper, there have been a total of 694,672,356 cases and 6,911,814 deaths on account of 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 particle measures between 60 to 140 nanometers in diameter and boasts a positive-sense, single-stranded RNA genome spanning a length of 29891 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, reveals 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]. 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].

The SARS-CoV-2 virus primarily targets the respiratory system in humans, although, in certain instances, infections affecting other organs have also been documented. Initial studies conducted on cases originating in Wuhan, China, have provided insights into a spectrum of symptoms, including but not limited to fever, dry cough, shortness of breath, headache, vertigo, fatigue, vomiting, and diarrhea. It's important to note that not all individuals exhibit the same symptoms, and the nature and severity of these symptoms can vary significantly from person to person [13]. When a virus undergoes genetic alterations, it is described as having undergone mutations. These genetic modifications result in the emergence of virus genomes that differ from one another in terms of their genetic makeup, and these distinct genetic versions are referred to as variants. Variants that display variations in their observable characteristics are designated as strains [14]. On January 10, 2020, the Global Influenza Surveillance and Response System (GISAID) took a pivotal step by making the genomic sequences of SARS-CoV-2 publicly accessible. Subsequently, the GISAID database has made available over 5 million genetic sequences of SARS-CoV-2 collected from 194 countries and territories for research purposes. This extensive database has become an invaluable resource for understanding the virus's genetic diversity and its implications for public health on a global scale [15,16].

1.2. Concept of "Long COVID"

The term "Long COVID" originally surfaced in social media to describe the persistence of symptoms for several weeks or even months following the initial SARS-CoV-2 infection [17,18]. "Long COVID" serves as a comprehensive label encompassing a range of symptoms that persist, regardless of one's viral status, long after acquiring the SARS-CoV-2 infection. It is also commonly referred to as "post-COVID syndrome" [19]. The manifestation of these symptoms can take on various patterns, either persisting continuously or recurring intermittently [20]. These symptoms may encompass the persistence of one or more of the acute COVID symptoms or the emergence of new ones. Notably, a significant portion of individuals experiencing post-COVID syndrome test negative for PCR, indicating microbiological recovery. In simpler terms, post-COVID syndrome represents the time gap between microbiological recovery and clinical recuperation [20,21]. In many cases, individuals with Long COVID demonstrate signs of biochemical and radiological improvement. Depending on the duration of these persistent symptoms, post-COVID or Long COVID can be categorized into two stages: "post-acute COVID," where symptoms extend beyond three weeks but less than 12 weeks, and "chronic COVID," where symptoms persist for more than 12 weeks. Consequently, within the population infected with SARS-CoV-2, the presence of one or more symptoms, whether they persist continuously or reappear intermittently, and whether they are new or reminiscent of acute COVID symptoms beyond the expected period of clinical recovery, defines the condition known as post-COVID syndrome or Long COVID [22-24].

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 encompassed fatigue (53.1%), a decline in overall quality of life (44.1%), difficulty breathing (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 to perform daily activities. Moreover, patients 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 [24]. 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 [23,25]. According to a recent meta-analysis, the five most frequently observed symptoms of Long COVID-19 include fatigue (58%), headaches (44%), attention-related issues (27%), hair loss (25%), and difficulty breathing (24%) [28].

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 emerge as invaluable sources of information, capable of offering deep insights into the characteristics and status of such outbreaks [29]. Leveraging the power of text mining, health-related data can be gleaned from platforms like Twitter [30]. The comprehensive information contained within Twitter data grants researchers access to substantial user-generated content, facilitating early response strategies and informed decision-making. The realm of social media data text mining assumes a pivotal role in monitoring diseases and gauging public awareness of health concerns, thereby enabling proactive disease forecasting [31]. It's worth emphasizing that text analysis of Twitter data has become a focal point in the realm of medical informatics research [32].

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 [33]. 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. Of these social media platforms, Twitter stands out as a prominent communication channel during disease outbreaks [34]. 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 infectious disease outbreaks [35-39], societal problems [40-44], and humanitarian issues [45-49], as can be seen from several prior works in these fields which focused on sentiment analysis and other forms of content analysis of Tweets. In the wake of the COVID-19 outbreak, a burgeoning body of research has harnessed Twitter data to unravel public responses and discourse surrounding this global health crisis. However, it's worth noting that the literature concerning Long COVID remains very limited, beset by various constraints and challenges. Furthermore, no prior work in this field has focused on the analysis of Tweets where Twitter users self-reported Long COVID. To enrich the literature related to the understanding and interpretation of public views and reactions towards Long COVID and to address the research gaps in this field, this study presents several novel findings from a comprehensive analysis of a dataset comprising 1,244,051 Tweets about Long COVID, posted on Twitter between May 25, 2020, and January 31, 2023. The rest of this paper is organized as follows. A review of recent works and the existing challenges are discussed in Section 2. Section 3 presents the step-by-step methodology that was followed for this research project. The novel findings of this work are presented and discussed in Section 4. Section 5 concludes the paper by summarizing its scientific contributions and outlining the plans for future work in this area.

2. Literature Review

In this section, a review of recent research endeavors within this domain, specifically focusing on the analysis of social media conversations about COVID-19 and Long COVID, with Twitter as the central platform of interest, is presented. Shamrat et al. [50] applied the k-nearest neighbors (kNN) algorithm to categorize COVID-19-related tweets into three distinct classes—' positive,' 'negative,' and 'neutral'. The first step of their study involved filtering relevant tweets. Following this initial filtration process, the kNN algorithm was applied for an in-depth analysis of these tweets and their associated sentiments. Sontayasara et al. [51] utilized the capabilities of the Support Vector Machine (SVM) classifier to develop an algorithm tailored for sentiment analysis. Their work involved analysis of tweets wherein individuals articulated their intentions to visit Bangkok during the COVID-19 pandemic. By applying this SVM-based classifier, they successfully segregated the tweets into three distinct sentiment categories—'positive,' 'negative,' and 'neutral'. Asgari-Chenaghlu et al. [52] uncovered and analyzed the trending topics and the primary concerns related to the COVID-19 pandemic, as conveyed on Twitter. Their approach was characterized by detecting the pulse of Twitter discourse, pinpointing emergent themes, and unraveling the overarching concerns articulated by users. Amen et al. [53] proposed a framework that employed a directed acyclic graph model to analyze COVID-19-related tweets. This framework was specifically designed to detect anomalies in the context of events, as expressed on Twitter.

Lyu et al.'s work [54] centered around the development of an approach tailored to the examination of COVID-19-related tweets intricately intertwined with the guidelines set forth by the Centers for Disease Control and Prevention (CDC). Their overarching goal was to uncover a spectrum of public perceptions, encompassing concerns, attention, expectations, and more, all orbiting around the CDC's directives concerning COVID-19. The work of Al-Ramahi et al. [55] involved filtering and scrutinizing tweets about COVID-19 spanning the period from January 1, 2020, to October 27, 2020. In their work, they analyzed tweets wherein individuals vocalized their divergent viewpoints on the efficacy of mask-wearing as a strategy to curb the spread of COVID-19. Jain et al. [56] proposed a methodology engineered to analyze COVID-19-related tweets, with the aim of bestowing an influence score upon the users who posted these tweets. The crux of their study lay in the identification of Twitter's influential figures who actively engaged with the COVID-19 discourse, thereby shaping the narrative and impacting public opinion. Meanwhile, Madani et al. [57] delved into the realm of machine learning, crafting a robust classifier that used random forest-based approaches to analyze tweets. This classifier's primary functionality was to sift through tweets pertaining to COVID-19 and flag those that propagated fake news or misinformation. The classifier achieved an accuracy of 79%, underscoring its effectiveness in combatting the spread of misinformation during the pandemic.

Shokoohyar et al. [58] developed a comprehensive system designed to delve into the expansive realm of Twitter discourse, where individuals voiced their opinions concerning the COVID-19-induced lockdown measures in the United States. Their aim was to gain profound insights into the multifaceted array of sentiments, perspectives, and reactions related to this pivotal aspect of the pandemic response. Chehal et al. [59] developed specialized software 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. [60] developed a systematic approach geared towards the identification and examination of COVID-19-related tweets, where Twitter's 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 [61] centered on studying tweets where individuals shared heart-wrenching accounts of their loved ones succumbing to COVID-19. Their particular focus honed in on those instances where patients faced their final moments in isolation. This research illuminated the profoundly emotional and isolating aspects of the pandemic's impact. Koh et al. [62] 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. [63] zeroed in on tweets where individuals voluntarily reported their symptoms, experiences with testing sites, and recovery status in relation to COVID-19. Their work served as means for understanding the public's firsthand experiences and interactions with healthcare systems. In [64], Leung et al. aimed to decode 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 angle shed light on the peculiar buying trends and emotional undercurrents that manifested during the pandemic.

The broad focus of the research conducted by Pokharel [65] was to perform sentiment analysis 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 May 21, 2020, and May 31, 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 where emotions of fear, sadness, and even disgust surfaced. Vijay et al. [66] 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. [67] involved targeting Twitter data that was uniquely associated with Canada and centered around text containing 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 had a neutral emotion, 35% of the tweets had a negative emotion, and 25% of the tweets expressed a positive emotion. The primary objective of the investigation conducted by Sahir et al. [68] 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 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 a mere 1% of the tweets had a neutral sentiment. In a similar work [69], a dataset of Tweets about online learning during COVID-19 was developed.

The study conducted by Pristiyono et al. [70] aimed to delve into the perspectives and viewpoints held by the Indonesian population concerning the COVID-19 vaccine in the specific timeframe of January 2021. This process involved the utilization of RapidMiner for data collection. The results of the sentiment analysis showed that the percentage of negative tweets, positive tweets, and neutral tweets were 56%, 39%, and 1%, respectively. In a similar study [71], the researchers analyzed a corpus of tweets pertaining to COVID-19 from Ireland. This dataset spanned an entire year, commencing on January 1, 2020, and concluding on December 31, 2020. The primary objective of this study was to conduct sentiment analysis. The findings showed that a significant portion of the sentiments was 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. [72] used the Twarc2 tool to extract a corpus of 62,232 Tweets about Long COVID posted between March 25, 2022, and April 1, 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. [73] 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 December 11, 2021, and December 20, 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 works have the following limitations:

- Lack of focus on Long COVID: As discussed in this section, a wide range of research questions in the context of COVID-19, such as – tourism [51], trending topics [52], concerns [52], event analysis [53], views towards wearing masks [55], analysis of behaviors of influencers [56], misinformation detection [57], analysis of addiction [60], detection of loneliness [62], and panic buying [64], have been explored and investigated in the last few months by studying and analyzing relevant Tweets. However, none of these works [51-53, 55-57, 60, 62, 64] focused on the analysis of Tweets about Long COVID.

- Limitations in the few works that exist on Long COVID: While there have been a few works, such as [72,73], that focused on the analysis of Tweets about Long COVID, the primary limitation of such works is the limited time range (for example: March 25, 2022, to April 1, 2022 in [72] and December 11, 2021, to December 20, 2021 in [73]) of the Tweets that were analyzed. Such limited time ranges represented only a small fraction of the total time Long COVID has impacted the global population.

- The works in this field (for example: [50], [51], [58], [65-68,70,71]) that involved sentiment analysis of Tweets about COVID-19, did not investigate sentiments associated with tweets about Long COVID.

- Lack of focus on self-reporting of Long COVID: Analysis of self-reporting of symptoms related to various health conditions on Twitter has attracted the attention of researchers from different disciplines, as is evident from the recent works related to Twitter data analysis that focused on self-diagnosis or self-reporting of mental health problems [74], autism [75], dementia [76], depression [77], breast cancer [78], swine flu [79], flu [80], chronic stress [81], post-traumatic stress disorder [82], and dental issues [83], just to name a few. Since the outbreak of COVID-19, works in this field (such as [63]) have focused on developing approaches to collect and analyze tweets where people self-reported symptoms of COVID-19. However, none of the prior works in this field have focused on investigating tweets comprising self-reporting of Long COVID.

The work of this paper aims to address these limitations by analyzing a corpus of Tweets where people self-reported symptoms of Long COVID. The step-by-step methodology that was followed for this work is described in Section 3.

3. Methodology

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 [84]. 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 [85,86]. 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 [87]. It's worth noting that Wilson et al. [88] 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 [86]. 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.

The problem of Sentiment Analysis can be investigated using diverse approaches such as manual classification, Linguistic Inquiry and Word Count (LIWC), Affective Norms for English Words (ANEW), the General Inquirer (GI), SentiWordNet, and machine learning-oriented techniques-based algorithms like Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM). However, the specific methodology employed in this study was VADER, an acronym for Valence Aware Dictionary for sEntiment Reasoning. The rationale behind choosing VADER stems from its demonstrated prowess, surpassing manual classification in performance, and its ability to surmount the limitations encountered by comparable sentiment analysis techniques. The following represents an overview of the distinct features and properties of VADER [91]:

a) VADER sets itself apart from LIWC by exhibiting heightened sensitivity to sentiment expressions thriving in the context of analysis of social media texts.

b) The General Inquirer falls short in its coverage of sentiment-relevant lexical features that are commonplace in social text, a limitation that VADER adeptly addresses.

c) The ANEW lexicon proves less responsive to the sentiment-relevant lexical features typically encountered in social text. This isn't a limitation of VADER.

d) The SentiWordNet lexicon exhibits substantial noise, with a significant portion of its synsets lacking positive or negative polarity. This isn't a limitation of VADER.

e) The Naïve Bayes classifier hinges on a simplistic assumption of feature independence, a limitation circumvented by VADER's more nuanced approach.

f) The Maximum Entropy approach, devoid of the conditional independence assumption between features, factors in information entropy through feature weightings.

g) Machine learning classifiers, though robust, often demand extensive training data—a hurdle also faced by validated sentiment lexicons. Furthermore, machine learning classifiers rely on the training set to represent an extensive array of features.

VADER stands out by employing a sparse rule-based model, crafting a computational sentiment analysis engine tailored for social media-style text. It excels in its adaptability across various domains and requires no dedicated training data, relying instead on a versatile, valence-based, human-curated gold standard sentiment lexicon. VADER boasts efficiency, accommodating real-time analysis of streaming data, all while maintaining its computational performance, the time complexity of which is O(N). To add to this, VADER is accessible to all, with no subscription or purchase costs associated.

Beyond polarity detection (positive, negative, and neutral), VADER possesses the capability to gauge the intensity of sentiment expressed in texts.

To develop the system architecture for sentiment analysis of this research work, RapidMiner was used [92]. RapidMiner, originally known as Yet Another Learning Environment (YALE) [130], stands as a versatile data science platform empowering the conception, deployment, and application of an extensive array of algorithms and models within the realms of Machine Learning, Data Science, Artificial Intelligence, and Big Data. Its utility spans the domains of both scholarly research and the creation of pragmatic business-oriented applications and solutions. The comprehensive RapidMiner ecosystem encompasses a suite of integrated development environments, including (1) RapidMiner Studio, (2) RapidMiner Auto Model, (3) RapidMiner Turbo Prep, (4) RapidMiner Go, (5) RapidMiner Server, and (6) RapidMiner Radoop. It is worth mentioning here that in this paper, the specific reference to "RapidMiner" pertains exclusively to "RapidMiner Studio" and not to any of the alternative development environments affiliated with this software tool. RapidMiner Studio is designed as an open-core model, comprising a robust Graphical User Interface (GUI) that empowers developers to craft an array of applications and workflows while facilitating algorithm development and implementation. Within the RapidMiner development environment, specific operations or functions are denoted as "operators," and a coherent sequence of these "operators" (configured linearly, hierarchically, or through a blend of both approaches) directed towards a particular task or objective is termed a "process." RapidMiner streamlines the creation of these "processes" by offering a diverse selection of pre-built "operators" that can be directly employed. Furthermore, a specific class of "operators" is available to modify the defining characteristics of other "operators," adding a layer of flexibility and customization to the analytical process.

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 May 25, 2020, and January 31, 2023 [93]. 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 where individuals self-reported Long COVID on Twitter, 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 [74], autism [75], dementia [76], depression [77], breast cancer [78], swine flu [79], flu [80], chronic stress [81], post-traumatic stress disorder [82], and dental issues [83], 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". These Bag of Words were applied through a data filter to the Tweets to obtain a corpus of 7,348 Tweets where each Tweet involved self-reporting Long COVID. Thereafter, data preprocessing of these Tweets was performed. The data processing comprised the following steps:

a)   Removal of characters that are not alphabets.

b)   Removal of URLs.

c)   Removal of hashtags.

d)   Removal of user mentions.

e)   Detection of English words using tokenization.

f)   Stemming and Lemmatization.

g)   Removal of stop words

h)   Removal of numbers

i)   Replacing missing values.

Multiple "operators" were developed in RapidMiner to perform the data preprocessing. Thereafter, the Extract Sentiment operator in RapidMiner was customized to implement the VADER sentiment analysis approach on this corpus of 7,348 Tweets. The VADER methodology computed a compound sentiment score in the range of +4 to -4 for each of these Tweets. A rule-based approach was developed to identify the number of positive, negative, and neutral Tweets. This rule-based method functioned by analyzing the compound sentiment score of each Tweet. If this score was greater than 0, that Tweet was classified as a positive Tweet; if this score was less than 0, that Tweet was classified as a negative Tweet. Finally, if this score was 0, that Tweet was classified as a neutral Tweet. As the data processing and this rule-based approach for classification of the sentiment of a Tweet involved multiple "operators" working in a sequential manner, these were developed as "sub-processes" in the main "process" that was developed in RapidMiner. Figure 1 shows this main "process" that was developed in RapidMiner to perform the sentiment analysis of these Tweets. In this "process", "Dataset – F1" and "Dataset – F2" represent the two different files of this dataset [93]. This "process" consists of two "sub-processes" for data preprocessing and rule-based detection of sentiment classes, which are shown in Figure 2a,b, respectively. The results of running this "process" on all the Tweets in the dataset are presented and discussed in Section 4.
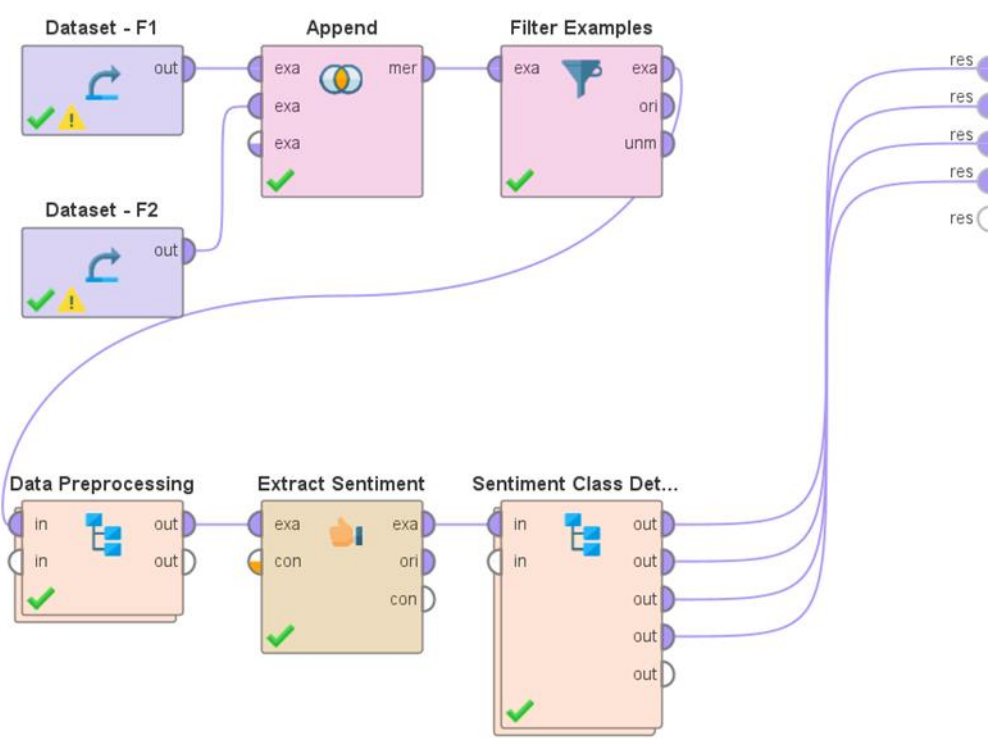

Figure 1. The main "process" that was developed in RapidMiner to perform the sentiment analysis of Tweets (posted on Twitter between May 25, 2020, and January 31, 2023) where individuals self-reported long COVID.
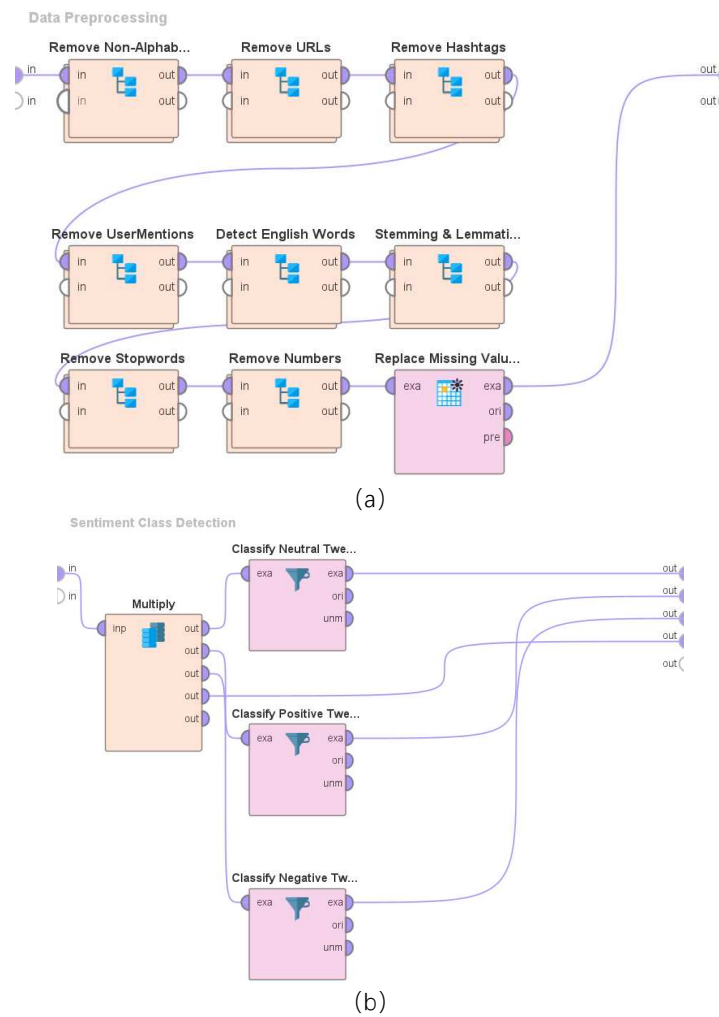
(a)

(b)

Figure 2. (**a**). Representation of the "operators" which comprised the data preprocessing "sub-process" shown in Figure 1. (**b**). Representation of the "operators" which comprised the sentiment class detection "sub-process" shown in Figure 1.

4. Results and Discussions

This section presents the results of performing sentiment analysis of Tweets (posted on Twitter between May 25, 2020, and January 31, 2023) where 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 self-reporting of Long COVID. Figure 3 shows a monthly variation of these Tweets during this time range, i.e., between May 25, 2020, and January 31, 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 where individuals self-reported Long COVID on Twitter, has been considerably high in 2022 as compared to the average number of Tweets per month in 2021.
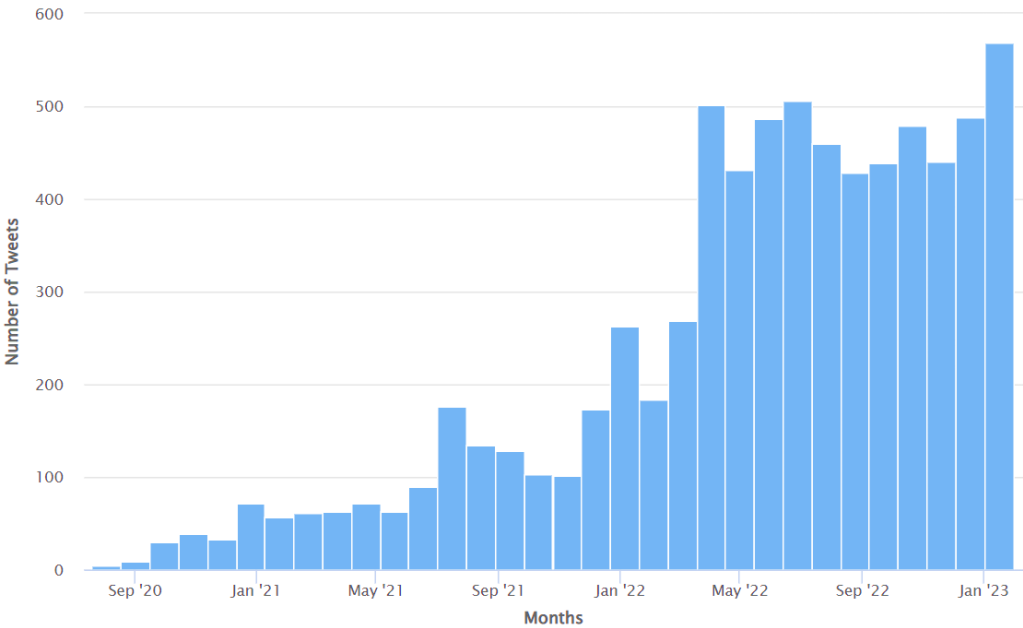
Figure 3. Representation of the monthly variation of Tweets (posted on Twitter between May 25, 2020, and January 31, 2023) where individuals self-reported long COVID.

Figure 4 shows the results of the RapidMiner "process" shown in Figure 1. As can be seen from Figure 4, the percentage of Tweets with positive, negative, and neutral sentiment were 43.12%, 42.65%, and 14.22%, respectively.
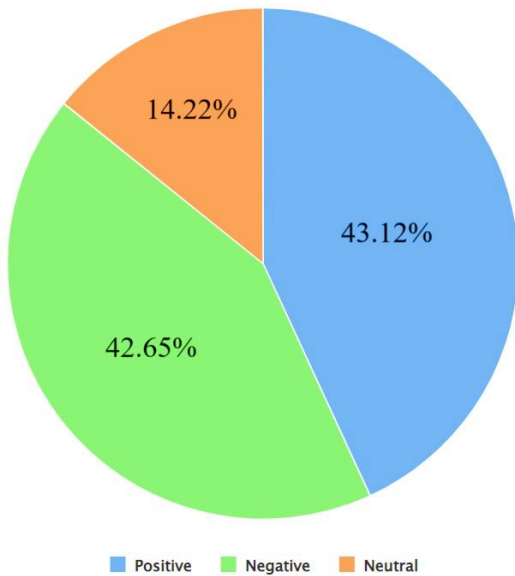

Figure 4. Representation of the results of the RapidMiner "process" shown in Figure 1.

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, where 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 5. As can be seen from Figure 5, 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.
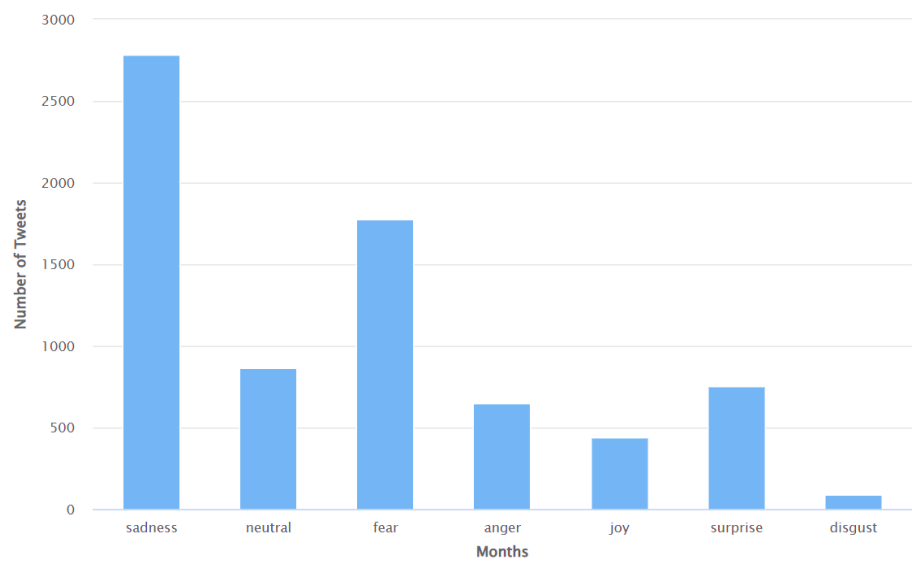
Figure 5. Representation of the results of analysis of Tweets to categorize the same into fine-grain sentiment classes.

5. Conclusions

In the last decade and a half, humanity has confronted several infectious disease outbreaks, each of which had a toll on both lives and the global economy. During those outbreaks, social media platforms emerged as invaluable resources of information, capable of providing profound insights into the nuances and dynamics of these outbreaks. Leveraging social media data for syndromic surveillance, with a dedicated emphasis on public health matters disseminated through web-based content, has been of paramount significance to researchers across different disciplines. The term "Long COVID" describes the persistence of symptoms of COVID-19 for several weeks or even months following the initial SARS-CoV-2 infection. Since the outbreak of COVID-19, social media platforms, such as Twitter, have experienced a tremendous increase in conversations related to Long COVID. Recent works in this field have focused on sentiment analysis of Tweets related to COVID-19 to unveil the multifaceted spectrum of emotions and viewpoints held by the Twitter community. However, most of these works did not focus on Long COVID, and the few works that focused on Long COVID analyzed Tweets that were posted on Twitter during a very limited time range. Furthermore, no prior work in this field has investigated Tweets where individuals self-reported Long COVID. The work presented in this paper aims to address these research challenges by presenting multiple novel findings from a comprehensive analysis of a dataset comprising 1,244,051 Tweets about Long COVID, posted on Twitter between May 25, 2020, and January 31, 2023. First, the analysis shows that the average number of Tweets per month where individuals self-reported Long COVID on Twitter, has been considerably high in 2022 as compared to the average number of Tweets per month in 2021. Second, findings of sentiment analysis using VADER show that the percentage of Tweets with positive, negative, and neutral sentiment were 43.12%, 42.65%, and 14.22%, respectively. Third, the analysis of sentiments associated with the Tweets also shows 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. Future work in this area would involve performing topic modeling of these Tweets to understand the specific topics that were represented in these discussions on Twitter in the context of long COVID.

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, doi:10.1056/nejmoa2001017.

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, doi:10.1056/nejmoa030747.

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, doi:10.1056/nejmoa030781.

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, doi:10.1056/nejmoa1211721.

5. COVID - Coronavirus Statistics - Worldometer Available online: https://www.worldometers.info/coronavirus/ (accessed on 3 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, doi:10.1038/s41586-020-2012-7.

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, doi:10.1016/s0140-6736(20)30154-9.

8. Zhou, Y.; Zeng, Y.; Tong, Y.; Chen, C. Ophthalmologic Evidence against the Interpersonal Transmission of 2019 Novel Coronavirus through Conjunctiva. bioRxiv 2020.

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.

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, doi:10.1080/22221751.2020.1725399.

11. Ceraolo, C.; Giorgi, F.M. Genomic Variance of the 2019-nCoV Coronavirus. J. Med. Virol. 2020, 92, 522–528, doi:10.1002/jmv.25700.

12. Gallagher, T.M.; Buchmeier, M.J. Coronavirus Spike Proteins in Viral Entry and Pathogenesis. Virology 2001, 279, 371–374, doi:10.1006/viro.2000.0757.

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, doi:10.1016/s0140-6736(20)30183-5.

14. Lauring, A.S.; Hodcroft, E.B. Genetic Variants of SARS-CoV-2—What Do They Mean? JAMA 2021, 325, 529, doi:10.1001/jama.2020.27124.

15. Khare, S.; GISAID Global Data Science Initiative (GISAID), Munich, Germany; Gurry, C.; Freitas, L.; B Schultz, M.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; TC Lee, R.; et al. GISAID's Role in Pandemic Response. China CDC Wkly. 2021, 3, 1049–1051, doi:10.46234/ccdcw2021.255.

16. GISAID - 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, doi:10.1136/bmj.n1648.

19. Nabavi, N. Long Covid: How to Define It and How to Manage It. BMJ 2020, m3489, doi:10.1136/bmj.m3489.

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, doi:10.1002/jmv.26465.

21. Greenhalgh, T.; Knight, M.; A'Court, C.; Buxton, M.; Husain, L. Management of Post-Acute Covid-19 in Primary Care. BMJ 2020, m3026, doi:10.1136/bmj.m3026.

22. Raveendran, A.V. Long COVID-19: Challenges in the Diagnosis and Proposed Diagnostic Criteria. Diabetes Metab. Syndr. 2021, 15, 145–146, doi:10.1016/j.dsx.2020.12.025.

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, doi:10.1093/cid/ciaa1330.

24. Falahi, S.; Kenarkoohi, A. COVID-19 Reinfection: Prolonged Shedding or True Reinfection? New Microbes New Infect. 2020, 38, 100812, doi:10.1016/j.nmni.2020.100812.

14

25. 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, doi:10.1001/jama.2020.12603.

26. 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, doi:10.1136/thoraxjnl-2020-216086.

27. 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, doi:10.15585/mmwr.mm6930e1.

28. 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, 1–12, doi:10.1038/s41598-021-95565-8.

29. Ginsberg, J.; Mohebbi, M.H.; Patel, R.S.; Brammer, L.; Smolinski, M.S.; Brilliant, L. Detecting Influenza Epidemics Using Search Engine Query Data. Nature 2009, 457, 1012–1014, doi:10.1038/nature07634.

30. Wolters Kluwer Health Available online: https://www.apjtm.org/text.asp?2020/13/8/378/279651 (accessed on 3 September 2023).

31. 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, doi:10.1371/journal.pone.0139701.

32. Chew, C.; Eysenbach, G. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. PLoS One 2010, 5, e14118, doi:10.1371/journal.pone.0014118.

33. 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, doi:10.3390/ijerph17062032.

34. Bartlett, C.; Wurtz, R. Twitter and Public Health. J. Public Health Manag. Pract. 2015, 21, 375–383, doi:10.1097/phh.0000000000000041.

35. 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, doi:10.2196/30451.

36. 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, doi:10.3349/ymj.2019.60.8.796.

37. 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, doi:10.2196/publichealth.5059.

38. 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, doi:10.1016/j.ajic.2016.04.253.

39. 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, doi:10.3390/idr14060087.

40. 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, doi:10.1007/s10639-022-11493-8.

41. Li, X.; Hasan, S.; Culotta, A. Identifying Hurricane Evacuation Intent on Twitter. Proceedings of the International AAAI Conference on Web and Social Media 2022, 16, 618–627, doi:10.1609/icwsm.v16i1.19320.

42. Lawelai, H.; Sadat, A.; Suherman, A. Democracy and Freedom of Opinion in Social Media: Sentiment Analysis on Twitter. prj 2022, 10, 40–48, doi:10.55678/prj.v10i1.585.

43. 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, 2021; pp. 322–327 ISBN 9783030553067.

44. 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, 2022; pp. 563–571 ISBN 9783030855390.

45. 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, doi:10.2196/35788.

46. Chang, R.-C.; Rao, A.; Zhong, Q.; Wojcieszak, M.; Lerman, K. #RoeOverturned: Twitter Dataset on the Abortion Rights Controversy. Proceedings of the International AAAI Conference on Web and Social Media 2023, 17, 997–1005, doi:10.1609/icwsm.v17i1.22207.

47.   Bhatia, K.V. Hindu Nationalism Online: Twitter as Discourse and Interface. Religions (Basel) 2022, 13, 739, doi:10.3390/rel13080739.

48.   Peña-Fernández, S.; Larrondo-Ureta, A.; Morales-i-Gras, J. Feminism, gender identity and polarization in TikTok and Twitter. Comunicar 2023, 31, 49–60, doi:10.3916/c75-2023-04.

49.   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, doi:10.1002/aepp.13258.

50.   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 Usi Ng NLP and Supervised KNN Classification Algorithm. Indones. J. Electr. Eng. Comput. Sci. 2021, 23, 463, doi:10.11591/ijeecs.v23.i1.pp463-470.

51.   Sontayasara, T.; Jariyapongpaiboon, S.; Promjun, A.; Seelpipat, N.; Saengtabtim, K.; Tang, J.; Leelawat, N.; Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand; International School of Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand; Disaster and Risk Management Information Systems Research Group, Chulalongkorn University, Bangkok, Thailand Twitter Sentiment Analysis of Bangkok Tourism during COVID-19 Pandemic Using Support Vector Machine Algorithm. J. Disaster Res. 2021, 16, 24–30, doi:10.20965/jdr.2021.p0024.

52.   Asgari-Chenaghlu, M.; Nikzad-Khasmakhi, N.; Minaee, S. Covid-Transformer: Detecting COVID-19 Trending Topics on Twitter Using Universal Sentence Encoder. arXiv [cs.CL] 2020.

53.   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, doi:10.1016/j.patcog.2021.108404.

54.   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, doi:10.2196/25108.

55.   Al-Ramahi, M.; Elnoshokaty, A.; El-Gayar, O.; Nasralah, T.; Wahbeh, A. Public Discourse against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data. JMIR Public Health Surveill. 2021, 7, e26780, doi:10.2196/26780.

56.   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, doi:10.1016/j.chaos.2020.110037.

57.   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, doi:10.1016/j.rinp.2021.104266.

58.   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, doi:10.1504/ijbsr.2021.117316.

59.   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, doi:10.1177/0020764020940741.

60.   Glowacki, E.M.; Wilcox, G.B.; Glowacki, J.B. Identifying #addiction Concerns on Twitter during the Covid-19 Pandemic: A Text Mining Analysis. Subst. Abus. 2021, 42, 39–46, doi:10.1080/08897077.2020.1822489.

61.   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, doi:10.1177/02692163211017026.

62.   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, doi:10.1016/j.jpsychires.2020.11.015.

63.   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 Infoveillance Study. JMIR Public Health Surveill. 2020, 6, e19509, doi:10.2196/19509.

64.   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, doi:10.3390/ijerph18031127.

65.   Pokharel, B.P. Twitter Sentiment Analysis during Covid-19 Outbreak in Nepal. SSRN Electron. J. 2020, doi:10.2139/ssrn.3624719.

66.   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); IEEE, 2020; pp. 1–7.

67.   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, doi:10.3390/ijerph18115993.

68.   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, doi:10.1088/1757-899x/1156/1/012011.

16

69. Thakur, N. A Large-Scale Dataset of Twitter Chatter about Online Learning during the Current COVID-19 Omicron Wave. Data (Basel) 2022, 7, 109, doi:10.3390/data7080109.

70. 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, doi:10.1088/1757-899x/1088/1/012045.

71. Lohar, P.; Xie, G.; Bendechache, 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 The 16th International Conference on Availability, Reliability and Security; ACM: New York, NY, USA, 2021.

72. Awoyemi, T.; Ebili, U.; Olusanya, A.; Ogunniyi, K.E.; Adejumo, A.V. Twitter Sentiment Analysis of Long COVID Syndrome. Cureus 2022, doi:10.7759/cureus.25901.

73. 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); IEEE, 2022; Vol. 1, pp. 140–148.

74. 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).

75. 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, doi:10.2196/12264.

76. Talbot, C.; O'Dwyer, S.; Clare, L.; Heaton, J.; Anderson, J. Identifying People with Dementia on Twitter. Dementia 2020, 19, 965–974, doi:10.1177/1471301218792122.

77. Almouzini, S.; Khemakhem, M.; Alageel, A. Detecting Arabic Depressed Users from Twitter Data. Procedia Comput. Sci. 2019, 163, 257–265, doi:10.1016/j.procs.2019.12.107.

78. 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. 2018, doi:10.48550/ARXIV.1805.09959.

79. 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: Berlin, Heidelberg, 2011; pp. 18–26 ISBN 9783642236341.

80. 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 Proceedings of the 2018 International Conference on Computing and Big Data; ACM: New York, NY, USA, 2018.

81. 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, 10.1097/CIN.0000000000000985, doi:10.1097/cin.0000000000000985.

82. Coppersmith, G.; Harman, C.; Dredze, M. Measuring Post Traumatic Stress Disorder in Twitter. Proceedings of the International AAAI Conference on Web and Social Media 2014, 8, 579–582, doi:10.1609/icwsm.v8i1.14574.

83. 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, doi:10.1186/s12903-021-01825-4.

84. Tsytsarau, M.; Palpanas, T. Survey on Mining Subjective Data on the Web. Data Min. Knowl. Discov. 2012, 24, 478–514, doi:10.1007/s10618-011-0238-6.

85. 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).

86. Liu, B. Sentiment Analysis and Opinion Mining; Springer Nature: Cham, Switzerland, 2022; ISBN 9783031021459.

87. Medhat, W.; Hassan, A.; Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Eng. J. 2014, 5, 1093–1113, doi:10.1016/j.asej.2014.04.011.

88. 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).

89. 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, doi:10.1016/j.eswa.2018.10.003.

90. 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, doi:10.1109/taffc.2020.2970399.

91. Hutto, C.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media 2014, 8, 216–225, doi:10.1609/icwsm.v8i1.14550.

92.   Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T. YALE: Rapid Prototyping for Complex Data Mining Tasks. In Proceedings of the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; ACM: New York, NY, USA, 2006.

93.   Fung, M. Twitter-Long COVID 2023 Available online: https://www.kaggle.com/datasets/matt0922/twitter-long-covid-2023 (accessed on 5 September 2023).