Preprints.org

Analyzing Public Reactions during the MPox Outbreak: Findings from Topic Modeling of Tweets

Nirmalya Thakur [*] , Yuvraj Nihal Duggal , Zihui Liu

Communication

Analyzing Public Reactions during the MPox Outbreak: Findings from Topic Modeling of Tweets

Nirmalya Thakur *, Yuvraj Nihal Duggal and Zihui Liu

Department of Computer Science, Emory University, Atlanta, GA 30322, USA; nirmalya.thakur@emory.edu (N.T.) yuvraj.nihal.duggal@emory.edu (Y.N.D), chloe.liu2@emory.edu (Z.L.)

* Correspondence: nirmalya.thakur@emory.edu

Abstract: In the last decade and a half, the world has experienced the outbreak of a range of viruses such as COVID-19, H1N1, flu, Ebola, Zika Virus, Middle East Respiratory Syndrome (MERS), Measles, and West Nile Virus, just to name a few. During these virus outbreaks, the usage and effectiveness of social media platforms increased significantly as such platforms served as virtual communities, enabling their users to share and exchange information, news, perspectives, opinions, ideas, and comments related to the outbreaks. Analysis of this Big Data of conversations related to virus outbreaks using concepts of Natural Language Processing such as Topic Modeling has attracted the attention of researchers from different disciplines such as Healthcare, Epidemiology, Data Science, Medicine, and Computer Science. The recent outbreak of the MPox virus has resulted in a tremendous increase in the usage of Twitter. Prior works in this field have primarily focused on the sentiment analysis and content analysis of these Tweets, and the few works that have focused on topic modeling have multiple limitations. This paper aims to address this research gap and makes two scientific contributions to this field. First, it presents the results of performing Topic Modeling on 601,432 Tweets about the 2022 Mpox outbreak, which were posted on Twitter between May 7, 2022, and March 3, 2023. The results indicate that the conversations on Twitter related to Mpox during this time range may be broadly categorized into four distinct themes - *Views and Perspectives about MPox*, *Updates on Cases and Investigations about Mpox*, *MPox and the LGBTQIA+ Community*, and *MPox and COVID-19*. Second, the paper presents the findings from the analysis of these Tweets. The results show that the theme that was most popular on Twitter (in terms of the number of Tweets posted) during this time range was - *Views and Perspectives about MPox*. It is followed by the theme of *MPox and the LGBTQIA+ Community*, which is followed by the themes of *MPox and COVID-19* and *Updates on Cases and Investigations about Mpox*, respectively. Finally, a comparison with prior works in this field is also presented to highlight the novelty and significance of this research work.

Keywords: MPox; big data; data analysis; data science; Twitter; natural language processing

1. Introduction

Monkeypox (MPox), caused by the monkeypox virus, which belongs to the Poxviridae family, Chordopoxvirinae subfamily, and Orthopoxvirus genus [1], is a re-emerging zoonotic disease. In the shape of a brick-like virion ranging from 200 nm to 250 nm, the MPox virus has a large genome of about 200 kilobase pairs encoding approximately 190 proteins [2]. Two clades of MPox, clade 1 and clade 2, show a 0.5% genomic difference, with clade 1 having a 1-12% case-fatality rate and clade 2 having a 0.1% case-fatality rate [3,4]. The first case of human MPox was recorded in a 9-month-old boy in the Democratic Republic of the Congo (DRC) in 1970 [1]. After the first case in 1970, 59 cases were reported in West and Central Africa in the next decade, with a 17% mortality rate in children under 10 [5–9]. The World Health Organization (WHO) monitored MPox cases post-1980. Between 1981 and 2017, there were multiple outbreaks of MPox in DRC due to clade 1, with the fatality rate being between 1-12%, primarily due to inadequate health systems [10–14]. From 2003-2022, few travel-related cases were reported outside endemic countries, but the number of cases was not very high [15–20]. However, a global outbreak of the MPox virus started on May 7, 2022 [21], and on July 23, 2022, the WHO declared MPox a Global Public Health Emergency (GPHE) [22]. This outbreak is

linked to a new lineage, B.1 (clade 2b), with a higher mutation rate. This outbreak has resulted in 110 countries reporting about 87,000 cases and 112 deaths so far [23–26].

The MPox virus can enter hosts via respiratory or dermal routes. As a result, infection may occur in airway epithelial cells, keratinocytes, fibroblasts, and endothelial cells [27–29]. The incubation period of MPox is 5–21 days, and common symptoms include fever (between 38.5°C and 40.5°C), headache, and myalgia. A distinguishing feature of the MPox infection is the presence of swelling at the maxillary, cervical, or inguinal lymph nodes [30]. Currently, no FDA-approved treatments for MPox exist. At present, in the United States, there are three vaccines - JYNNEOS, ACAM2000®, and APSV [31], that are available. Out of these three vaccines, JYNNEOS has been approved by the FDA for smallpox and monkeypox in adults at high risk [32]. Tecovirimat is effective against Orthopoxvirus in animals but untested in human Mpox [33]. Human trials indicate minor side effects and ongoing randomized controlled trials are assessing its safety and efficacy for MPox [34]. Other potential treatments for MPox include VIGIV, cidofovir, and brincidofovir, with proven in vitro and animal efficacy but limited availability [35].

Since the first case of this outbreak, various policy-making bodies of the world have taken measures to contain the spread of the MPox virus. For instance, the New York City Health + Hospitals (NYC H+H), an integrated healthcare system, has been pivotal in the fight against emerging pathogens in the New York City region of the United States. As a result of the work done by NYC H+H, a total of 99,079 New Yorkers have received at least one shot of the MPox vaccine [36]. However, the uncertainty and challenges surrounding asymptomatic transmission of viruses such as Mpox [37,38] and the absence or inadequacy of appropriate and effective transmission-based personal protective equipment (PPE), such as N95 masks, face shields, gowns, and extended cuff examination gloves, may result is a possible resurgence of MPox [39,40]. The fifth meeting of the WHO's International Health Regulations (IHR) Emergency Committee on the Multi-Country Outbreak of Mpox took place on May 10, 2023. At this meeting, the committee *acknowledged remaining uncertainties about the disease, regarding modes of transmission in some countries, poor quality of some reported data, and continued lack of effective countermeasures in the African countries, where mpox occurs regularly* [41]. On May 15, 2023, the US Centers for Disease Control and Prevention (CDC) warned about the potential resurgence of MPox cases in the US [42,43].

In today's Internet of Everything era [44], social media platforms provide a seamless and virtual means for users to connect, communicate, and collaborate with each other. The popularity of social media platforms has been growing exponentially in the last few years. At present, 4.9 billion people use social media, and this number is expected to rise to 5.85 billion by 2027 [45]. Among the numerous social media platforms that have been used by the public in the last decade and a half, Twitter has been highly popular amongst all age groups. Twitter has over 368 million active monthly users [46]. Twitter stands out as the social media site that journalists choose to use [47] and is among the sites with the highest global adoption rates [48]. Twitter has been highly popular amongst researchers for studying, analyzing, modeling, and interpreting social media communications related to a wide range of research problems from different domains. During the virus outbreaks that took place in the last few years, such as COVID-19, H1N1, flu, Ebola, Zika Virus, Middle East Respiratory Syndrome (MERS), Measles, and West Nile Virus, just to name a few [49], topic modeling of Tweets helped to understand the perception, preparedness, response, views, and opinions of the general public during these virus outbreaks. Prior works in this field have primarily focused on the sentiment analysis and content analysis of Tweets about this MPox outbreak. Since the first case of this outbreak, only a couple of studies have been published that have focused on topic modeling of the Tweets. However, these studies have multiple limitations centered around (1) the limited time range of the Tweets that were analyzed and (2) the limited number of Tweets that were analyzed. Addressing this research gap serves as the main motivation for this work. The rest of the paper is organized as follows. A comprehensive review of recent works in this field is presented in Section 2. Section 3 describes the methodology that was followed for this work. It is followed by Section 4, which presents the results and highlights the novel findings of this work. Section 4 is followed by the conclusion section that summarizes the contributions of this paper and outlines the scope for future work in this area.

2. Literature Review

Mining and analysis of Tweets for the investigation and exploration of different research questions has been of significant interest to researchers from a wide range of disciplines in the last few years. While misinformation presents some challenges, it is still crucial to understand web behavior on Twitter and its implications for real-world decision-making. Therefore, this section is section is divided into three parts. Section 2.1 presents a brief review of mining and analysis of Tweets for interdisciplinary research. Section 2.2 discusses the recent advances related to the study and analysis of Tweets in the field of Healthcare. Section 3.3 outlines the latest works that focused on the mining and analysis of Tweets about the MPox outbreak.

2.1. A Brief Review of Recent Works Related to the Mining and Analysis of Tweets for Interdisciplinary Research

Shaheer et al. [50] analyzed tweets calling for tourism boycotts in China, Kerala (India), Spain, and South Africa due to concerns about animal abuse and identified three strategies used to mobilize support for the boycotts. Abu Samah et al. [51] proposed a web-based dashboard to visualize customer sentiment towards Malaysian airline companies, as expressed on Twitter. Similar to the tourism industry, the entertainment industry and politics have also benefitted from the mining and analysis of Tweets. Bodaghi et al. [52] explored the differences between the web behaviors of actors who spread fake news and those who spread the truth on Twitter. This study showed that while fake news has much better modularity and intra-to inter-links ratio, truth tweeters generally have higher page rank centrality. Ante's study [53] focused on the analysis of Elon Musk's Twitter activity on cryptocurrency to highlight the significance of the so-called "Musk Effect", suggesting the seriousness of issues like market manipulation by influential individuals. Collins et al. [54] conducted a comparative analysis of over 2,000 tweets from the first two U.S. presidents of the "Twitter era", Barack Obama and Donald Trump, to assess the impact of their online correspondence on America's image abroad and its soft power. An interesting finding of this study was that the tone of presidential tweets can have significant and divergent effects on the perception of the U.S., even internationally. Beyond the border of the United States, Berrocal-Gonzalo et al. [55] found that *politainment*, the phenomenon of trivializing political information for entertainment purposes, manifested on Twitter during the Spanish general elections in April 2019, which prevented the creation of meaningful debates or interactions surrounding the elections on the platform.

The Big Data of conversations on Twitter is also considerably informative regarding the unremitting controversies underlying human society and human rights. Following the United States Supreme Court decision in Dobbs vs. Jackson Women's Health Organization that overturned abortion rights, Chang et al. [56] presented a large-scale Twitter dataset collected on the abortion rights debate in the U.S. Similarly, Peña-Fernández et al. [57] analyzed the polarization produced in social media debates regarding the rights of feminism and transgender, specifically the use of the term "TERF" (trans-exclusionary radical feminist) on Twitter. The findings of this work show that online debates are poorly inclusive, suggesting the prevalence of community isolation. Goetz et al. [58] analyzed sentiments in food security-related Tweets in the U.S. during the early stages of the COVID-19 pandemic, from which they found that keywords of negative emotions were statistically correlated with contemporaneous food insufficiency rates reported in the Household Pulse Survey. Tao et al. [59] conducted a comparative study of posts on Twitter and Weibo regarding the Russian-Ukrainian War to reveal the differences in the topics of posts between the two platforms and to call for humanitarianism and peace. As can be seen from this brief review, mining, and analysis of Tweets holds the potential for the investigation of research questions across different disciplines. In the context of different virus outbreaks that the world has witnessed in the last decade and a half, healthcare-based research using Tweets has emerged as a crucial utilization of this vast potential of mining and analyzing Tweets. Some recent works are briefly reviewed in Section 2.2, which is followed by a dedicated review of recent works related to the analysis of Tweets about the MPox outbreak.

2.2. A Brief Review of Recent Works Related to the Mining and Analysis of Tweets for Healthcare Research

The Big Data of conversations and information exchange from social media platforms, specifically Twitter, has the potential to improve the efficiency, accuracy, and coverage of the healthcare systems in different geographic regions. Skovgaard et al. [60] analyzed Twitter discussions on personalized medicine. Their study revealed an intriguing distinction in attitudes between the professionals and the general private users, with the former considering personalized medicine as a promising future and the latter being concerned about new infrastructures and their implementation. Thakur et al. [61,62] developed a framework to address loneliness and social isolation in the elderly using Twitter data. Cevik et al. [63] performed a comprehensive study to analyze the sentiments of Tweets about Parkinson's disease. Kesler et al. [64] applied topic modeling and qualitative content analysis to comments related to cancer-related cognitive impairment (CRCI), revealing the importance of coping mechanisms. Klein et al. [65,66] demonstrated the potential of using Twitter data to identify the start and end of the 40-week prenatal period, making it a valuable resource for observational studies on potential risk factors in pregnancy. Thackeray et al. [67] analyzed how Twitter is used during Breast Cancer Awareness Month (BCAM). The findings showed that while organizations and celebrities emphasized fundraisers, early detection, and diagnoses, the general public constituted the majority of the tweets that did not promote any specific preventive behavior.

Tweets provide a timely and authentic record of public perception and understanding of human health crises. The comprehensive study by Russell et al. [68] highlighted the need for regulatory changes to restrict online advertising volume about gambling activities during the COVID-19 lockdown. To investigate how the Dengue epidemic was reflected on Twitter, Gomide et al. [69] proposed an active surveillance methodology based on volume, location, time, and public perception. The analysis found a high correlation between the number of cases reported by official statistics and the number of tweets posted during the same period. The work done by Radzikowski et al. [70] reported that news organizations had a higher impact than health organizations in communicating health-related information. By examining the use of Twitter data to track public sentiment and disease activity related to H1N1 or swine flu, Signorini et al. [71] found that estimates of influenza-like illness derived from Twitter chatter accurately tracked reported disease levels from governmental organizations. Similar conclusions have been drawn from studies that focused on the recent outbreaks of listeriosis and cholera [72,73]. With the realization of Twitter's ability to provide real-time information, Voss et al. [74] collected and analyzed tweets tagged with #WestNileVirus and #WNV and concluded that unusually higher temperatures and mosquito activities led to an increase in tweet numbers about WNV. In response to the 2015 U.S. Salmonella outbreak in cucumbers imported from Mexico, Bolotova et al. [75] were able to draw some insightful conclusions by analyzing Twitter communications at the time of CDC's official announcements and the official release of the first recall of cucumbers. To inform health promotion efforts, Porat et al. [76] analyzed the content and source of popular tweets related to a diphtheria case in Spain. The most notable conclusion from their study is their suggestion for healthcare organizations to collaborate with popular journalists, news outlets, and science authors to address public concerns and misinformation through the outlet of social media like Twitter. As can be seen from this brief review, the study, analysis, and interpretation of multimodal characteristics of Tweets about different virus outbreaks has helped in the timely advancement of research in the field of Healthcare. Section 2.3 specifically highlights the recent works in this field that focused on the analysis of Tweets about MPox.

2.3. Review of Recent Works related to the Mining and Analysis of Tweets about MPox

Knudsen et al. [77] studied 262 Tweets to describe MPox risks to students. The results showed that credentialed Twitter users were 4.6 times more likely to tweet inaccurate information about MPox. Zuhanda et al. [78] studied 5,000 Tweets about MPox posted on August 5, 2022, to perform sentiment analysis. The results showed that 51.92% of the Tweets had a negative sentiment, and 48.08% of the Tweets had a positive sentiment. Ortiz-Martínez et al. [79] performed a study of 100 top Tweets about MPox posted on May 24, 2022. The findings showed that most of the Tweets were posted by informal individuals or groups (60%), followed by healthcare or public health (32%), and

news outlets or journalists (8%). The work by Rahmanian et al. [80] involved studying 384,560 Tweets posted between May 16, 2022, and May 22, 2022. The findings showed that the majority of these Tweets were posted by individuals from the United States and Canada. Cooper et al. [81] studied Tweets containing the word "monkeypox" posted between May 1, 2022, and July 23, 2022. The results showed that a total of 48330 Tweets were posted by LGBTQ+ self-identified advocates or allies. The work of Ng et al. [82] focused on studying 352,182 Tweets about MPox posted between May 6, 2022, and July 23, 2022. The authors performed topic modeling of these Tweets and derived three themes - concerns of safety, stigmatization of minority communities, and a general lack of faith in public institutions. Bengesi et al. [83] mined over 500,000 multilingual tweets related to MPox and performed sentiment analysis. Olusegun et al. [84] studied 800,000 Tweets about MPox and used NRCLexicon to predict and measure the emotional significance of each Tweet. Farahat et al. [85] performed sentiment analysis on a total of 8532 Tweets about MPox posted between May 22, 2022, and August 5, 2022. The findings of sentiment analysis showed that 48% of the Tweets were neutral, 37% of the Tweets were positive, and 15% of the Tweets were negative. Sv et al. [86] studied a total of 556,402 Tweets about MPox posted between June 1, 2022, and June 25, 2022. The results of topic modeling showed that among the Tweets about MPox that had negative sentiments, there was a range of topics that were represented, such as deaths caused by the MPox virus, the severity of the virus, lesions caused by the virus, whether the virus is airborne, vaccines for the virus, and whether the virus will lead to the next pandemic. Mohbey et al. [87] developed a CNN and LSTM-based model to perform sentiment analysis of Tweets about MPox, and the accuracy of the model was found to be 94%. A dataset of Tweets about the MPox outbreak was developed by Nia et al. [88]. The work of Iparraguirre-Villanueva [89] focused on the detection of polarity in conversations on Twitter about MPox. The results showed that 45.42% of people expressed neither positive nor negative opinions, while 19.45% expressed negative and fearful feelings about MPox. AL-Ahdal [90] studied 15,936 Tweets about MPox posted by individuals from Germany. The results showed that the public displayed an impersonal feeling toward MPox.

As can be seen from this review and the research questions that were investigated in these papers, the major focus areas of these works have been sentiment analysis and content analysis. Only a couple of works have focused on the topic modeling of Tweets about MPox. However, those works have multiple limitations centered around the limited time range of the Tweets that were analyzed and the limited number of Tweets that were analyzed (discussed in detail in Section 4). This study aims to address this research gap. Therefore, topic modeling of 601,432 Tweets about the 2022 Mpox outbreak, which was posted on Twitter between May 7, 2022, and March 3, 2023, was performed in this study. Section 3 outlines the step-by-step methodology that was followed for the system design and implementation. The results and novel contributions of this work are presented and discussed in Section 4.

3. Methodology

This section presents the methodology that was followed for the system design and implementation. This section is divided into three parts. In Section 3.1, a technical overview of RapidMiner [91] is presented as RapidMiner was used for this work. Section 3.2 presents the description of the topic modeling architecture that was used in this work. Section 3.3 outlines the steps that were followed for the implementation of this topic modeling architecture, along with the specifics of the system design.

3.1. Technical Overview of RapidMiner

RapidMiner is a Data Science software platform that allows the development and implementation of different algorithms. It enables its users to visually design data workflows and build predictive models using a graphical user interface (GUI). RapidMiner allows the development of applications, workflows, and algorithms, which are known as "processes" that consist of multiple "operators", which may be built-in and/or user-defined. Every "process" and each of its "operators" are associated with a specific functionality to make the entire "process" work. RapidMiner provides

a wide range of built-in "operators" that may be directly used for the implementation of various tasks. There also exist certain "operators" that may be used to modify the functionality or features of other "operators". The platform also allows developers to create their own "operators" which can be shared via the RapidMiner Marketplace. RapidMiner is developed on a client-server model with public and private cloud infrastructures. There is a free version and an enterprise version. The free version has a data processing limit of 10,000 rows for any "process", and the enterprise version does not have any limits on the number of rows in a data file that it can process [91,92]. For this research work, the educational license of RapidMiner (available to researchers in academia upon request) was used in RapidMiner Studio 10.1.001. With the educational license, RapidMiner Studio can be used to process any number of rows in any dataset. The following represent a few notable characteristics of RapidMiner Studio [93,94]:

1. It supplies pre-built "operators" encompassing distinct functions that can be directly employed or customized for the creation and execution of algorithms and applications related to Machine Learning, Data Science, Artificial Intelligence, and Big Data.
2. RapidMiner is developed using Java, which ensures that RapidMiner "workflows" retain the Write Once Run Anywhere (WORA) attribute of Java.
3. The platform permits the installation of various extensions to facilitate seamless connectivity and integration of RapidMiner "workflows" with other software and hardware environments.
4. Scripts developed in programming languages, like Python and R, can also be imported into a RapidMiner "workflow" to supplement its functionalities.
5. The software enables the creation of new "operators" and effortless dissemination of the same within the RapidMiner community.
6. RapidMiner consists of "operators" that enable it to establish connections with social media platforms, such as Twitter and Facebook. Such connections facilitate the extraction of tweets, comments, posts, reactions, and other relevant social media interactions.

3.2. Description of the Topic Modeling Architecture for System Design

Latent Dirichlet Allocation (LDA) [95] is a generative probabilistic model used in the field of Natural Language Processing and Machine Learning. It is commonly used for topic modeling, which is the task of identifying topics within a collection of documents. In LDA, the mixture of topics is derived from a consistent Dirichlet prior, which is the same across all documents. The procedure [96] for creating a corpus is outlined as follows (in this context, the focus is on smoothed LDA). Thereafter, the likelihood of generating a corpus can be represented as shown in Equation (1).

1. Select a multinomial distribution $\emptyset_z$ for each topic z from a Dirichlet distribution with parameter β
2. For each document d, pick a multinomial distribution θ_d also from a Dirichlet distribution with parameter α
3. In document d, for each token word w, pick a topic z $\in$ {1….K} from the multinomial distribution θ_d
4. Pick a word w from the multinomial distribution θ_z

$$P(Doc_1, \ldots, Doc_N | \alpha, \beta) = \int \int \prod_{z=1}^{K} P(\phi_z | \beta) \prod_{d=1}^{N} P(\phi_d | \alpha) \left(\prod_{i=1}^{N_d} \sum_{z_i=1}^{K} P(z_i | \theta) P(w_i | z, \theta) \right) d\theta d\phi \qquad (1)$$

To use language models for information retrieval in an LDA, an approach using the query likelihood model is used, where each document is scored by the likelihood of its model generating a query Q. This is shown in Equations (2) and (3). In Equation (2), D represents a model for documents, Q stands for the query, and q denotes an individual term within the query Q. P(Q|D) signifies the probability of the document model generating the query terms, following the assumption of 'bag-of-words'. This assumption considers that terms are independent when given the documents. $P(q_i|D)$ is specified by the document model with Dirichlet smoothing. In Equation (3), P(w|D) is the maximum likelihood estimate of word w in document D, and P'(w|coll) is the maximum likelihood

estimate of the same word w in the entire collection. However, directly implementing the LDA model hurts retrieval performance. So, in a prior work in this field [96], the original document model (Equation 3) was combined with the LDA model to construct a new LDA-based document model, as shown in Equation (4). The LDA model introduces a novel document representation centered around topics. After obtaining the posterior estimates for θ and φ, the word probability within a document can be computed using Equation (5), where $\hat{\theta}$ and $\hat{\phi}$ are the posterior estimates of θ and $\emptyset$ [96].

$$P(Q|D) = \prod_{q \in Q} P(q|D) \tag{2}$$

$$P(w|D) = \frac{N_d}{N_d + \mu} P_{ML}(w|D) + \left(1 - \frac{N_d}{N_d + \mu}\right) P_{ML}(w|coll) \tag{3}$$

$$P(w|D) = \lambda \left(\frac{N_d}{N_d + \mu} P_{ML}(w|D) + \left(1 - \frac{N_d}{N_d + \mu}\right) P_{ML}(w|coll)\right) + (1 - \lambda) P_{lda}(w|D) \tag{4}$$

$$P(w|d, \hat{\theta}, \hat{\phi}) = \sum_{z=1}^{K} P(w|z, \hat{\phi}) P(z|\hat{\theta}, d) \tag{5}$$

In this research work, SparseLDA was implemented, as prior work has shown that it is 20 times faster than the traditional LDA [97]. In the SparseLDA framework [97], given an observed word type w, the probability of topic z in document d can be computed using Equation (6).

$$P(z = t|w)\alpha \left(\alpha_t + n_{(t|d)}\right) \frac{\beta + n_{w|t}}{\beta V + n_{.|t}} \tag{6}$$

However, Equation (6) requires the calculation of the unnormalized weight q(z) for all topics to determine the normalizing constant for the distribution $\sum_z q(z)$. A simpler approach involves storing a significant portion of the computation needed to calculate the normalization constant. Through the reorganization of terms within the numerator, Equation (5) can be portioned into three distinct sections, as shown in Equations (7) to (10). Here, the first term is constant for all documents, and the second term is independent of the current word type. Moreover, the $\sum_z q(z)$ corresponds to the sum across topics for each of the three components in Equation (7) [97].

$$P(z = t|w)\alpha \frac{\alpha_t \beta}{\beta V + n_{.|t}} + \frac{n_{t|d}\beta}{\beta V + n_{.|t}} + \frac{(\alpha_t + n_{t|d})n_{w|t}}{\beta V + n_{.|t}} \tag{7}$$

$$s = \sum_t \left(\frac{\alpha_t \beta}{\beta V + n_{.|t}}\right) \tag{8}$$

$$r = \sum_t \left(\frac{n_{t|d}\beta}{\beta V + n_{.|t}}\right) \tag{9}$$

$$q = \sum_t \left(\frac{(\alpha_t + n_{t|d})n_{w|t}}{\beta V + n_{.|t}}\right) \tag{10}$$

This divides the full sampling mass into three buckets. Now, $U \sim \mathcal{U}(0, s + r + q)$ can be sampled. If U < s, it would imply hitting the smoothing-only bucket. Thereafter, the process involves stepping through each topic and calculating and adding $\frac{\beta \alpha_t}{\beta V + n_{.|t}}$ for that topic until it is greater than 'x'. For the document bucket, s < r < (s+r), the process involves iterating through the set of topics that satisfies $n_{t|d \neq 0}$. The constant 's' only changes when the hyperparameters are updated. Conversely, the constant 'r' is solely influenced by document-topic counts. This permits the computation of 'r' once at the start of each document and subsequently modifies it by subtracting and adding values related to the prior and current topics in each Gibbs update. This process takes constant time,

independent of the number of topics. The topic word constant, 'q', changes with the value of w, so old computations cannot be easily recycled. However, the performance can be significantly improved by splitting q into two components, as shown in Equation (11). With this Equation, the coefficient $\frac{(\alpha_t+n_{t|d})}{\beta V+n_{\cdot|t}}$ can be cached for every topic. Calculating 'q' for a specific 'w' involves performing a single multiplication operation for each topic, where $n_{w|t}$ is non-zero. Given that $n_{t|d}$ = 0 for all topics within any document, the coefficients vector will predominantly comprise only $\frac{(\alpha_t)}{\beta V+n_{\cdot|t}}$. Therefore, this allows the optimization of the LDA model by storing these coefficients across documents, refreshing values only for topics with non-zero counts in the current document, and reverting these values to α-only values upon finishing the sampling process for that document [97].

$$q = \sum_t \left[\frac{(\alpha_t + n_{t|d})}{\beta V + n_{\cdot|t}} \times n_{w|t}\right] \tag{11}$$

3.2. Description of the System Design and Implementation

This section describes the system design and its implementation, as well as the dataset that was used for performing this research work. The dataset that was used comprises 601,431 Tweet IDs of Tweets about MPox posted between May 7, 2022, and March 3, 2023 [98]. This dataset complies with the FAIR principles of scientific data management [99]. This dataset contains only Tweet IDs, and the standard procedure for working with such datasets is that the dataset is hydrated to obtain the text of the Tweets and related information. However, this dataset was developed by the first author of this paper, so all the Tweets were already available, and hydration of the Tweet IDs was not necessary. Figure 1 shows the system design in RapidMiner. This is a "process" that was developed in RapidMiner Studio 10.1.001 (with the Educational License) to set up this system, and this "process" comprises different "operators" with different functionalities.
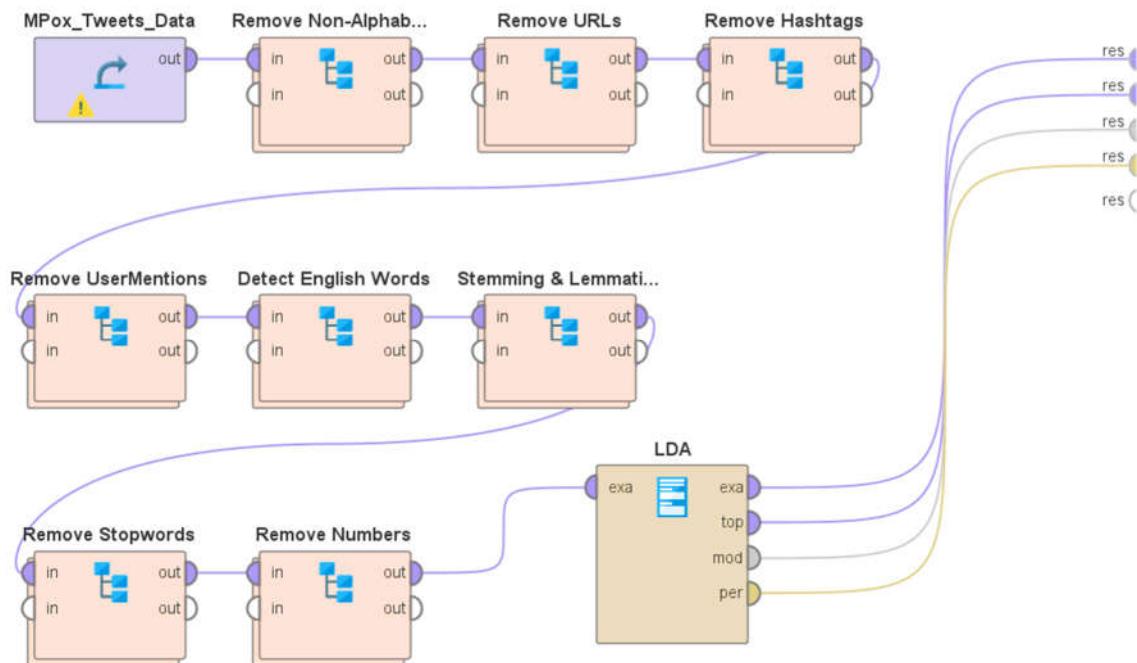
Figure 1. System Design in RapidMiner for performing Topic Modeling.

In this Figure, the "MPox_Tweets-Data" "operator" represents the Tweets from the dataset described earlier in this section. These Tweets were imported into RapidMiner Studio to develop this "process". All 601,431 Tweets about MPox posted between May 7, 2022, and March 3, 2023, were used for the development of this LDA model. Thereafter, separate "operators" were developed to perform the different steps of the data processing. The data processing comprised the following steps,

and for each of these steps, a separate "operator" was developed in this RapidMiner "process". For steps (a), (b), (c), and (d) of the data preprocessing, different regular expressions were developed and applied to define the functionalities of these "operators".

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

After completion of the data pre-processing, an LDA model was developed and implemented in RapidMiner as per the architecture of the parallel topic model and SparseLDA (described in Section 3.2) by customizing and utilizing the "Extract Topics from Data (LDA)" operator in RapidMiner Studio 10.1.001. The number of iterations for optimization was set to 1000, and the frequency of hyperparameter optimization was set to 10. As discussed in prior works in this field [100–103], the average coherence value of an LDA model serves as a key indicator for the determination of the optimal number of topics. So, this "process" (shown in Figure 1) was repeatedly run by varying the number of topics from 2 to 50, and the average coherence value of the model was computed and recorded for each of these runs. The results of the same to deduce the optimal number of topics, as well as the specific topics that were identified in the Tweets, are presented in Section 4.

4. Results and Discussions

This section presents the results of this work. As stated in Section 3.3, the LDA model (shown in Figure 1) was run by varying the number of topics from 2 to 50 to determine the optimal number of topics based on the analysis of the average coherence value for each run. Table 1 represents the average coherence value of this LDA model from each run. This table was compiled by varying the number of topics from 2 to 50. An analysis of the same is presented in Figure 2.

Table 1. Average coherence values of the LDA Model shown in Figure 1 for different number of topics.

Number of Topics	Average Coherence Value
2	-6.1450
3	-6.0560
4	-4.6730
5	-5.2120
6	-5.7230
7	-5.8700
8	-6.6150
9	-6.8800
10	-6.1840
11	-5.6000
12	-5.4140
13	-5.3280
14	-6.7830
15	-6.0380
16	-5.6930
17	-5.6520
18	-6.5670
19	-6.0470
20	-6.0610

21	-6.3420
22	-5.5790
23	-6.0700
24	-5.9090
25	-6.7030
26	-6.6010
27	-5.9930
28	-5.9870
29	-5.8120
30	-6.0040
31	-5.8810
32	-6.0350
33	-5.8860
34	-6.2010
35	-5.7920
36	-6.0450
37	-6.5680
38	-6.3470
39	-6.1800
40	-6.2180
41	-6.4490
42	-6.1700
43	-6.2120
44	-6.3390
45	-5.8690
46	-6.2330
47	-6.1720
48	-5.9840
49	-6.1210
50	-5.9280
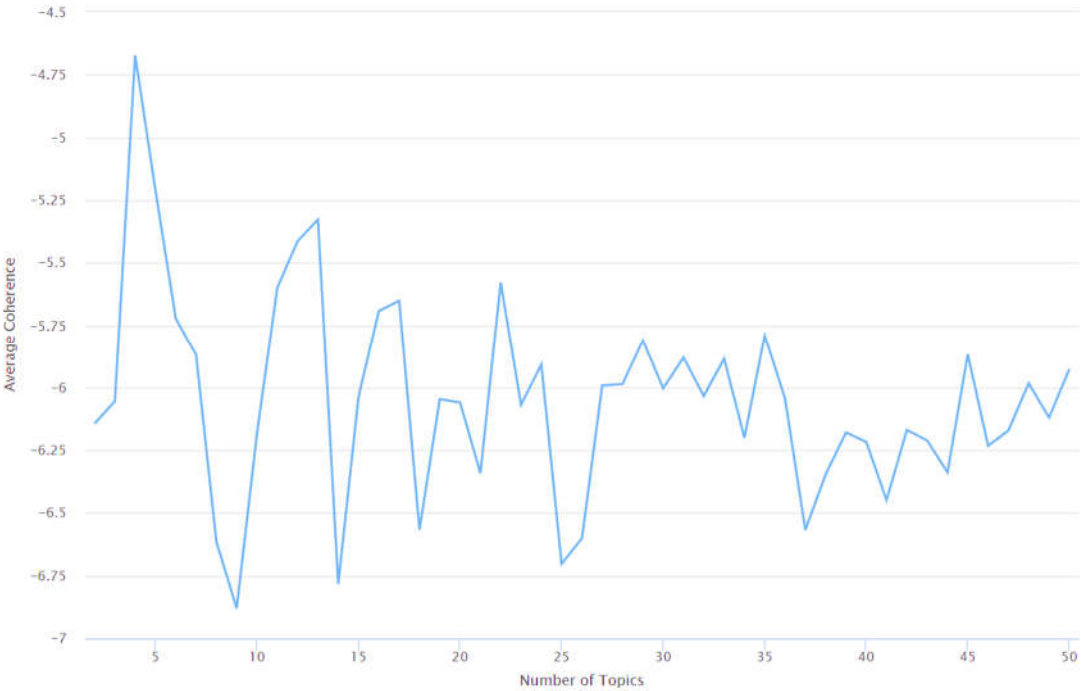
Figure 2. Analysis of the average coherence values of the LDA Model for different number of topics.

From Table 1 and Figure 2, the optimal number of topics was determined to be 4, as the LDA model produced the highest coherence score for the same. It is worth mentioning here that negative values of coherence scores are not unusual for an LDA model that follows a system architecture as described in this paper, and a recent work [104] followed a similar system architecture for the LDA model that was developed and the authors of that work obtained negative values for the coherence scores for all the number of topics, and the lowest value of the coherence score reported was as low as about -11.5.

After determining that the optimal number of topics for this LDA model is 4, this model was run by setting the number of topics as 4, and the characteristics from the output were observed and analyzed. For each Tweet, the RapidMiner "process" computed a confidence value for each topic and then predicted a topic for that Tweet based on the highest confidence value. This is shown in Figure 3. To avoid an output table comprising 601,431 rows, this Figure shows a random selection of 17 rows from this output table. In Figure 3, the attributes – *confidence(Topic_0)*, *confidence(Topic_1)*, *confidence(Topic_2)*, and *confidence(Topic_3)*, represent the confidence values of each Tweet belonging to Topics 0, 1, 2, and 3, as computed by the LDA model. The last attribute – *Tweet_Text* in this Figure, shows the Tweet (after data preprocessing) that was analyzed. For each Tweet, the highest of these confidence values was used to predict the topic for the same. For instance, in the first row in Figure 3, it can be seen that *confidence(Topic_0)* has the highest value, so the predicted topic for this Tweet was Topic 0. In a similar manner, the LDA model predicted the topics for all the Tweets of this dataset. The Tweets belonging to each of these topics – Topic 0, Topic 1, Topic 2, and Topic 3 were studied to understand the underlying themes of conversation that represented each of these topics.

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	Tweet_Text
117	116	Topic_0	0.739	0.062	0.143	0.056	also been preoccupied wit...
118	117	Topic_1	0.049	0.674	0.028	0.250	Tomorrow on Morning wit...
119	118	Topic_0	0.708	0.023	0.258	0.011	I already done enough mo...
120	119	Topic_0	0.852	0.018	0.090	0.040	The monkey pox they bee...
121	120	Topic_0	0.867	0.027	0.054	0.052	That s what they said abo...
122	121	Topic_0	0.685	0.013	0.064	0.239	people have to take of Co...
123	122	Topic_0	0.414	0.384	0.175	0.027	First it will be that monkey ...
124	123	Topic_3	0.345	0.022	0.139	0.495	don t sweat the Monkey Pox
125	124	Topic_0	0.739	0.018	0.084	0.160	All you are getting monkey...
126	125	Topic_0	0.798	0.008	0.036	0.158	This Monkey Pox I d rather...
127	126	Topic_3	0.221	0.013	0.098	0.668	anyone have any Monkey ...
128	127	Topic_0	0.504	0.010	0.173	0.313	I have no idea what s goin...
129	128	Topic_0	0.908	0.015	0.041	0.037	if that monkey pox hit the I ...
130	129	Topic_2	0.473	0.005	0.491	0.032	Why are u people showin...
131	130	Topic_0	0.887	0.013	0.053	0.047	The same who he already ...
132	131	Topic_3	0.235	0.022	0.066	0.677	Bunch of Monkey Pox now
133	132	Topic_0	0.602	0.013	0.043	0.342	MONKEY POX I wonder if t...
134	133	Topic_0	0.881	0.010	0.049	0.060	I it because i was riding th...

Figure 3. A random selection of 17 rows from the output table of the developed LDA model.

Based on this study, the broad themes which represented these topics were observed to be "Views and Perspectives about MPox", "Updates on Cases and Investigations about Mpox", "MPox and the LGBTQIA+ Community", and "MPox and COVID-19". Table 2 represents a random selection of five Tweets for each of these Topics. In Table 2, these Tweets are presented in "as is" form, i.e., in the manner in which they were originally posted on Twitter to provide better context as compared to the preprocessed version of these Tweets.

Table 2. Representation of five Tweets (selected randomly) for each Topic – Topic 0, Topic 1, Topic 2, and Topic 3.

Topic 0, Theme: Views and Perspectives about MPox

Tweet #	Original Text of the Tweet
Tweet #1	@vancemurphy @pfizer @moderna_tx @US_FDA Well, you know the new thing is monkey pox, right? Vaccines are so yesterday.
Tweet #2	Its annoys me how they use pictures of black peoples hands when they discuss monkey pox
Tweet #3	The pics of monkey pox looks exactly like shingles
Tweet #4	@masthahh1 Are there any stats on the people who have gotten monkey pox? Were they all vaccinated?
Tweet #5	Looking at the state of the UK. I'd be more worried about Monkey Pox catching a dose of Englishman!

Topic 1, Theme: Updates on Cases and Investigations about MPox

Tweet #1	BREAKING: Health department investigating possible monkey pox case in NYC
Tweet #2	New York health officials are investigating a potential case of monkeypox after a patient tested positive for the family of viruses associated with the rare illness.
Tweet #3	U.S. government officials are placing orders for millions of doses of monkeypox vaccines amid a worldwide outbreak and a possible case in New York City, the Independent reports.
Tweet #4	WHO is convening an Emergency Committee meeting out of concern for international spread of monkeypox, a high consequence infection. They will likely discuss whether to declare monkeypox a Public Health Emergency of International Concern (PHEIC)
Tweet #5	The UK Health Security Agency said the new cases of the rare monkeypox infection do not have known connections with the previous confirmed cases announced on 14 May and a case on 7 May

Topic 2, Theme: MPox and the LGBTQIA+ Community

| Tweet #1	@CraigbryCraig @BreezerGalway Moneypox has been known about since 1958. Majority of case are in gay males. No need to freak out
Tweet #2	. Gay? Had "close" contact with someone whose in the hospital now in Montreal. Apparently majority in Montreal who contracted the Monkey Pox were gay 35-50 year old men. AIDS started in the gay community too. Something about monkeying around ...
Tweet #3	@jmcrookston Just to be SUPER CLEAR, what I mean by this, is that no, monkeypox isn't a "gay disease". I'm queer and super not okay with the way the media is framing this the same way HIV/AIDS was framed in the 70s/80s.
Tweet #4	@jeffreyatucker @ezralevant Some knowledge about Monkey pox, it's mostly for gay. Not a threat.
Tweet #5	@EnemyInAState @TimothyVollmer Absolutely agree only other events won't have the stigma attached which is happening with monkey pox so many people are convinced it's a gay disease because there's no context

Topic 3, Theme: MPox and COVID-19

Tweet #1	@COVIDnewsfast Transmission of Monkey Pox is not the same as Covid!
Tweet #2	MUST WATCH: Amazing Polly exposes 2021 DAVOS pandemic event for a May 15 2022 release of Monkey-Pox! BOOM! This Monkey Pox, like COVID, is being exploited to push a global government. Amazing Polly catches them.
Tweet #3	@ANCParliament What are your plans on preventing Monkey Pox that has "accidentally" been released in the United States of America from coming into South Africa before it becomes a big Issue like Covid-19?
Tweet #4	WW3 is on the horizon, Covid-19, and Monkey Pox about to be released into the world we'll be lucky if any humans survive? Nice work Joe. #LetsGoBrandon
Tweet #5	Monkey Pox is coming! Covid did not do the trick.

[1] These Tweets are presented here in "as is" form. These Tweets do not represent or reflect the views, or opinions, or beliefs, or political stance of the authors of this paper.

Figure 4 shows an analysis of the number of Tweets per topic. As can be seen from Figure 4, Topic 0, or the theme of *Views and Perspectives about Mpox,* was most popular on Twitter (in terms of the number of Tweets posted) during this time range of May 7, 2022, to March 3, 2023. It was followed by Topic 2, or the theme of *MPox and the LGBTQIA+ Community*. This was followed by Topic 3 (or the

13

theme of *MPox and COVID-19*) and Topic 1 (or the theme of *Updates on Cases and Investigations about Mpox*), respectively.
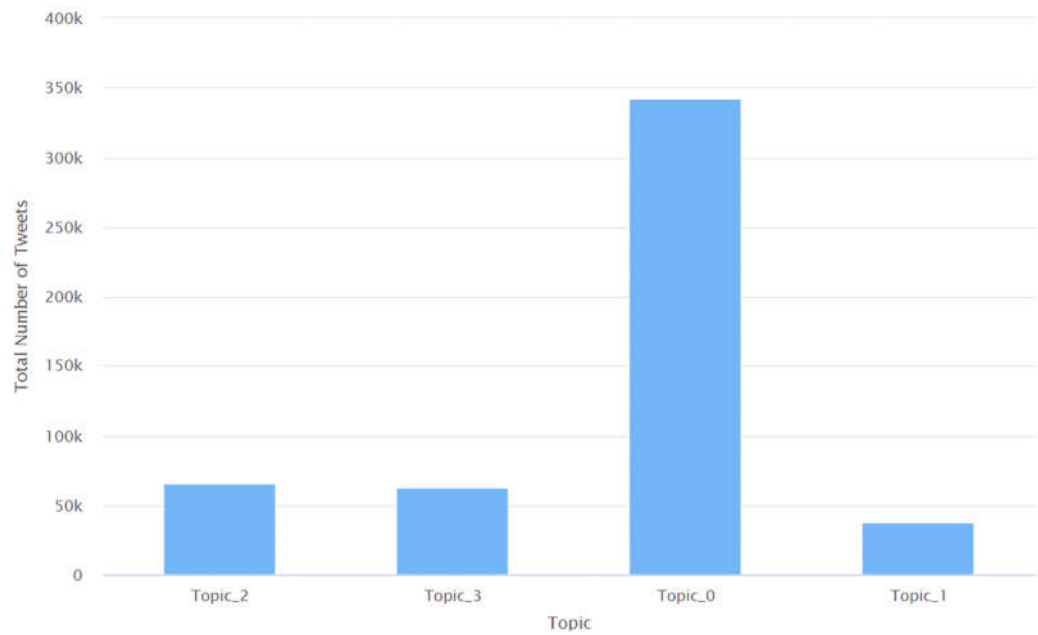
Figure 4. Analysis of the number of Tweets posted per topic.

Next, a comparative study with prior works in this field is presented. Table 3 presents a summary of recent works in this field that focused on the mining and analysis of Tweets about MPox. As can be seen from Table 3, most of the works in this field have focused on Sentiment Analysis and Content Analysis of Tweets. At the same time, a couple of works [82,86] also exist in this field where topic modeling of Tweets about Mpox was performed. As can be seen from Table 3, topic modeling of Tweets about MPox is a research area that remains less explored and less investigated. A comparison of the different characteristics of this work with these two works [82,86] is presented in Table 4.

Table 3. Summary and categorization of the recent works in this field.

Work	Sentiment Analysis	Content Analysis	Topic Modeling	Dataset Development
Knudsen et al. [77]		✓		
Zuhanda et al. [78]	✓			
Ortiz-Martínez et al. [79]		✓		
Rahmanian et al. [80]		✓		
Cooper et al. [81]	✓	✓		
Ng et al. [82]			✓	
Bengesi et al. [83]	✓			
Olusegun et al. [84]	✓			
Farahat et al. [85]	✓	✓		
Sv et al. [86]			✓	
Mohney et al. [87]	✓			
Nia et al. [88]				✓
Iparraguirre-Villanueva [89]	✓			
AL-Ahdal [90]		✓		

Table 4. Comparison of specific characteristics of this work with two similar works in this field that also focused on topic modeling of Tweets.

Work	Number of Tweets Analyzed	Time Range of the Tweets
Ng et al. [82]	352,182 Tweets	May 6, 2022, to July 23, 2022
Sv et al. [86]	556,402 Tweets	June 1, 2022, to June 25, 2022
Thakur et al. [this work]	601,432 Tweets	May 7, 2022, to March 3, 2023

Table 4 outlines the limitations in a couple of similar works [82,86] in this field and highlights how the work of this paper addresses the same. Specifically, the limitations in these two works can be listed as follows:

1. Limited time range of the analyzed Tweets: The time range of the Tweets that were analyzed in these works represents Tweets that were posted only during certain months of the 2022 MPox outbreak. One of the works [82] included Tweets that were posted on the day the first case of the 2022 MPox outbreak was recorded (May 7, 2022), but the other work [86] didn't. Furthermore, none of these works have analyzed Tweets posted after July 23, 2022.

2. Limited number of Tweets: The number of Tweets that were investigated in these works are 352,182 and 556,402, respectively. These numbers represent a fraction of the total number of Tweets that have been posted since the first recorded case of the 2022 Mpox outbreak on May 7, 2022.

These limitations that exist in similar works in this field are addressed in this paper. First, the dataset that was used for developing the LDA model comprises Tweets about MPox, which were posted on Twitter between May 7, 2022, and March 3, 2023 – a time range that is greater than the time ranges of the similar works shown in Table 4. Second, a total of 601,432 Tweets were analyzed in this study. This is much higher than the number of Tweets which were analyzed in similar works in this field. Thus, to summarize, the time range of the Tweets and the number of Tweets that were analyzed in this study further support the scientific contributions of this work.

5. Conclusions

In the last decade and a half, the world has experienced the outbreak of a range of viruses such as COVID-19, H1N1, flu, Ebola, Zika Virus, Middle East Respiratory Syndrome (MERS), Measles, and West Nile Virus, just to name a few. In today's Internet of Everything era, the popularity of social media platforms has been growing exponentially. Social media platforms have served as virtual communities during the outbreak of such viruses in the past, allowing people from different parts of the world to share and exchange information, news, perspectives, opinions, ideas, and comments related to the outbreaks. Researchers from different disciplines have analyzed this Big Data of conversations related to virus outbreaks on social media platforms such as Twitter using concepts such as Topic Modeling to understand the underlying themes of conversations and information exchange of the general public. The recent outbreak of the MPox virus has resulted in a tremendous increase in the utilization of social media platforms such as Twitter. Prior works in this field have primarily focused on sentiment analysis and content analysis of Tweets about MPox, and a couple of works in this field that have focused on topic modeling have multiple limitations. This paper aims to address this research gap and makes two scientific contributions to this field. First, it presents the results of performing topic modeling of 601,432 Tweets about the 2022 MPox outbreak, which was posted on Twitter between May 7, 2022, and March 3, 2023. These results indicate that the conversations related to MPox during this time range may be broadly categorized into four distinct themes - *Views and Perspectives about MPox*, *Updates on Cases and Investigations about MPox*, *MPox and the LGBTQIA+ Community*, and *MPox and COVID-19*. Second, the paper presents the findings from the analysis of the Tweets that focused on these topics. The results show that the theme that was most popular on Twitter (in terms of the number of Tweets posted) during this time range was *Views and Perspectives about MPox*. It is followed by the theme of *MPox and the LGBTQIA+ Community*. This theme is followed by the themes of *MPox and COVID-19* and *Updates on Cases and Investigations about Mpox*, respectively. As per the best knowledge of the authors, no similar work has been done in this field thus far. Future work in this area would involve collecting more Tweets and repeating this study

in a few months to evaluate and interpret any variations in terms of the themes of conversations on Twitter related to MPox.

Supplementary Materials: Not Applicable.

Author Contributions: Conceptualization, N.T.; methodology, N.T. and Y.N.D.; software, N.T.; validation, N.T.; formal analysis, N.T.; investigation, N.T.; resources, N.T. and Y.N.D.; data curation, N.T.; writing—original draft preparation, N.T., Y.N.D, and Z.L.; writing—review and editing, N.T., Y.N.D, and Z.L.; visualization, N.T.; supervision, N.T.; project administration, N.T.; funding acquisition, Not Applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data analyzed in this study are publicly available at https://dx.doi.org/10.21227/16ca-c879

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McCollum, A.M.; Damon, I.K. Human Monkeypox. Clin. Infect. Dis. 2014, 58, 260–267, doi:10.1093/cid/cit703.
2. Chen, N.; Li, G.; Liszewski, M.K.; Atkinson, J.P.; Jahrling, P.B.; Feng, Z.; Schriewer, J.; Buck, C.; Wang, C.; Lefkowitz, E.J.; et al. Virulence Differences between Monkeypox Virus Isolates from West Africa and the Congo Basin. Virology 2005, 340, 46–63, doi:10.1016/j.virol.2005.05.030.
3. Beer, E.M.; Rao, V.B. A Systematic Review of the Epidemiology of Human Monkeypox Outbreaks and Implications for Outbreak Strategy. PLoS Negl. Trop. Dis. 2019, 13, e0007791, doi:10.1371/journal.pntd.0007791.
4. Likos, A.M.; Sammons, S.A.; Olson, V.A.; Frace, A.M.; Li, Y.; Olsen-Rasmussen, M.; Davidson, W.; Galloway, R.; Khristova, M.L.; Reynolds, M.G.; et al. A Tale of Two Clades: Monkeypox Viruses. J. Gen. Virol. 2005, 86, 2661–2672, doi:10.1099/vir.0.81215-0.
5. Heymann, D.L.; Szczeniowski, M.; Esteves, K. Re-Emergence of Monkeypox in Africa: A Review of the Past Six Years. Br. Med. Bull. 1998, 54, 693–702, doi:10.1093/oxfordjournals.bmb.a011720.
6. Mandja, B.-A.M.; Brembilla, A.; Handschumacher, P.; Bompangue, D.; Gonzalez, J.-P.; Muyembe, J.-J.; Mauny, F. Temporal and Spatial Dynamics of Monkeypox in Democratic Republic of Congo, 2000–2015. Ecohealth 2019, 16, 476–487, doi:10.1007/s10393-019-01435-1.
7. Rimoin, A.W.; Mulembakani, P.M.; Johnston, S.C.; Lloyd Smith, J.O.; Kisalu, N.K.; Kinkela, T.L.; Blumberg, S.; Thomassen, H.A.; Pike, B.L.; Fair, J.N.; et al. Major Increase in Human Monkeypox Incidence 30 Years after Smallpox Vaccination Campaigns Cease in the Democratic Republic of Congo. Proc. Natl. Acad. Sci. U. S. A. 2010, 107, 16262–16267, doi:10.1073/pnas.1005769107.
8. Breman, J.G.; Henderson, D.A. Poxvirus Dilemmas — Monkeypox, Smallpox, and Biologic Terrorism. N. Engl. J. Med. 1998, 339, 556–559, doi:10.1056/nejm199808203390811.
9. Guarner, J.; del Rio, C.; Malani, P.N. Monkeypox in 2022—What Clinicians Need to Know. JAMA 2022, 328, 139, doi:10.1001/jama.2022.10802.
10. Nguyen, P.-Y.; Ajisegiri, W.S.; Costantino, V.; Chughtai, A.A.; MacIntyre, C.R. Reemergence of Human Monkeypox and Declining Population Immunity in the Context of Urbanization, Nigeria, 2017–2020. Emerg. Infect. Dis. 2021, 27, 1007, doi:10.3201/eid2704.203569.
11. R Pebody Human Monkeypox in Kasai Oriental, Democratic Republic of Congo, February 1996 – October 1997: Preliminary Report. Wkly. releases (1997–2007) 1997, 1, doi:10.2807/esw.01.32.01015-en.
12. Foster, S.O.; Brink, E.W.; Hutchins, D.L.; Pifer, J.M.; Lourie, B.; Moser, C.R.; Cummings, E.C.; Kuteyi, O.E.K.; Eke, R.E.A.; Titus, J.B.; et al. Human Monkeypox. Bulletin of the World Health Organization 1972, 46, 569.
13. Fuller, T.; Thomassen, H.A.; Mulembakani, P.M.; Johnston, S.C.; Lloyd-Smith, J.O.; Kisalu, N.K.; Lutete, T.K.; Blumberg, S.; Fair, J.N.; Wolfe, N.D.; et al. Using Remote Sensing to Map the Risk of Human Monkeypox Virus in the Congo Basin. Ecohealth 2011, 8, 14–25, doi:10.1007/s10393-010-0355-5.
14. Jezek, Z.; Szczeniowski, M.; Paluku, K.M.; Mutombo, M. Human Monkeypox: Clinical Features of 282 Patients. J. Infect. Dis. 1987, 156, 293–298, doi:10.1093/infdis/156.2.293.

15. Erez, N.; Achdout, H.; Milrot, E.; Schwartz, Y.; Wiener-Well, Y.; Paran, N.; Politi, B.; Tamir, H.; Israely, T.; Weiss, S.; et al. Diagnosis of Imported Monkeypox, Israel, 2018. Emerg. Infect. Dis. 2019, 25, 980–983, doi:10.3201/eid2505.190076.

16. Yong, S.E.F.; Ng, O.T.; Ho, Z.J.M.; Mak, T.M.; Marimuthu, K.; Vasoo, S.; Yeo, T.W.; Ng, Y.K.; Cui, L.; Ferdous, Z.; et al. Imported Monkeypox, Singapore. Emerg. Infect. Dis. 2020, 26, 1826–1830, doi:10.3201/eid2608.191387.

17. Reed, K.D.; Melski, J.W.; Graham, M.B.; Regnery, R.L.; Sotir, M.J.; Wegner, M.V.; Kazmierczak, J.J.; Stratman, E.J.; Li, Y.; Fairley, J.A.; et al. The Detection of Monkeypox in Humans in the Western Hemisphere. N. Engl. J. Med. 2004, 350, 342–350, doi:10.1056/nejmoa032299.

18. Zachary, K.C.; Shenoy, E.S. Monkeypox Transmission Following Exposure in Healthcare Facilities in Nonendemic Settings: Low Risk but Limited Literature. Infect. Control Hosp. Epidemiol. 2022, 43, 920–924, doi:10.1017/ice.2022.152.

19. Adler, H.; Gould, S.; Hine, P.; Snell, L.B.; Wong, W.; Houlihan, C.F.; Osborne, J.C.; Rampling, T.; Beadsworth, M.B.J.; Duncan, C.J.A.; et al. Clinical Features and Management of Human Monkeypox: A Retrospective Observational Study in the UK. Lancet Infect. Dis. 2022, 22, 1153–1162, doi:10.1016/s1473-3099(22)00228-6.

20. Vaughan, A.; Aarons, E.; Astbury, J.; Brooks, T.; Chand, M.; Flegg, P.; Hardman, A.; Harper, N.; Jarvis, R.; Mawdsley, S.; et al. Human-to-Human Transmission of Monkeypox Virus, United Kingdom, October 2018. Emerg. Infect. Dis. 2020, 26, 782–785, doi:10.3201/eid2604.191164.

21. Saxena, S.K.; Ansari, S.; Maurya, V.K.; Kumar, S.; Jain, A.; Paweska, J.T.; Tripathi, A.K.; Abdel-Moneim, A.S. Re‐emerging Human Monkeypox: A Major Public‐health Debacle. J. Med. Virol. 2023, 95, doi:10.1002/jmv.27902.

22. Kozlov, M. Monkeypox Declared a Global Emergency: Will It Help Contain the Outbreaks? Nature 2022, doi:10.1038/d41586-022-02054-7.

23. 2022 Mpox Outbreak Global Map Available online: https://www.cdc.gov/poxvirus/mpox/response/2022/world-map.html (accessed on 31 August 2023).

24. Multi-Country Outbreak of Mpox, External Situation Report #22 - 11 May 2023 Available online: https://www.who.int/publications/m/item/multi-country-outbreak-of-mpox--external-situation-report--22---11-may-2023 (accessed on 31 August 2023).

25. Joint ECDC-WHO Regional Office for Europe Mpox Surveillance Bulletin Available online: https://monkeypoxreport.ecdc.europa.eu/ (accessed on 31 August 2023).

26. Mpox (Monkeypox) Available online: https://www.who.int/news-room/fact-sheets/detail/monkeypox (accessed on 31 August 2023).

27. Mark Elwood, J. Smallpox and Its Eradication. Journal of Epidemiology and Community Health 1989, 43, 92.

28. Liu, L.; Xu, Z.; Fuhlbrigge, R.C.; Peña-Cruz, V.; Lieberman, J.; Kupper, T.S. Vaccinia Virus Induces Strong Immunoregulatory Cytokine Production in Healthy Human Epidermal Keratinocytes: A Novel Strategy for Immune Evasion. J. Virol. 2005, 79, 7363–7370, doi:10.1128/jvi.79.12.7363-7370.2005.

29. MacLeod, D.T.; Nakatsuji, T.; Wang, Z.; di Nardo, A.; Gallo, R.L. Vaccinia Virus Binds to the Scavenger Receptor MARCO on the Surface of Keratinocytes. J. Invest. Dermatol. 2015, 135, 142–150, doi:10.1038/jid.2014.330.

30. Charniga, K.; Masters, N.B.; Slayton, R.B.; Gosdin, L.; Minhaj, F.S.; Philpott, D.; Smith, D.; Gearhart, S.; Alvarado-Ramy, F.; Brown, C.; et al. Estimating the Incubation Period of Monkeypox Virus during the 2022 Multi-National Outbreak. bioRxiv 2022.

31. Vaccines Available online: https://www.cdc.gov/smallpox/clinicians/vaccines.html (accessed on 31 August 2023).

32. FDA Approves First Live, Non-Replicating Vaccine to Prevent Smallpox and Monkeypox Available online: https://www.fda.gov/news-events/press-announcements/fda-approves-first-live-non-replicating-vaccine-prevent-smallpox-and-monkeypox (accessed on 31 August 2023).

33. Berhanu, A.; Prigge, J.T.; Silvera, P.M.; Honeychurch, K.M.; Hruby, D.E.; Grosenbach, D.W. Treatment with the Smallpox Antiviral Tecovirimat (ST-246) Alone or in Combination with ACAM2000 Vaccination Is Effective as a Postsymptomatic Therapy for Monkeypox Virus Infection. Antimicrob. Agents Chemother. 2015, 59, 4296–4300, doi:10.1128/aac.00208-15.

34. Grosenbach, D.W.; Honeychurch, K.; Rose, E.A.; Chinsangaram, J.; Frimm, A.; Maiti, B.; Lovejoy, C.; Meara, I.; Long, P.; Hruby, D.E. Oral Tecovirimat for the Treatment of Smallpox. N. Engl. J. Med. 2018, 379, 44–53, doi:10.1056/nejmoa1705688.

35. O'Shea, J.; Filardo, T.D.; Morris, S.B.; Weiser, J.; Petersen, B.; Brooks, J.T. Interim Guidance for Prevention and Treatment of Monkeypox in Persons with HIV Infection — United States, August 2022. MMWR Morb. Mortal. Wkly. Rep. 2022, 71, 1023–1028, doi:10.15585/mmwr.mm7132e4.

36. Piccolo, A.J.L.; Chan, J.; Cohen, G.M.; Mgbako, O.; Pitts, R.A.; Postelnicu, R.; Wallach, A.; Mukherjee, V. Critical Elements of an Mpox Vaccination Model at the Largest Public Health Hospital System in the United States. Vaccines (Basel) 2023, 11, 1138, doi:10.3390/vaccines11071138.

37. Beeson, A.; Styczynski, A.; Hutson, C.L.; Whitehill, F.; Angelo, K.M.; Minhaj, F.S.; Morgan, C.; Ciampaglio, K.; Reynolds, M.G.; McCollum, A.M.; et al. Mpox Respiratory Transmission: The State of the Evidence. Lancet Microbe 2023, 4, e277–e283, doi:10.1016/s2666-5247(23)00034-4.

38. CDC Detection & Transmission of Mpox Virus during the 2022 Clade IIb Out Available online: https://www.cdc.gov/poxvirus/mpox/about/science-behind-transmission.html (accessed on 31 August 2023).

39. Mohanto, S.; Faiyazuddin, M.; Dilip Gholap, A.; Jogi, D.; Bhunia, A.; Subbaram, K.; Gulzar Ahmed, M.; Nag, S.; Shabib Akhtar, M.; Bonilla-Aldana, D.K.; et al. Addressing the Resurgence of Global Monkeypox (Mpox) through Advanced Drug Delivery Platforms. Travel Med. Infect. Dis. 2023, 102636, doi:10.1016/j.tmaid.2023.102636.

40. JYNNEOS Vaccine Coverage by Jurisdiction Risk Assessment of Mpox Resurgence and Vaccination Considerations Available online: https://icpcovid.com/sites/default/files/2023-04/Ep%20327-8%20B%20Risk%20of%20Mpox%20Resurgence%20and%20Continued%20Vaccination%20Efforts%20_%20Mpox%20_%20Poxvirus%20_%20CDC.pdf (accessed on 31 August 2023).

41. Fifth Meeting of the International Health Regulations (2005) (IHR) Emergency Committee on the Multi-Country Outbreak of Mpox (Monkeypox) Available online: https://www.who.int/news/item/11-05-2023-fifth-meeting-of-the-international-health-regulations-(2005)-(ihr)-emergency-committee-on-the-multi-country-outbreak-of-monkeypox-(mpox) (accessed on 31 August 2023).

42. Health Alert Network (HAN) - 00490 Available online: https://emergency.cdc.gov/han/2023/han00490.asp (accessed on 31 August 2023).

43. Howard, J. CNN. May 15 2023,.

44. Miraz, M.H.; Ali, M.; Excell, P.S.; Picking, R. A Review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). In Proceedings of the 2015 Internet Technologies and Applications (ITA); IEEE, 2015; pp. 219–224.

45. Belle Wong, J.D. Top Social Media Statistics and Trends of 2023 Available online: https://www.forbes.com/advisor/business/social-media-statistics/ (accessed on 31 August 2023).

46. Twitter: Number of Users Worldwide 2024 Available online: https://www.statista.com/statistics/303681/twitter-users-worldwide/ (accessed on 31 August 2023).

47. Hutchinson, A. New Study Shows Twitter Is the Most Used Social Media Platform among Journalists Available online: https://www.socialmediatoday.com/news/new-study-shows-twitter-is-the-most-used-social-media-platform-among-journa/626245/ (accessed on 31 August 2023).

48. Biggest Social Media Platforms 2023 Available online: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (accessed on 31 August 2023).

49. 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, doi:10.3390/bdcc7020116.

50. Shaheer, I.; Carr, N.; Insch, A. Rallying Support for Animal Welfare on Twitter: A Tale of Four Destination Boycotts. Tourism Recreation Res. 2023, 48, 384–398, doi:10.1080/02508281.2021.1936411.

51. Abu Samah, K.A.F.; Amirah Misdan, N.F.; Hasrol Jono, M.N.H.; Riza, L.S. The Best Malaysian Airline Companies Visualization through Bilingual Twitter Sentiment Analysis: A Machine Learning Classification. JOIV Int. J. Inform. Vis. 2022, 6, 130, doi:10.30630/joiv.6.1.879.

52. Bodaghi, A.; Oliveira, J. The Theater of Fake News Spreading, Who Plays Which Role? A Study on Real Graphs of Spreading on Twitter. Expert Syst. Appl. 2022, 189, 116110, doi:10.1016/j.eswa.2021.116110.

53. Ante, L. How Elon Musk's Twitter Activity Moves Cryptocurrency Markets. Technol. Forecast. Soc. Change 2023, 186, 122112, doi:10.1016/j.techfore.2022.122112.

54. Collins, S.; DeWitt, J. Words Matter: Presidents Obama and Trump, Twitter, and U.s. Soft Power. World Aff. 2023, 186, 530–571, doi:10.1177/00438200231161631.

55. Berrocal-Gonzalo, S.; Zamora-Martínez, P.; González-Neira, A. Politainment on Twitter: Engagement in the Spanish Legislative Elections of April 2019. Media Commun. 2023, 11, 163–175, doi:10.17645/mac.v11i2.6292.

56. 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.

57. 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.

58. 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.

59. Tao, W.; Peng, Y. Differentiation and Unity: A Cross-Platform Comparison Analysis of Online Posts' Semantics of the Russian–Ukrainian War Based on Weibo and Twitter. Commun. Public 2023, 8, 105–124, doi:10.1177/20570473231165563.

60. Skovgaard, L.; Grundtvig, A. Who Tweets What about Personalised Medicine? Promises and Concerns from Twitter Discussions in Denmark. Digit. Health 2023, 9, 205520762311698, doi:10.1177/20552076231169832.

61. 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.

62. 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.

63. Cevik, F.; Kilimci, Z.H. Analysis of Parkinson's Disease Using Deep Learning and Word Embedding Models. acperpro 2019, 2, 786–797, doi:10.33793/acperpro.02.03.86.

64. Kesler, S.R.; Henneghan, A.M.; Thurman, W.; Rao, V. Identifying Themes for Assessing Cancer-Related Cognitive Impairment: Topic Modeling and Qualitative Content Analysis of Public Online Comments. JMIR Cancer 2022, 8, e34828, doi:10.2196/34828.

65. Klein, A.Z.; Kunatharaju, S.; O'Connor, K.; Gonzalez-Hernandez, G. Pregex: Rule-Based Detection and Extraction of Twitter Data in Pregnancy. J. Med. Internet Res. 2023, 25, e40569, doi:10.2196/40569.

66. Klein, A.Z.; O'Connor, K.; Levine, L.D.; Gonzalez-Hernandez, G. Using Twitter Data for Cohort Studies of Drug Safety in Pregnancy: Proof-of-Concept with β-Blockers. JMIR Form. Res. 2022, 6, e36771, doi:10.2196/36771.

67. Thackeray, R.; Burton, S.H.; Giraud-Carrier, C.; Rollins, S.; Draper, C.R. Using Twitter for Breast Cancer Prevention: An Analysis of Breast Cancer Awareness Month. BMC Cancer 2013, 13, doi:10.1186/1471-2407-13-508.

68. Russell, A.M.T.; Hing, N.; Bryden, G.M.; Thorne, H.; Rockloff, M.J.; Browne, M. Gambling Advertising on Twitter before, during and after the Initial Australian COVID-19 Lockdown. J. Behav. Addict. 2023, 12, 557–570, doi:10.1556/2006.2023.00020.

69. Gomide, J.; Veloso, A.; Meira, W., Jr; Almeida, V.; Benevenuto, F.; Ferraz, F.; Teixeira, M. Dengue Surveillance Based on a Computational Model of Spatio-Temporal Locality of Twitter. In Proceedings of the Proceedings of the 3rd International Web Science Conference; ACM: New York, NY, USA, 2011.

70. 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.

71. Signorini, A.; Segre, A.M.; Polgreen, P.M. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.s. during the Influenza A H1N1 Pandemic. PLoS One 2011, 6, e19467, doi:10.1371/journal.pone.0019467.

72. Hirschfeld, D. Twitter Data Accurately Tracked Haiti Cholera Outbreak. Nature 2012, doi:10.1038/nature.2012.9770.

73.   van der Vyver, A.G. The Listeriosis Outbreak in South Africa: A Twitter Analysis of Public Reaction Available online: http://www.icmis.net/icmis18/ICMIS18CD/pdf/S198-final.pdf (accessed on 31 August 2023).

74.   Sugumaran, R.; Voss, J. Real-Time Spatio-Temporal Analysis of West Nile Virus Using Twitter Data. In Proceedings of the Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications; ACM: New York, NY, USA, 2012.

75.   Bolotova, Y.V.; Lou, J.; Safro, I. Detecting and Monitoring Foodborne Illness Outbreaks: Twitter Communications and the 2015 U.S. Salmonella Outbreak Linked to Imported Cucumbers. arXiv [stat.AP] 2017.

76.   Porat, T.; Garaizar, P.; Ferrero, M.; Jones, H.; Ashworth, M.; Vadillo, M.A. Content and Source Analysis of Popular Tweets Following a Recent Case of Diphtheria in Spain. Eur. J. Public Health 2019, 29, 117–122, doi:10.1093/eurpub/cky144.

77.   Knudsen, B.; Høeg, T.B.; Prasad, V. Analysis of Tweets Discussing the Risk of Mpox among Children and Young People in School (May-Oct 2022): Public Health Experts on Twitter Consistently Exaggerated Risks and Infrequently Reported Accurate Information. bioRxiv 2023.

78.   Zuhanda, M.K. Analysis of Twitter User Sentiment on the Monkeypox Virus Issue Using the Nrc Lexicon Available online: https://www.iocscience.org/ejournal/index.php/mantik/article/view/3502 (accessed on 31 August 2023).

79.   Ortiz-Martínez, Y.; Sarmiento, J.; Bonilla-Aldana, D.K.; Rodríguez-Morales, A.J. Monkeypox Goes Viral: Measuring the Misinformation Outbreak on Twitter. J. Infect. Dev. Ctries. 2022, 16, 1218–1220, doi:10.3855/jidc.16907.

80.   Rahmanian, V.; Jahanbin, K.; Jokar, M. Using Twitter and Web News Mining to Predict the Monkeypox Outbreak. Asian Pac. J. Trop. Med. 2022, 15, 236, doi:10.4103/1995-7645.346083.

81.   Cooper, L.N.; Radunsky, A.P.; Hanna, J.J.; Most, Z.M.; Perl, T.M.; Lehmann, C.U.; Medford, R.J. Analyzing an Emerging Pandemic on Twitter: Monkeypox. Open Forum Infect. Dis. 2023, 10, doi:10.1093/ofid/ofad142.

82.   Ng, Q.X.; Yau, C.E.; Lim, Y.L.; Wong, L.K.T.; Liew, T.M. Public Sentiment on the Global Outbreak of Monkeypox: An Unsupervised Machine Learning Analysis of 352,182 Twitter Posts. Public Health 2022, 213, 1–4, doi:10.1016/j.puhe.2022.09.008.

83.   Bengesi, S.; Oladunni, T.; Olusegun, R.; Audu, H. A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion from Twitter Tweets. IEEE Access 2023, 11, 11811–11826, doi:10.1109/access.2023.3242290.

84.   Olusegun, R.; Oladunni, T.; Audu, H.; Houkpati, Y.A.O.; Bengesi, S. Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach. IEEE Access 2023, 11, 49882–49894, doi:10.1109/access.2023.3277868.

85.   Farahat, R.A.; Yassin, M.A.; Al-Tawfiq, J.A.; Bejan, C.A.; Abdelazeem, B. Public Perspectives of Monkeypox in Twitter: A Social Media Analysis Using Machine Learning. New Microbes New Infect. 2022, 49–50, 101053, doi:10.1016/j.nmni.2022.101053.

86.   Sv, P.; Ittamalla, R. What Concerns the General Public the Most about Monkeypox Virus? – A Text Analytics Study Based on Natural Language Processing (NLP). Travel Med. Infect. Dis. 2022, 49, 102404, doi:10.1016/j.tmaid.2022.102404.

87.   Mohbey, K.K.; Meena, G.; Kumar, S.; Lokesh, K. A CNN-LSTM-Based Hybrid Deep Learning Approach to Detect Sentiment Polarities on Monkeypox Tweets. arXiv [cs.CV] 2022.

88.   Nia, Z.M.; Bragazzi, N.L.; Wu, J.; Kong, J.D. A Twitter Dataset for Monkeypox, May 2022. Data Brief 2023, 48, 109118, doi:10.1016/j.dib.2023.109118.

89.   Iparraguirre-Villanueva, O.; Alvarez-Risco, A.; Herrera Salazar, J.L.; Beltozar-Clemente, S.; Zapata-Paulini, J.; Yáñez, J.A.; Cabanillas-Carbonell, M. The Public Health Contribution of Sentiment Analysis of Monkeypox Tweets to Detect Polarities Using the CNN-LSTM Model. Vaccines (Basel) 2023, 11, 312, doi:10.3390/vaccines11020312.

90.   AL-Ahdal, T.; Coker, D.; Awad, H.; Reda, A.; Żuratyński, P.; Khailaie, S. Improving Public Health Policy by Comparing the Public Response during the Start of COVID-19 and Monkeypox on Twitter in Germany: A Mixed Methods Study. Vaccines (Basel) 2022, 10, 1985, doi:10.3390/vaccines10121985.

91. 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.

92. Hofmann, M.; Klinkenberg, R. RapidMiner: Data Mining Use Cases and Business Analytics Applications; CRC Press: Boca Raton, FL, 2016; ISBN 9781498759861.

93. Thakur, N.; Han, C.Y. Multimodal Approaches for Indoor Localization for Ambient Assisted Living in Smart Homes. Information (Basel) 2021, 12, 114, doi:10.3390/info12030114.

94. Dwivedi, S.; Kasliwal, P.; Soni, S. Comprehensive Study of Data Analytics Tools (RapidMiner, Weka, R Tool, Knime). In Proceedings of the 2016 Symposium on Colossal Data Analysis and Networking (CDAN); IEEE, 2016; pp. 1–8.

95. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; Zhao, L. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. Multimed. Tools Appl. 2019, 78, 15169–15211, doi:10.1007/s11042-018-6894-4.

96. Wei, X.; Croft, W.B. LDA-Based Document Models for Ad-Hoc Retrieval. In Proceedings of the Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval; ACM: New York, NY, USA, 2006.

97. Yao, L.; Mimno, D.; McCallum, A. Efficient Methods for Topic Model Inference on Streaming Document Collections. In Proceedings of the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining; ACM: New York, NY, USA, 2009.

98. 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.

99. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. Sci. Data 2016, 3, 1–9, doi:10.1038/sdata.2016.18.

100. Syed, S.; Spruit, M. Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA); IEEE, 2017; pp. 165–174.

101. Omar, M.; On, B.-W.; Lee, I.; Choi, G.S. LDA Topics: Representation and Evaluation. J. Inf. Sci. 2015, 41, 662–675, doi:10.1177/0165551515587839.

102. Amoualian, H.; Lu, W.; Gaussier, E.; Balikas, G.; Amini, M.R.; Clausel, M. Topical Coherence in LDA-Based Models through Induced Segmentation. In Proceedings of the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 1799–1809.

103. Ajinaja, M.O.; Adetunmbi, A.O.; Ugwu, C.C.; Popoola, O.S. Semantic Similarity Measure for Topic Modeling Using Latent Dirichlet Allocation and Collapsed Gibbs Sampling. Iran J. Comput. Sci. 2023, 6, 81–94, doi:10.1007/s42044-022-00124-7.

104. Xue, J.; Chen, J.; Chen, C.; Zheng, C.; Li, S.; Zhu, T. Public Discourse and Sentiment during the COVID 19 Pandemic: Using Latent Dirichlet Allocation for Topic Modeling on Twitter. PLoS One 2020, 15, e0239441, doi:10.1371/journal.pone.0239441.