

# A Large-Scale Dataset of Twitter Chatter about Online Learning during the Current COVID-19 Omicron Wave

Nirmalya Thakur 

Department of Electrical Engineering and Computer Science, University of Cincinnati,  
Cincinnati, OH 45221-0030, USA; thakurna@mail.uc.edu

**Abstract:** The COVID-19 Omicron variant, reported to be the most immune-evasive variant of COVID-19, is resulting in a surge of COVID-19 cases globally. This has caused schools, colleges, and universities in different parts of the world to transition to online learning. As a result, social media platforms such as Twitter are seeing an increase in conversations related to online learning in the form of tweets. Mining such tweets to develop a dataset can serve as a data resource for different applications and use-cases related to the analysis of interest, views, opinions, perspectives, attitudes, and feedback towards online learning during the current surge of COVID-19 cases caused by the Omicron variant. Therefore, this work presents a large-scale, open-access Twitter dataset of conversations about online learning from different parts of the world since the first detected case of the COVID-19 Omicron variant in November 2021. The dataset is compliant with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter, as well as with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management. The paper also briefly outlines some potential applications in the fields of Big Data, Data Mining, Natural Language Processing, and their related disciplines, with a specific focus on online learning during this Omicron wave that may be studied, explored, and investigated by using this dataset.



**Citation:** Thakur, N. A Large-Scale Dataset of Twitter Chatter about Online Learning during the Current COVID-19 Omicron Wave. *Data* **2022**, *7*, 109. <https://doi.org/10.3390/data7080109>

Academic Editors: Antonio Sarasa Cabezuelo and Ramón González del Campo Rodríguez Barbero

Received: 9 June 2022

Accepted: 2 August 2022

Published: 4 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Dataset:** <https://doi.org/10.5281/zenodo.6837118>

**Dataset License:** CC-BY 4.0

**Keywords:** COVID-19; COVID; omicron; online learning; remote learning; online education; Twitter; dataset; tweets; social media; big data

## 1. Introduction

The first cases of the COVID-19 pandemic, caused by the SARS-CoV-2 virus, were recorded in a seafood market in Wuhan, China, in December 2019 [1]. Since then, the virus has been found in all the countries of the world. At the time of writing this paper, globally, there have been 535,342,382 cases with 6,320,324 deaths [2]. Since the initial cases in China, the SARS-CoV-2 virus has undergone multiple mutations, and as a result, multiple variants have been detected in different parts of the world. Some of these include: Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Epsilon (B.1.427 B.1.429), Eta (B.1.525), Iota (B.1.526), Kappa (B.1.617.1), Zeta (P.2), Mu (B.1.621, B.1.621.1), and Omicron (B.1.1.529, BA.1, BA.1.1, BA.2, BA.3, BA.4 and BA.5) [3]. Out of all these variants, the Omicron variant, first detected on 24 November 2021 from a sample collected on 9 November 2021, was classified as a Variant of Concern (VOC) by the World Health Organization (WHO) on 26 November 2021 [4]. The Omicron variant has a spike protein that contains 30 mutations [5]. It has been reported to be the most immune-evasive variant of COVID-19 and to present very strong resistance against antibody-based or plasma-based treatments [6]. According to WHO, the new cases due to this Omicron variant have been “off the charts”

and are setting new records in terms of COVID-19 cases all over the world [7]. The Omicron variant currently accounts for 86% of the COVID-19 cases worldwide [8], and some of the countries that have recorded the most cases due to the SARS-CoV-2 Omicron variant include the United Kingdom (1,138,814 cases), USA (945,470 cases), Germany (245,120 cases), Denmark (218,106 cases), France (110,959 cases), Canada (92,341 cases), Japan (71,056 cases), India (56,125 cases), Australia (46,576 cases), Sweden (43,400 cases), Israel (39,908 cases), Poland (33,436 cases), and Brazil (32,880 cases) [9].

Since the beginning of the pandemic, many countries, such as India [10], the United States [11], the United Kingdom [12], Spain [13], Greece [14], Italy [15], Austria [16], Nigeria [17], China [18], New Zealand [19], Ireland [20], Germany [21], South Africa [22], Australia [23], France, [24], Norway [25], and several more [26], went on a complete lockdown with work from home and remote work guidelines that affected a multitude of industries and sectors. Out of all these sectors that were impacted by the nationwide lockdowns and the associated guidelines in different parts of the world, the education sector was an important one. On a global scale, universities, colleges, and schools had to switch to online education, which required its faculty, administrators, staff, and students to become familiarized with online learning and the associated tools and platforms that were necessary for this new norm of education. Due to the worldwide adoption and familiarization with various forms of tools, platforms, software, and hardware necessary for online education, the online education market is rapidly booming and is expected to reach more than USD 350 billion by 2025 [27]. Online learning may be broadly defined as *“learning experiences in synchronous or asynchronous environments using different devices (e.g., mobile phones, laptops, etc.) with internet access. In these environments, students can be anywhere (independent) to learn and interact with instructors and other students”* [28]. Online learning has a range of synonyms, and some of the most commonly used synonyms include remote education, online education, virtual education, remote learning, e-learning, distance education, virtual learning, asynchronous learning, and blended learning [28].

On a global scale, more than 43,518,726 students were affected due to in-person school closures due to COVID-19 [29]. The closing of universities, colleges, and schools was recorded in 188 countries [30], and 90% of the countries reported a switch to one or more forms of online learning [31]. Despite these promising numbers, 31% (463 million) of students in schools (in preprimary to secondary education) could not adopt online learning either due to lack of technologies, training, or accessibility, and 75% of students who belonged to the poorest households could not switch to the technologies required for online learning [31].

With the advancements in vaccine research and other forms of treatment of COVID-19 toward the later part of 2020 [32–34] and in compliance with the recommendations from various local and national policy-making bodies, different universities, colleges, and schools started to transition to hybrid (both online and in-person) learning as well as completely in-person learning [35]. However, this was associated with several challenges [36], including a surge of COVID-19 cases in students, educators, and staff members, an increase in stress and anxiety in both students and their parents, the need for allocation of funds by these educational institutions to conduct classes in a socially distant manner, and for procurement of hand sanitizers and disinfectants. Despite these challenges, education continued in both hybrid and in-person forms for a few months. However, due to the recent global surge in COVID-19 cases due to the Omicron variant [7–9], many educational institutions all over the world have transitioned back to online learning since the beginning of 2022, and several are in the process of transitioning to online learning over the next few months [37–42].

The modern-day Internet of Everything lifestyle [43] is characterized by people spending more time on the internet than ever before, with a specific focus on social media platforms. The use of social media platforms has skyrocketed in the recent past [44]. Social media usage characteristics include conversations on diverse topics such as recent issues, global challenges, emerging technologies, news, current events, politics, family, relationships, and career opportunities [45]. Twitter, one such social media platform, used by

people of almost all age groups [46,47], has been rapidly gaining popularity in all parts of the world and is currently the second most visited social media platform [48]. At present, there are about 192 million daily active users on Twitter, and approximately 500 million tweets are posted on Twitter every day [49]. Mining of social media conversations, such as Tweets, to develop datasets has been of significant interest to the scientific community in the areas of Big Data, Data Mining, and Natural Language Processing, as can be seen from these recent works where relevant Tweets were mined to develop Twitter datasets on the 2020 US Presidential Election [50], 2022 Russia–Ukraine war [51], climate change [52], natural hazards [53], European Migration Crisis [54], movies [55], toxic behavior amongst adolescents [56], music [57], civil unrest [58], drug safety [59], and Inflammatory Bowel Disease [60].

In the context of the recent surge of COVID-19 cases due to the Omicron variant and its impact on the education sector, there has been a significant increase in conversations on Twitter related to online learning. Mining such conversations to develop a dataset would serve as a rich data resource for the investigation of different research questions in the fields of Big Data, Data Mining, Data Science, and Natural Language Processing, with a central focus on analyzing tweets related to online learning during this time.

Previous works [61–90] (discussed in Section 2) related to online learning since the outbreak of COVID-19 have focused on analyzing multiple factors related to online learning only in certain geographic regions, mostly by using surveys, and not on a global scale by analyzing conversations from all over the world, such as Tweets. Prior works on the development of Twitter datasets related to COVID-19 have also not focused on mining relevant tweets related to online learning during the ongoing COVID-19 Omicron wave. To address these limitations, this work proposes a dataset of more than 50,000 Tweet IDs (that correspond to the same number of Tweets) about online learning that was posted on Twitter from 9 November 2021 to 13 July 2022, which is publicly available at <https://doi.org/10.5281/zenodo.6837118>. The earliest date was selected as 9 November 2021, as the Omicron variant was detected for the first time in a sample that was collected on this date. The most recent date, at the time of resubmission of this journal paper after the completion of the first round of peer review and the subsequent editorial decision, was 13 July 2022.

The rest of the paper is organized as follows. Section 2 presents an overview of recent works in this field. The methodology that was followed for the development of this dataset is presented in Section 3. Section 4 provides the description of the dataset. Section 5 briefly discusses a few potential applications of this dataset. The conclusion and scope for future work are presented in Section 6, which is followed by references.

## 2. Literature Review

There has been a significant amount of research related to online learning since the global outbreak of COVID-19. The work by Muhammad et al. [61] was a research study that examined the attitudes of Pakistani higher education students toward compulsory digital and distance learning courses during COVID-19. In [62], Rasmitadila et al. presented a study that explored the perceptions of primary school teachers towards online learning during COVID-19. Data were collected through surveys and semi-structured interviews, and 67 teachers in primary schools participated in this study. The work by Irawan et al. [63] aimed to identify the impact of student psychology on online learning during the COVID-19 pandemic. The research method used a qualitative research type of phenomenology. The research subjects were 30 students of Mulawarman University, a university in Indonesia, who were interviewed via telephone. The work of Baticulon et al. [64] was to identify barriers to online learning from the perspective of medical students in the Philippines. The authors sent out an electronic survey to the students who participated in this study. The qualitative study presented by Hussein et al. [65] aimed to investigate the attitudes of undergraduate students towards online learning during the first few weeks of the mandatory shift to online learning caused by COVID-19. Students from two general

English courses at a university located in the United Arab Emirates were asked to write semi-guided essays and the associated data were analyzed by the authors. The work of Famularsih et al. [66] focused on studying the utilization of online learning applications in English as a Foreign Language (EFL) classrooms. The participants of this study were 35 students from a university in Salatiga, Indonesia. The data were gathered through surveys and semi-structured interviews.

The study by Sutarto et al. [67] focused on understanding the strategies used by teachers of SDIT Rabbi Radhiyya Curup, a school in Indonesia, to increase students' interest and responses to online learning during COVID-19. The data were collected by conducting semi-structured interviews, which were analyzed using the Miles and Huberman model. Almusharraf et al.'s [68] work aimed to evaluate the level of postsecondary student satisfaction with online learning platforms and learning experiences during the COVID-19 pandemic in Saudi Arabia. Quantitative research was carried out in this study by using a survey that was sent out to 283 students enrolled at a higher education institution in Saudi Arabia. These data were analyzed using SPSS. Al-Salman et al. [69] investigated the influence of digital technology, instructional and assessment quality, economic status and psychological state, and course type on Jordanian university students' attitudes towards online learning during the COVID-19 emergency transition to online learning. A total of 4037 undergraduate students from four universities participated in this study.

The aim of Bolatov et al.'s work [70] was to compare the differences between the mental state of students switching to online learning and the mental state of the students who were still using traditional learning. This study included medical students from Astana Medical University, a university in Kazakhstan. The work by Agormedah et al. [71] explored the responses of students to online learning in higher education in Ghana. The sample size of this study involved 467 students. The findings indicated that a majority of the students had a positive response to the transition to online learning. The work of Moawad et al. [72] aimed to identify the academic stressors by analyzing the worries and fears that students at the College of Education in King Saud University, a university in Saudi Arabia, experienced during the time of COVID-19. The results showed that the issue with the highest percentage of stress among students was their uncertainty over the end-of-semester exams and assessments. The work by Khan et al. [73] discussed various digital education methods, approaches, and systems that could be implemented by the education system of Bangladesh during COVID-19. The purpose of the study performed by Catalano et al. [74] was to determine teacher perceptions of students' access and participation in online learning, as well as concerns about educational outcomes among different groups of learners. The work of Kapasia et al. [75] aimed to assess the impact of the nationwide lockdown on account of COVID-19 on undergraduate and postgraduate students in West Bengal, a state in India. The authors conducted an online survey that included 232 students. In [76], Burns et al. performed a conceptual analysis on student wellbeing at universities in the United Kingdom with a specific focus on the psychosocial impact the pandemic had on students. Küsel et al. [77] performed a study to evaluate German university students' readiness for using digital media and online learning in their tertiary education and compared the findings with the results from the same study performed on students in the United States. A total of 72 students from universities in Germany and 176 students from universities in the United States were a part of this study. Darayseh et al. [78] analyzed the impact of COVID-19 on modes of teaching, with a specific focus on science education in schools in the United Arab Emirates. Questionnaires were deployed through an online platform, and a total of 62 science teachers participated in this study. Tsekhmister et al. [79] conducted a study to evaluate the effectiveness of virtual reality technology and online teaching systems among medical students of Bogomolets National Medical University, a university in Ukraine. The study was performed using a questionnaire that contained 15 questions with five options to comprehensively evaluate these technologies.

Arsaliev et al.'s work [80] aimed to investigate whether an online format was effective in providing education for ethnocultural competence development. A combination of

digital surveys, tests, questionnaires, and online class interviews were used in this study that involved 120 students at Southern Federal University, a university in Russia. Cárdenas-Cruz et al.'s [81] work aimed to facilitate the acquisition of specific transversal skills of undergraduate students at the University of Granada in Spain during the outbreak by means of an integrated online working system. Papouli et al. [82] aimed to explore Greek social-work students' views on the use of digital technology during their stay at home due to the coronavirus lockdown. A total of 550 students from different universities in Greece participated in this study. In [83], Parmigiani et al. designed a qualitative study aimed at investigating the factors affecting e-inclusion during COVID-19. A total of 785 teachers at the University of Genoa, a university in Italy, participated in this study. Resch et al. [84] focused on analyzing the effects of COVID-19 on university students' social and academic integration, based on Tinto's integration theory. A total of 640 university students in Austria completed an online survey pertaining to academic and social integration in this study. The purpose of the study by Noah et al. [85] was to examine the impacts of Google classroom as an online learning delivery platform in a secondary school during the COVID-19 pandemic in Nigeria. The study included 140 participants. Chen et al. [86] studied user satisfaction in the context of using online education platforms in China during COVID-19. The work used a combination of questionnaires and a back propagation neural network.

Drane et al. [87] performed a comprehensive review of existing works to present the impact of 'learning at home' on the educational outcomes of vulnerable children in Australia during the COVID-19 pandemic. The work of Mukuna et al. [88] explored the perceived challenges of online teaching encountered by educators in a school in the Thabo Mofutsanyana District in South Africa. A total of six educators participated in this study. In [89], Hsiao presented the results of a study to explore the influences of course type and gender on distance learning performance. A total of 18,085 students from a university in Taiwan comprised the sample size of this study. Nafrees et al. [90] performed an analysis to determine the factors of awareness of students about online learning among undergraduate students at Southeastern University, a university in Sri Lanka. The study comprised about 400 questionnaires, and a total of 310 responses from students were analyzed by the authors. The findings showed that most students preferred to use WebEx over other platforms for their online education due to the user-friendliness of WebEx.

In terms of mining relevant conversations related to a specific topic on Twitter since the outbreak of COVID-19, the prior works in this field have focused on the development of datasets for healthcare misinformation [91], misleading information [92], vaccine misinformation [93], patient identification [94], updates related to vaccine development [95], and rumors related to COVID-19 [96].

Despite these emerging works in the fields of online learning and the development of Twitter datasets, there exist multiple limitations. First, these works in the field of online learning have been confined to studying or analyzing the success or failure, degrees of acceptance, and associated factors related to online learning in specific geographic regions in countries such as Pakistan [61], Indonesia [62,63,66,67], Philippines [64], UAE [65], Saudi Arabia [68,72], Jordan [69], Kazakhstan [70], Ghana [71], Bangladesh [73], the United States [74,77], India [75], the United Kingdom [76], Germany [77], the UAE [78], Ukraine [79], Russia [80], Spain [81], Greece [82], Italy [83], Austria [84], Nigeria [85], China [86], Australia [87], South Africa [88], Taiwan [89], and Sri Lanka [90], and not on a global level. Second, due to the lack of datasets such as Twitter conversations related to online learning from global users, the data that were analyzed in these studies were mostly in the form of surveys that were conducted in these respective geographic regions. Third, the Twitter datasets related to COVID-19 [91–96] do not focus on online learning and the ongoing chatter on Twitter about the same amidst the global rise of COVID-19 cases due to the Omicron variant. The dataset proposed in this paper aims to address all these limitations.

### 3. Methodology

This section describes the methodology that was followed for the development of this dataset, which is available at <https://doi.org/10.5281/zenodo.6837118>. The dataset contains a total of 52,984 Tweet IDs that correspond to the same number of tweets about online learning, which were publicly posted on Twitter from 9 November 2021 to 13 July 2022. This section also outlines how this work and the associated dataset development is in compliance with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter, as well as follows the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management. These are discussed in Sections 3.1–3.3, respectively.

#### 3.1. Process for Dataset Development

As this work focuses on developing a Twitter dataset, the privacy policy, developer agreement, and guidelines for content redistribution of Twitter [97,98] were thoroughly studied, and after studying the same, it was concluded that mining relevant tweets from Twitter to develop a dataset (comprising only Tweet IDs) is in compliance with all these policies of Twitter. Therefore, this dataset contains only Tweet IDs and does not contain any other information related to the respective Tweets that were mined. A detailed explanation of this compliance is mentioned in Section 3.2.

The tweets were collected by using the Search Twitter “operator” [99] available in RapidMiner studio [100] and the Advanced Search feature of the Twitter API. RapidMiner is a data science platform that allows the development, implementation, and testing of various algorithms, processes, and applications in the fields of Big Data, Data Mining, Data Science, Artificial Intelligence, Machine Learning, and their related areas. There are various RapidMiner products available such as RapidMiner Studio, RapidMiner AI Hub, and RapidMiner Radoop. For this work, the RapidMiner studio, version 9.10, was downloaded and installed on a laptop with the Microsoft Windows 10 Home operating system with Intel (R) Pentium (R) Silver N5030 CPU @ 1.10GHz, 1101 Mhz, 4 Core (s), and 4 Logical Processor (s). In the RapidMiner platform, “process” and “operator” are two commonly used terminologies. An “operator” represents a specific function or operation, for instance, fetch data from a social media platform such as Twitter based on a specific set of guidelines or to perform a specific operation on a dataset. RapidMiner has a number of in-built “operators”. It also allows users to develop “operators” from scratch. A collection of “operators” that are connected in a logical and executable sequence to achieve a desired purpose is called a “process”. A “process” may also contain just one “operator” if the complete functionality of the “process” can be found in one in-built or user-defined “operator”. The Search Twitter “operator”, an in-built “operator” of RapidMiner, works by connecting with the Twitter API and by complying with the Twitter API standard search policies [101] to fetch tweets between two given dates that contain one or more keywords or phrases which are provided as input to this “operator”. As there are different keywords that Twitter users can use to refer to both COVID-19, the Omicron variant, and online learning; therefore, a bag of words was developed based on studying commonly used synonyms, phrases, and terms used to refer to online learning [102], COVID-19 and the Omicron variant [103]. These synonyms, terms, and phrases, all of which were included in the data collection process, are shown in Table 1.

**Table 1.** List of commonly used synonyms, terms, and phrases for online learning and COVID-19.

Terminology	List of Synonyms and Terms
COVID-19	Omicron, COVID, COVID19, coronavirus, coronavirus pandemic, COVID-19, corona, corona outbreak, omicron variant, SARS-CoV-2, corona virus
online learning	online education, online learning, remote education, remote learning, e-learning, elearning, distance learning, distance education, virtual learning, virtual education, online teaching, remote teaching, virtual teaching, online class, online classes, remote class, remote classes, distance class, distance classes, virtual class, virtual classes, online course, online courses, remote course, remote courses, distance course, distance courses, virtual course, virtual courses, online school, virtual school, remote school, online college, online university, virtual college, virtual university, remote college, remote university, online lecture, virtual lecture, remote lecture, online lectures, virtual lectures, remote lectures

There are various forms of educational structures and educational systems followed by different countries all over the world. For instance, in the United States, early childhood education is followed by primary school (also called elementary school), middle school, secondary school (also called high school), and then postsecondary (tertiary) education. Postsecondary education includes nondegree programs that lead to certificates and diplomas plus six degree levels: associate, bachelor, first professional, master, advanced intermediate, and research doctorate. The US system does not offer a second or higher doctorate but does offer postdoctorate research programs [104]. A different educational structure is followed in India [105]. The school system in India has four levels: lower primary school (age 6 to 10), upper primary school (age 11 and 12), high school (age 13 to 15), and higher secondary school (age 17 and 18). The lower primary school is divided into five “standards”, upper primary school into two, high school into three, and higher secondary school into two. Another different educational structure can be seen in the United Kingdom (UK). The education system in the UK is divided into four main parts, primary education, secondary education, further education, and higher education. Children in the UK have to legally attend primary and secondary education, which runs from about five years old until the student is 16 years old. The education system in the UK is also split into “key stages”: Key Stage 1 (age 5 to 7), Key Stage 2 (age 7 to 11), Key Stage 3 (age 11 to 14), and Key Stage 4 (age 14 to 16) [106]. This study focuses on collecting tweets about online education or online learning on a global scale (and not tweets originating from any specific country specific to its educational structure or educational system). So, a comprehensive list of keywords (as shown in Table 1) was developed that would most commonly be used to refer to online education or online learning in different parts of the world, irrespective of the educational structure followed in that specific geographic region. The effectiveness of this approach can be seen from the different worldwide educational systems that are the subject matters of the tweets present in the dataset proposed as a result of this work. For instance, in this dataset, Tweet ID: 1458685065152450565 refers to online education in India; Tweet ID: 1462489169079513090 refers to online education in the United States; Tweet ID: 1462475208644874242 refers to online education in Pakistan; Tweet ID: 1462373712389238787 refers to online education in Indonesia; Tweet ID: refers to online education in the UK; Tweet ID: 1462357217479434241 refers to online education in Ukraine; Tweet ID: 1462512737402109952 refers to online education in Nigeria; Tweet ID: 1462315144411856897 refers to online education in Spain; Tweet ID: 1462411445035941891 refers to online education in Malaysia, and so on.

Tweets were searched using this “process” that comprised the Search Twitter “operator” in a way that it consisted of at least one synonym, term, or phrase used to refer to COVID-19 and at least one synonym, term, or phrase used to refer to online learning. The Search Twitter “operator” is not case-sensitive, so it returned the tweets based on keyword matching by ignoring the case (uppercase or lowercase).

The output of this RapidMiner “process” comprised multiple attributes such as the Tweet ID, Tweet Source (the source used to post the Tweet such as Twitter for Android, Twitter for IOS, etc.), Text of the Tweet, Retweet count, and the username of the Twitter user who posted the Tweet, all of which is public information that can be mined in compliance with the guidelines set forth in the Twitter API standard search policies. However, as per the developer policy, privacy policy, and content redistribution guidelines of Twitter, all the attributes other than the Tweet IDs were deleted by using data filters. Therefore, the dataset consists of only Tweet IDs. These Tweet IDs were grouped into different .txt files based on the timeline of the associated tweets. The description and details of these dataset files are presented in Section 4.

The complete information associated with a tweet, such as the text of a tweet, username, user ID, timestamp, retweet count, etc., can be obtained from a Tweet ID by following a process known as hydration of Tweet ID [107]. Researchers in the field of Big Data, Data Mining, and Natural Language Processing, with a specific focus on Twitter research, have developed multiple tools for the hydration of Tweet IDs. Some of the most commonly used tools include the Hydrator app [108], Social Media Mining Toolkit [109], and Twarc [110], all of which work by complying with the policies of accessing the Twitter API. Any of these tools can be used on this dataset to obtain the associated information, such as the text of a tweet, username, user ID, timestamp, and retweet count for all the Tweet IDs. A step-by-step process on how to use one of these tools, the Hydrator app, for hydrating all the Tweet IDs in this dataset is mentioned in Appendix A.

A couple of things are worth mentioning here. First, Twitter allows users the option to delete a tweet, which would mean that there would be no retrievable Tweet text and other related information (upon hydration) for a Tweet ID of a deleted tweet. All the Tweet IDs available in this dataset correspond to tweets that have not been deleted at the time of writing this paper. Second, the Twitter API’s search feature does not return an exhaustive list of tweets that were posted in a specific date range. So, it is possible that multiple tweets that might have been posted in between this date range were not returned by the Twitter API’s search feature when the data collection was performed and are thus not a part of this dataset.

### 3.2. Compliance with Twitter Policies

The privacy policy of Twitter [97] states “*Twitter is public and Tweets are immediately viewable and searchable by anyone around the world*”. To add, the Twitter developer agreement [98] defines tweets as “*public data*”. The guidelines for Twitter content redistribution [98] state “*If you provide Twitter Content to third parties, including downloadable datasets or via an API, you may only distribute Tweet IDs, Direct Message IDs, and/or User IDs (except as described below)*”. It also states “*We also grant special permissions to academic researchers sharing Tweet IDs and User IDs for non-commercial research purposes. Academic researchers are permitted to distribute an unlimited number of Tweet IDs and/or User IDs if they are doing so on behalf of an academic institution and for the sole purpose of non-commercial research*”. Therefore, it may be concluded that mining relevant tweets from Twitter to develop a dataset (comprising only Tweet IDs) and to share the same is in compliance with the privacy policy, developer agreement, and content redistribution guidelines of Twitter.

### 3.3. Compliance with FAIR

This section outlines how this dataset is compliant with the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management [111]. The dataset is findable, as it has a unique and permanent DOI, which was assigned by Zenodo. The dataset is accessible online. It is interoperable due to the use of .txt files for data representation that can be downloaded, read, and analyzed across different computer systems and applications. The dataset is reusable as the associated tweets and related information, such as user ID, username, retweet count, etc., for all the Tweet IDs, can be

obtained by the process of hydration in compliance with Twitter policies (Appendix A) for data analysis and interpretation.

#### 4. Data Description

This section provides a detailed description of this dataset. The raw version of the dataset comprised 67,319 tweets. This included multiple duplicate tweets. The duplicate tweets were recorded mostly because several Twitter users used a list of different hashtags referring to either online learning and/or the Omicron variant of COVID-19 in the same tweet, probably for increased audience engagement. For instance, as per the methodology described in Section 3, Tweet ID: 1464533235367510019 was captured twice, as it contains two synonyms (“omicron” and “covid”) from the list of synonyms presented in Table 1. Therefore, after the data collection process was completed as described in Section 3, data preprocessing and data cleaning were performed using RapidMiner to remove duplicate tweets. After the removal of duplicate tweets, the dataset comprised 52,984 Tweet IDs corresponding to the same number of tweets about online learning posted on Twitter between 9 November 2021 (the sample collected on this date was the first case of Omicron) to 13 July 2022 (the most recent date at the time of resubmission of this paper to this journal after the completion of the first round of peer review and the subsequent editorial decision). The dataset is available at <https://doi.org/10.5281/zenodo.6837118>. The dataset comprises nine .txt files. Table 2 presents the description of each of these dataset files along with the number of Tweet IDs present in each of them. As can be seen from Table 2, the greatest number of Tweets were posted in January 2022. The fact that the tweets of only 13 days in July 2022 were mined is the likely reason why July 2022 accounts for the least number of tweets as per this table.

**Table 2.** Description of all the files present in this dataset that comprises tweets about online learning during the current COVID-19 Omicron Wave.

Filename	No. of Tweet IDs	Date Range of the Associated Tweets
TweetIDs_November_2021.txt	1283	1 November 2021 to 30 November 2021
TweetIDs_December_2021.txt	10,545	1 December 2021 to 31 December 2021
TweetIDs_January_2022.txt	23,078	1 January 2022 to 31 January 2022
TweetIDs_February_2022.txt	4751	1 February 2022 to 28 February 2022
TweetIDs_March_2022.txt	3434	1 March 2022 to 31 March 2022
TweetIDs_April_2022.txt	3355	1 April 2022 to 30 April 2022
TweetIDs_May_2022.txt	3120	1 May 2022 to 31 May 2022
TweetIDs_June_2022.txt	2361	1 June 2022 to 30 June 2022
TweetIDs_July_2022.txt	1057	1 June 2022 to 13 July 2022

Table 3 presents some characteristic features of this dataset. As can be seen from Table 3, the tweets are present in 34 different languages in this dataset. The most common language is English (50,539 Tweets), which is followed by Indonesian (527 Tweets), Tagalog (525 Tweets), Estonian (364 Tweets), Spanish (236 Tweets), Hindi (179 Tweets), and 28 other languages. All these tweets were posted on 237 different days between 9 November 2021 and 13 July 2022. The highest number of Tweets was recorded on 5 January 2022 (2067 Tweets), which is followed by 6 January 2022 (1592 Tweets), 3 January 2022 (1465 Tweets), 4 January 2022 (1355 Tweets), and the other dates. A total of 17,950 distinct Twitter users posted these tweets, who have a total follower count of 4,345,192,697. The combined favorite count and retweet count of all the tweets present in this dataset are 3,273,263 and 556,980, respectively. A total of 5722 Tweets present in this dataset were posted by Twitter users with a verified Twitter account, and the remaining Tweets came from an unverified Twitter account. The number of distinct URLs that can be found embedded in these Tweets is 7869. The URL that occurs the greatest number of times (30 times) in the Tweets points to a list of online courses for COVID-19 safety at work [112]. The URL

that occurs the second greatest number of times (29 times) is a YouTube video that is also an online course on COVID-19 [113].

**Table 3.** Characteristic features of this dataset that comprises tweets about online learning during the current COVID-19 Omicron Wave.

Characteristic Feature	Count
Languages in which the Tweets are available	34
Distinct days when the Tweets were posted	237
Distinct users who posted the Tweets	17,950
Total follower count of all the Twitter users who posted the Tweets	4,345,192,697
Number of Tweets from a verified Twitter account	5722
Number of Tweets from an unverified Twitter account	47,262
Total favorite count of all the Tweets	3,273,263
Total retweet count of all the Tweets	556,980
Distinct URLs embedded in the Tweets	7869

## 5. Potential Applications: Brief Overview

This dataset of more than 50,000 Tweet IDs is expected to help advance interdisciplinary research in different fields such as Big Data, Data Science, Data Mining, Natural Language Processing, Healthcare, and their related disciplines. A few potential applications and use-case scenarios that may be investigated using this dataset include performing sentiment analysis [114], performing aspect-based sentiment analysis [115], predicting popular tweets [116], detecting sarcasm [117], developing topic modeling [118], tracking retweeting patterns [119], ranking tweets [120], performing content value analysis [121], tracking credibility of information [122], detecting conspiracy theories [123], predicting emoji usage patterns [124], studying the relevance of information [125], detecting satire [126], detecting deception [127], extracting categorical topics and emerging issues [128], characterizing Twitter users [129], and detection of Twitter user demographics [130] in the context of Twitter chatter related to online learning during the current Omicron wave of COVID-19.

## 6. Conclusions

The outbreak of COVID-19 led to schools, colleges, and universities in almost all parts of the world closing and transitioning to online learning. The development of vaccines and other forms of treatment towards the end of 2020 led to some of these educational institutions reopening and starting to function in a hybrid as well as in a completely in-person manner. The recent surge of COVID-19 cases globally due to the Omicron variant, the most immune-evasive variant of COVID-19 that presents very strong resistance against antibody-based or plasma-based treatments, has resulted in several such educational institutions switching to online learning once again. This has led to an increase in the number of online conversations, specifically on Twitter, related to online learning since the first detected case of the Omicron variant in November 2021. Mining such tweets to develop a dataset would serve as a data resource for interdisciplinary research related to the analysis of interest, views, opinions, perspectives, attitudes, and feedback towards online learning during the current surge of COVID-19 cases caused due to this variant. The prior works in this field did not focus on the development of a similar data resource. Therefore, this work presents an open-access dataset of more than 50,000 Tweet IDs (that correspond to the same number of tweets) about online learning posted on Twitter between 9 November 2021 (the sample collected on this date was the first case of Omicron) and 13 July 2022 (the most recent date at the time of resubmission of this journal paper after the completion of the first round of peer review and the subsequent editorial decision). The dataset is compliant with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter, as well as with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management. The paper also briefly outlines a

few potential research directions that may be investigated using this dataset. Future work on this project would involve updating the dataset with more recent tweets to ensure that the scientific community has access to the recent data in this regard.

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

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are publicly available at <https://doi.org/10.5281/zenodo.6837118>.

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

## Appendix A

The following is the step-by-step process for using the Hydrator app [105] to hydrate this dataset or, in other words, to obtain the text of the tweet, user ID, username, retweet count, language, tweet URL, source, and other public information related to all the Tweet IDs present in this dataset. The Hydrator app works in compliance with the policies for accessing and calling the Twitter API.

1. Download and install the desktop version of the Hydrator app [131].
2. Click on the “Link Twitter Account” button on the Hydrator app to connect the app to an active Twitter account.
3. Click on the “Add” button to upload one of the dataset files (in .txt format, such as TweetIDs\_June\_2022.txt). This process adds the dataset file to the Hydrator app.
4. If the file upload is successful, the Hydrator app will show the total number of Tweet IDs present in the file. For instance, for the file, “TweetIDs\_June\_2022.txt”, the app would show the Number of Tweet IDs as 2361.
5. Provide details for the respective fields: Title, Creator, Publisher, and URL in the app, and click on “Add Dataset” to add this dataset to the app.
6. The app would automatically redirect to the “Datasets” tab. Click on the “Start” button to start hydrating the Tweet IDs. During the hydration process, the progress indicator would increase, indicating the number of Tweet IDs that have been successfully hydrated and the number of Tweet IDs that are pending hydration.
7. After the hydration process ends, a .jsonl file would be generated by the app that the user can choose to save on the local storage.
8. The app would also display a “CSV” button in place of the “Start” button. Clicking on this “CSV” button would generate a .csv file with detailed information about the tweets, which would include the text of the tweet, user ID, username, retweet count, language, tweet URL, source, and other public information related to the tweet.
9. Repeat steps 3–8 for hydrating all the files of this dataset.

## References

1. Wu, Y.-C.; Chen, C.-S.; Chan, Y.-J. Overview of the 2019 Novel Coronavirus (2019-NCoV): The Pathogen of Severe Specific Contagious Pneumonia (SSCP): The Pathogen of Severe Specific Contagious Pneumonia (SSCP). *J. Chin. Med. Assoc.* **2020**, *83*, 217–220. [CrossRef]
2. COVID Live. Coronavirus Statistics—Worldometer. Available online: <https://www.worldometers.info/coronavirus/> (accessed on 6 June 2022).
3. CDC. SARS-CoV-2 Variant Classifications and Definitions. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html> (accessed on 6 June 2022).
4. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. Available online: [https://www.who.int/news/item/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern) (accessed on 6 June 2022).
5. Gobeil, S.M.-C.; Henderson, R.; Stalls, V.; Janowska, K.; Huang, X.; May, A.; Speakman, M.; Beaudoin, E.; Manne, K.; Li, D.; et al. Structural diversity of the SARS-CoV-2 Omicron spike. *Mol. Cell* **2022**, *82*, 2050–2068.e6. [CrossRef] [PubMed]
6. Schmidt, F.; Muecksch, F.; Weisblum, Y.; Da Silva, J.; Bednarski, E.; Cho, A.; Wang, Z.; Gaebler, C.; Caskey, M.; Nussenzweig, M.C.; et al. Plasma Neutralization of the SARS-CoV-2 Omicron Variant. *N. Engl. J. Med.* **2022**, *386*, 599–601. [CrossRef] [PubMed]

7. Feiner, L. WHO Says Omicron Cases Are “off the Charts” as Global Infections Set New Records. Available online: <https://www.cnbc.com/2022/01/12/who-says-omicron-cases-are-off-the-charts-as-global-infections-set-new-records.html> (accessed on 6 June 2022).
8. Weekly Epidemiological Update on COVID-19—22 March 2022. Available online: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-march-2022> (accessed on 6 June 2022).
9. SARS-CoV-2 Omicron Variant Cases Worldwide 2022. Available online: <https://www.statista.com/statistics/1279100/number-omicron-variant-worldwide-by-country/> (accessed on 6 June 2022).
10. Lancet, T. India under COVID-19 lockdown. *Lancet* **2020**, *395*, 1315. [CrossRef]
11. Surano, F.V.; Porfiri, M.; Rizzo, A. Analysis of lockdown perception in the United States during the COVID-19 pandemic. *Eur. Phys. J. Spéc. Top.* **2021**, *231*, 1625–1633. [CrossRef]
12. Jallow, H.; Renukappa, S.; Suresh, S. The impact of COVID-19 outbreak on United Kingdom infrastructure sector. *Smart Sustain. Built Environ.* **2020**, *10*, 581–593. [CrossRef]
13. Tejedor, S.; Cervi, L.; Pérez-Escoda, A.; Jumbo, F.T. Digital Literacy and Higher Education during COVID-19 Lockdown: Spain, Italy, and Ecuador. *Publications* **2020**, *8*, 48. [CrossRef]
14. Fountoulakis, K.N.; Apostolidou, M.K.; Atsiova, M.B.; Filippidou, A.K.; Florou, A.K.; Gousiou, D.S.; Katsara, A.R.; Mantzari, S.N.; Padouva-Markoulaki, M.; Papatriantafyllou, E.I.; et al. Self-reported changes in anxiety, depression and suicidality during the COVID-19 lockdown in Greece. *J. Affect. Disord.* **2021**, *279*, 624–629. [CrossRef]
15. Guzzetta, G.; Riccardo, F.; Marziano, V.; Poletti, P.; Trentini, F.; Bella, A.; Andrianou, X.; Del Manso, M.; Fabiani, M.; Bellino, S.; et al. The Impact of a Nation-Wide Lockdown on COVID-19 Transmissibility in Italy. *arXiv* **2020**, arXiv:2004.12338.
16. Probst, T.; Stipl, P.; Pieh, C. Changes in Provision of Psychotherapy in the Early Weeks of the COVID-19 Lockdown in Austria. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3815. [CrossRef]
17. Oyediran, W.O.; Omoare, A.M.; Owoyemi, M.A.; Adejobi, A.O.; Fasasi, R.B. Prospects and limitations of e-learning application in private tertiary institutions amidst COVID-19 lockdown in Nigeria. *Heliyon* **2020**, *6*, e05457. [CrossRef] [PubMed]
18. Lau, H.; Khosrawipour, V.; Kocbach, P.; Mikolajczyk, A.; Schubert, J.; Bania, J.; Khosrawipour, T. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *J. Travel Med.* **2020**, *27*, taaa037. [CrossRef]
19. Chan, D.Z.; Stewart, R.A.; Kerr, A.J.; Dicker, B.; Kyle, C.V.; Adamson, P.D.; Devlin, G.; Edmond, J.; El-Jack, S.; Elliott, J.M.; et al. The impact of a national COVID-19 lockdown on acute coronary syndrome hospitalisations in New Zealand (ANZACS-QI 55). *Lancet Reg. Health West. Pac.* **2020**, *5*, 100056. [CrossRef]
20. Fahy, S.; Moore, J.; Kelly, M.; Flannery, O.; Kenny, P. Analysing the variation in volume and nature of trauma presentations during COVID-19 lockdown in Ireland. *Bone Jt. Open* **2020**, *1*, 261–266. [CrossRef] [PubMed]
21. LeMenager, T.; Neissner, M.; Koopmann, A.; Reinhard, I.; Georgiadou, E.; Müller, A.; Kiefer, F.; Hillemacher, T. COVID-19 Lockdown Restrictions and Online Media Consumption in Germany. *Int. J. Environ. Res. Public Health* **2020**, *18*, 14. [CrossRef] [PubMed]
22. Stiegler, N.; Bouchard, J.-P. South Africa: Challenges and successes of the COVID-19 lockdown. *Ann. Med. Psychol.* **2020**, *178*, 695–698. [CrossRef] [PubMed]
23. Matheson, A.; McGannon, C.J.; Malhotra, A.; Palmer, K.R.; Stewart, A.E.; Wallace, E.M.; Mol, B.W.; Hodges, R.J.; Rolnik, D.L. Prematurity Rates During the Coronavirus Disease 2019 (COVID-19) Pandemic Lockdown in Melbourne, Australia. *Obstet. Gynecol.* **2021**, *137*, 405–407. [CrossRef]
24. Di Domenico, L.; Pullano, G.; Sabbatini, C.E.; Boëlle, P.-Y.; Colizza, V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Med.* **2020**, *18*, 240. [CrossRef]
25. Lehmann, S.; Skogen, J.C.; Haug, E.; Mæland, S.; Fadnes, L.T.; Sandal, G.M.; Hysing, M.; Bjørknes, R. Perceived consequences and worries among youth in Norway during the COVID-19 pandemic lockdown. *Scand. J. Public Health* **2021**, *49*, 755–765. [CrossRef]
26. Onyeaka, H.; Anumudu, C.K.; Al-Sharify, Z.T.; Egele-Godswill, E.; Mbaegbu, P. COVID-19 pandemic: A review of the global lockdown and its far-reaching effects. *Sci. Prog.* **2021**, *104*, 368504211019854. [CrossRef]
27. Research and Markets Ltd. Online Education Market & Global Forecast, by End User, Learning Mode (Self-Paced, Instructor Led), Technology, Country, Company. Available online: <https://www.researchandmarkets.com/reports/4876815/> (accessed on 15 July 2022).
28. Singh, V.; Thurman, A. How Many Ways Can We Define Online Learning? A Systematic Literature Review of Definitions of Online Learning (1988–2018). *Am. J. Distance Educ.* **2019**, *33*, 289–306. [CrossRef]
29. Education: From Disruption to Recovery, UNSECO Report. Available online: <https://en.unesco.org/covid19/educationresponse> (accessed on 6 June 2022).
30. Education and COVID-19. Available online: <https://data.unicef.org/topic/education/covid-19/> (accessed on 6 June 2022).
31. COVID-19: Are Children Able to Continue Learning during School Closures? Available online: <https://data.unicef.org/resources/remote-learning-reachability-factsheet/> (accessed on 6 June 2022).
32. Stasi, C.; Fallani, S.; Voller, F.; Silvestri, C. Treatment for COVID-19: An overview. *Eur. J. Pharmacol.* **2020**, *889*, 173644. [CrossRef] [PubMed]
33. Peng, Y.; Tao, H.; Satyanarayanan, S.K.; Jin, K.; Su, H. A Comprehensive Summary of the Knowledge on COVID-19 Treatment. *Aging Dis.* **2021**, *12*, 155–191. [CrossRef] [PubMed]

34. Bartoli, A.; Gabrielli, F.; Alicandro, T.; Nascimbeni, F.; Andreone, P. COVID-19 treatment options: A difficult journey between failed attempts and experimental drugs. *Intern. Emerg. Med.* **2021**, *16*, 281–308. [CrossRef] [PubMed]
35. Reopening Schools after COVID-19 Closures Considerations for States. Available online: <http://files.eric.ed.gov/fulltext/ED609236.pdf> (accessed on 6 June 2022).
36. Gunawan, M.; Setiawan, A.A.; Leonita, I. Neville School Reopening during COVID-19 Pandemic: Is It Safe? A Systematic Review. Available online: <https://jamsa.amsa-international.org/index.php/main/article/view/380> (accessed on 1 August 2022).
37. Lockdowns, School Closures Return to Mainland China. Available online: <https://www.usnews.com/news/education-news/articles/2022-03-14/lockdowns-school-closures-return-to-mainland-china> (accessed on 6 June 2022).
38. Sachdev, C. India Postpones In-School Learning as Omicron Surges. Available online: <https://theworld.org/stories/2022-01-07/india-postpones-school-learning-omicron-surges> (accessed on 6 June 2022).
39. School Systems around the World Debate New Closures as Omicron Spreads. Available online: <https://www.washingtonpost.com/world/2022/01/07/global-school-closures-omicron/> (accessed on 6 June 2022).
40. Nearly 6000 Public Schools in Japan at Least Partially Closed Amid Omicron Wave. Available online: <https://www.japantimes.co.jp/news/2022/02/04/national/school-closures-omicron/> (accessed on 8 June 2022).
41. Khan, N. Hong Kong to Shut Schools to Fight Omicron; Foreigners Rush to Leave. Available online: <https://www.wsj.com/articles/hong-kong-sets-all-schools-for-covid-19-response-centers-11645530595> (accessed on 1 August 2022).
42. Collin Binkley (Associated Press). Dozens of US Colleges Starting Semester Online. Available online: <https://www.10tv.com/article/news/nation-world/colleges-online-omicron-covid-remote-learning/507-63ea4bd0-9ccf-40cd-a373-e54da40e2fdb> (accessed on 6 June 2022).
43. Snyder, T.; Byrd, G. The Internet of Everything. *Computer* **2017**, *50*, 8–9. [CrossRef]
44. Boulianne, S. Social media use and participation: A meta-analysis of current research. *Inf. Commun. Soc.* **2015**, *18*, 524–538. [CrossRef]
45. Kavada, A. Social Media as Conversation: A Manifesto. *Soc. Media Soc.* **2015**, *1*, 205630511558079. [CrossRef]
46. Liu, Y.; Singh, L.; Mneimneh, Z. A Comparative Analysis of Classic and Deep Learning Models for Inferring Gender and Age of Twitter Users. In Proceedings of the 2nd International Conference on Deep Learning Theory and Applications, Online, 7–9 July 2021; SciTePress—Science and Technology Publications: Setúbal, Portugal, 2021.
47. Özbaş-Anbarlı, Z. Living in digital space: Everyday life on Twitter. *Commun. Soc.* **2021**, *34*, 31–47. [CrossRef]
48. Gruzds, A.; Wellman, B.; Takhteyev, Y. Imagining Twitter as an Imagined Community. *Am. Behav. Sci.* **2011**, *55*, 1294–1318. [CrossRef]
49. Aslam, S. Twitter by the Numbers (2022): Stats, Demographics & Fun Facts. Available online: <https://www.Omnicoagency.com> (accessed on 13 July 2022).
50. Chen, E.; Deb, A.; Ferrara, E. #Election2020: The first public Twitter dataset on the 2020 US Presidential election. *J. Comput. Soc. Sci.* **2021**, *5*, 1–18. [CrossRef]
51. Haq, E.-U.; Tyson, G.; Lee, L.-H.; Braud, T.; Hui, P. Twitter Dataset for 2022 Russo-Ukrainian Crisis. *arXiv* **2022**, arXiv:2203.02955. [CrossRef]
52. Effrosynidis, D.; Karasakalidis, A.I.; Sylaios, G.; Arampatzis, A. The climate change Twitter dataset. *Expert Syst. Appl.* **2022**, *204*, 117541. [CrossRef]
53. Meng, L.; Dong, Z.S. Natural Hazards Twitter Dataset. *arXiv* **2020**, arXiv:2004.14456.
54. Urchs, S.; Wendlinger, L.; Mitrovic, J.; Granitzer, M. MMoveT15: A Twitter Dataset for Extracting and Analysing Migration-Movement Data of the European Migration Crisis 2015. In Proceedings of the 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Napoli, Italy, 12–14 June 2019; IEEE: New York, NY, USA, 2019; pp. 146–149.
55. Dooms, S.; De Pessemer, T.; Martens, L. MovieTweatings: A Movie Rating Dataset Collected from Twitter. In Proceedings of the Workshop on Crowdsourcing and Human Computation for Recommender Systems (CrowdRec 2013), Held in Conjunction with the 7th ACM Conference on Recommender Systems (RecSys 2013), Hong Kong, China, 12 October 2013.
56. Wijesiriwardene, T.; Inan, H.; Kursuncu, U.; Gaur, M.; Shalin, V.L.; Thirunarayan, K.; Sheth, A.; Arpinar, I.B. ALONE: A Dataset for Toxic Behavior among Adolescents on Twitter. In *Social Informatics*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; pp. 427–439. ISBN 9783030609740.
57. Zangerle, E.; Pichl, M.; Gassler, W.; Specht, G. #nowplaying Music Dataset: Extracting Listening Behavior from Twitter. In Proceedings of the First International Workshop on Internet-Scale Multimedia Management—WISMM'14, Orlando, FL, USA, 7 November 2014; ACM Press: New York, NY, USA, 2014.
58. Sech, J.; DeLucia, A.; Buczak, A.L.; Dredze, M. Civil Unrest on Twitter (CUT): A Dataset of Tweets to Support Research on Civil Unrest. In Proceedings of the Sixth Workshop on Noisy User-Generated Text (W-NUT 2020), Online, 19 November 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 215–221.
59. Tekumalla, R.; Banda, J.M. A Large-Scale Twitter Dataset for Drug Safety Applications Mined from Publicly Existing Resources. *arXiv* **2020**, arXiv:2003.13900.
60. Stemmer, M.; Parmet, Y.; Ravid, G. What Are IBD Patients Talking about on Twitter? In *ICT for Health, Accessibility and Wellbeing*; Springer International Publishing: Cham, Switzerland, 2021; pp. 206–220. ISBN 9783030942083.

61. Adnan, M.; Anwar, K. Online Learning amid the COVID-19 Pandemic: Students' Perspectives. *J. Pedagog. Sociol. Psychol.* **2020**, *2*, 45–51. [[CrossRef](#)]
62. Rasmitadila, R.; Aliyyah, R.R.; Rachmadtullah, R.; Samsudin, A.; Syaodih, E.; Nurtanto, M.; Tambunan, A.R.S. The Perceptions of Primary School Teachers of Online Learning during the COVID-19 Pandemic Period: A Case Study in Indonesia. *J. Ethn. Cult. Stud.* **2020**, *7*, 90–109. [[CrossRef](#)]
63. Irawan, A.W.; Dwisona, D.; Lestari, M. Psychological Impacts of Students on Online Learning during the Pandemic COVID-19. *KONSELI J. Bimbingan. Konseling (E-J.)* **2020**, *7*, 53–60. [[CrossRef](#)]
64. Baticulon, R.E.; Sy, J.J.; Alberto, N.R.I.; Baron, M.B.C.; Mabulay, R.E.C.; Rizada, L.G.T.; Tiu, C.J.S.; Clarion, C.A.; Reyes, J.C.B. Barriers to Online Learning in the Time of COVID-19: A National Survey of Medical Students in the Philippines. *Med. Sci. Educ.* **2021**, *31*, 615–626. [[CrossRef](#)]
65. Hussein, E.; Daoud, S.; Alrabaiiah, H.; Badawi, R. Exploring undergraduate students' attitudes towards emergency online learning during COVID-19: A case from the UAE. *Child. Youth Serv. Rev.* **2020**, *119*, 105699. [[CrossRef](#)]
66. Famularsih, S. Students' Experiences in Using Online Learning Applications Due to COVID-19 in English Classroom. *Stud. Learn. Teach.* **2020**, *1*, 112–121. [[CrossRef](#)]
67. Sutarto, S.; Sari, D.P.; Fathurrochman, I. Teacher strategies in online learning to increase students' interest in learning during COVID-19 pandemic. *J. Konseling Pendidik.* **2020**, *8*, 129. [[CrossRef](#)]
68. Almusharraf, N.; Khahro, S. Students Satisfaction with Online Learning Experiences during the COVID-19 Pandemic. *Int. J. Emerg. Technol. Learn. (ijET)* **2020**, *15*, 246. [[CrossRef](#)]
69. Al-Salman, S.; Haider, A.S. Jordanian University Students' Views on Emergency Online Learning during COVID-19. *Online Learn.* **2021**, *25*, 286–302. [[CrossRef](#)]
70. Bolatov, A.K.; Seisembekov, T.Z.; Askarova, A.Z.; Baikanova, R.K.; Smailova, D.S.; Fabbro, E. Online-Learning due to COVID-19 Improved Mental Health Among Medical Students. *Med. Sci. Educ.* **2020**, *31*, 183–192. [[CrossRef](#)] [[PubMed](#)]
71. Agormedah, E.K.; Henaku, E.A.; Ayite, D.M.K.; Ansah, E.A. Online Learning in Higher Education during COVID-19 Pandemic: A case of Ghana. *J. Educ. Technol. Online Learn.* **2020**, *3*, 183–210. [[CrossRef](#)]
72. Moawad, R.A. Online Learning during the COVID-19 Pandemic and Academic Stress in University Students. *Rev. Rom. Pentru Educ. Multidimens.* **2020**, *12*, 100–107. [[CrossRef](#)]
73. Khan, M.M.; Rahman, S.M.T.; Islam, S.T.A. Online Education System in Bangladesh during COVID-19 Pandemic. *Creat. Educ.* **2021**, *12*, 441–452. [[CrossRef](#)]
74. Catalano, A.J.; Torff, B.; Anderson, K.S. Transitioning to online learning during the COVID-19 pandemic: Differences in access and participation among students in disadvantaged school districts. *Int. J. Inf. Learn. Technol.* **2021**, *38*, 258–270. [[CrossRef](#)]
75. Kapasia, N.; Paul, P.; Roy, A.; Saha, J.; Zaveri, A.; Mallick, R.; Barman, B.; Das, P.; Chouhan, P. Impact of lockdown on learning status of undergraduate and postgraduate students during COVID-19 pandemic in West Bengal, India. *Child. Youth Serv. Rev.* **2020**, *116*, 105194. [[CrossRef](#)]
76. Burns, D.; Dagnall, N.; Holt, M. Assessing the Impact of the COVID-19 Pandemic on Student Wellbeing at Universities in the United Kingdom: A Conceptual Analysis. *Front. Educ.* **2020**, *5*, 582882. [[CrossRef](#)]
77. Küsel, J.; Martin, F.; Markic, S. University Students' Readiness for Using Digital Media and Online Learning—Comparison between Germany and the USA. *Educ. Sci.* **2020**, *10*, 313. [[CrossRef](#)]
78. Al Darayseh, A.S. The Impact of COVID-19 Pandemic on Modes of Teaching Science in UAE Schools. *J. Educ. Pract.* **2020**, *11*, 110–115. [[CrossRef](#)]
79. Tsekhmister, Y.V.; Konovalova, T.; Tsekhmister, B.Y.; Agrawal, A.; Ghosh, D. Evaluation of Virtual Reality Technology and Online Teaching System for Medical Students in Ukraine During COVID-19 Pandemic. *Int. J. Emerg. Technol. Learn.* **2021**, *16*, 127–139. [[CrossRef](#)]
80. Arsaliev, S.M.-K.; Andrienko, A.S. The Development of Ethnocultural Competence of University Students during COVID-19 Pandemic in Russia. In Proceedings of the 2020 3rd International Seminar on Education Research and Social Science (ISERSS 2020), Kuala Lumpur, Malaysia, 24–26 December 2021; Atlantis Press: Paris, France, 2021.
81. Cárdenas-Cruz, A.; Gómez-Moreno, G.; Matas-Lara, A.; Romero-Palacios, P.J.; Parrilla-Ruiz, F.M. An example of adaptation: Experience of virtual clinical skills circuits of internal medicine students at the Faculty of Medicine, University of Granada (Spain) during the COVID-19 pandemic. *Med. Educ. Online* **2022**, *27*, 2040191. [[CrossRef](#)] [[PubMed](#)]
82. Papouli, E.; Chatzifotiou, S.; Tsairidis, C. The use of digital technology at home during the COVID-19 outbreak: Views of social work students in Greece. *Soc. Work Educ.* **2020**, *39*, 1107–1115. [[CrossRef](#)]
83. Parmigiani, D.; Benigno, V.; Giusto, M.; Silvaggio, C.; Sperandio, S. E-inclusion: Online special education in Italy during the COVID-19 pandemic. *Technol. Pedagog. Educ.* **2020**, *30*, 111–124. [[CrossRef](#)]
84. Resch, K.; Alnahdi, G.; Schwab, S. Exploring the effects of the COVID-19 emergency remote education on students' social and academic integration in higher education in Austria. *High. Educ. Res. Dev.* **2022**, 1–15. [[CrossRef](#)]
85. Oyarinde, O.N.; Komolafe, O.G. Impact of Google Classroom as an Online Learning Delivery during COVID-19 Pandemic: The Case of a Secondary School in Nigeria. *J. Educ. Soc. Behav. Sci.* **2020**, *33*, 53–61. [[CrossRef](#)]
86. Chen, T.; Peng, L.; Yin, X.; Rong, J.; Yang, J.; Cong, G. Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic. *Healthcare* **2020**, *8*, 200. [[CrossRef](#)]

87. Drane, C.; Vernon, L.; O'shea, S. The Impact of "Learning at Home" on the Educational Outcomes of Vulnerable Children in Australia during the COVID-19 Pandemic. Available online: [https://www.ncsehe.edu.au/wp-content/uploads/2020/04/NCSEHE\\_V2\\_Final\\_literaturereview-learningathome-covid19-final\\_30042020.pdf](https://www.ncsehe.edu.au/wp-content/uploads/2020/04/NCSEHE_V2_Final_literaturereview-learningathome-covid19-final_30042020.pdf) (accessed on 6 June 2022).
88. Mukuna, K.R.; Aloka, P.J.O. Exploring Educators' Challenges of Online Learning in COVID-19 at a Rural School, South Africa. *Int. J. Learn. Teach. Educ. Res.* **2020**, *19*, 134–149. [[CrossRef](#)]
89. Hsiao, Y.-C. Impacts of course type and student gender on distance learning performance: A case study in Taiwan. *Educ. Inf. Technol.* **2021**, *26*, 6807–6822. [[CrossRef](#)] [[PubMed](#)]
90. Nafrees, A.; Roshan, A.; Baanu, A.N.; Nihma, M.F.; Shibly, F. Awareness of Online Learning of Undergraduates during COVID-19 with special reference to South Eastern University of Sri Lanka. *J. Physics Conf. Ser.* **2020**, *1712*, 012010. [[CrossRef](#)]
91. Cui, L.; Lee, D. CoAID: COVID-19 Healthcare Misinformation Dataset. *arXiv* **2020**, arXiv:2006.00885.
92. Elhadad, M.K.; Li, K.F.; Gebali, F. COVID-19-FAKES: A Twitter (Arabic/English) Dataset for Detecting Misleading Information on COVID-19. In *Advances in Intelligent Networking and Collaborative Systems*; Springer International Publishing: Cham, Switzerland, 2021; pp. 256–268. ISBN 9783030577957.
93. Hayawi, K.; Shahriar, S.; Serhani, M.; Taleb, I.; Mathew, S. ANTI-Vax: A novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health* **2021**, *203*, 23–30. [[CrossRef](#)] [[PubMed](#)]
94. Nasser, N.; Karim, L.; El Ouadrhiri, A.; Ali, A.; Khan, N. n-Gram based language processing using Twitter dataset to identify COVID-19 patients. *Sustain. Cities Soc.* **2021**, *72*, 103048. [[CrossRef](#)]
95. DeVerna, M.R.; Pierrri, F.; Truong, B.T.; Bollenbacher, J.; Axelrod, D.; Loynes, N.; Torres-Lugo, C.; Yang, K.-C.; Menczer, F.; Bryden, J. CoVaxxy: A Collection of English-Language Twitter Posts about COVID-19 Vaccines. *arXiv* **2021**, arXiv:2101.07694.
96. Cheng, M.; Wang, S.; Yan, X.; Yang, T.; Wang, W.; Huang, Z.; Xiao, X.; Nazarian, S.; Bogdan, P. A COVID-19 Rumor Dataset. *Front. Psychol.* **2021**, *12*, 644801. [[CrossRef](#)]
97. Privacy Policy. Available online: [https://twitter.com/en/privacy/previous/version\\_15](https://twitter.com/en/privacy/previous/version_15) (accessed on 6 June 2022).
98. Developer Agreement and Policy. Available online: <https://developer.twitter.com/en/developer-terms/agreement-and-policy> (accessed on 6 June 2022).
99. RapidMiner GmbH Search Twitter—RapidMiner Documentation. Available online: [https://docs.rapidminer.com/latest/studio/operators/data\\_access/applications/twitter/search\\_twitter.html](https://docs.rapidminer.com/latest/studio/operators/data_access/applications/twitter/search_twitter.html) (accessed on 6 June 2022).
100. Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T. YALE: Rapid Prototyping for Complex Data Mining Tasks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD'06, Philadelphia, PA, USA, 20–23 August 2006; ACM Press: New York, NY, USA, 2006.
101. Using Standard Search. Available online: <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/guides/standard-operators> (accessed on 6 June 2022).
102. Anohina, A. Analysis of the Terminology Used in the Field of Virtual Learning. *J. Educ. Technol. Soc.* **2005**, *8*, 91–102.
103. Ma, H.; Shen, L.; Sun, H.; Xu, Z.; Hou, L.; Wu, S.; Fang, A.; Li, J.; Qian, Q. COVID term: A bilingual terminology for COVID-19. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 231. [[CrossRef](#)]
104. Structure of U.S. Education. Available online: <https://www2.ed.gov/about/offices/list/ous/international/usnei/us/edlite-structure-us.html> (accessed on 15 July 2022).
105. The Education System in India. Available online: <https://www.gnu.org/education/edu-system-india.en.html> (accessed on 15 July 2022).
106. British Education System. Available online: <https://www.brightworldguardianships.com/en/guardianship/british-education-system/> (accessed on 15 July 2022).
107. Lamsal, R. Hydrating Tweet IDs. Available online: <https://theneuralblog.com/hydrating-tweet-ids/> (accessed on 6 June 2022).
108. Hydrator: Turn Tweet IDs into Twitter JSON & CSV from Your Desktop! Available online: <https://github.com/DocNow/hydrator> (accessed on 6 June 2022).
109. Tekumalla, R.; Banda, J.M. Social Media Mining Toolkit (SMMT). *Genom. Inform.* **2020**, *18*, e16. [[CrossRef](#)]
110. Twarc: A Command Line Tool (and Python Library) for Archiving Twitter JSON. Available online: <https://github.com/DocNow/twarc> (accessed on 6 June 2022).
111. 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*, 160018. [[CrossRef](#)] [[PubMed](#)]
112. Coronavirus (COVID 19) Online Training Certificate Courses for Workplace, Employees, Workers Australia. Available online: <https://www.sentrient.com.au/covid-19-coronavirus-courses> (accessed on 15 July 2022).
113. Chew, P. LearnT-SMARET Online Course (18-11-2021). COVID-19: Peter Chew Pandemic to Endemic Strategy. Available online: <https://www.youtube.com/watch?v=zLkUPY5Kt6c> (accessed on 15 July 2022).
114. Carvalho, J.; Plastino, A. On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis. *Artif. Intell. Rev.* **2020**, *54*, 1887–1936. [[CrossRef](#)]
115. Wang, J.; Xu, B.; Zu, Y. Deep Learning for Aspect-Based Sentiment Analysis. In Proceedings of the 2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), Chongqing, China, 9–11 July 2021; IEEE: New York, NY, USA, 2021; pp. 267–271.

116. Hong, L.; Dan, O.; Davison, B.D. Predicting Popular Messages in Twitter. In Proceedings of the 20th international conference companion on World Wide Web—WWW'11, Hyderabad, India, 28 March–1 April 2011; ACM Press: New York, NY, USA, 2011.
117. Bouazizi, M.; Ohtsuki, T.O. A Pattern-Based Approach for Sarcasm Detection on Twitter. *IEEE Access* **2016**, *4*, 5477–5488. [[CrossRef](#)]
118. Alvarez-Melis, D.; Saveski, M. Topic Modeling in Twitter: Aggregating Tweets by Conversations. In Proceedings of the Tenth International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016.
119. Boyd, D.; Golder, S.; Lotan, G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, Honolulu, HI, USA, 5–8 January 2010; IEEE: New York, NY, USA, 2010; pp. 1–10.
120. Uysal, I.; Croft, W.B. User Oriented Tweet Ranking: A Filtering Approach to Microblogs. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management—CIKM'11, Glasgow, UK, 24–28 October 2011; ACM Press: New York, NY, USA, 2011.
121. André, P.; Bernstein, M.; Luther, K. Who Gives a Tweet?: Evaluating Microblog Content Value. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work—CSCW'12, Seattle, WA, USA, 11–15 February 2012; ACM Press: New York, NY, USA, 2012.
122. Ito, J.; Song, J.; Toda, H.; Koike, Y.; Oyama, S. Assessment of Tweet Credibility with LDA Features. In Proceedings of the 24th International Conference on World Wide Web—WWW'15 Companion, Florence, Italy, 18–22 May 2015; ACM Press: New York, NY, USA, 2015.
123. Stephens, M. A geospatial infodemic: Mapping Twitter conspiracy theories of COVID-19. *Dialogues Hum. Geogr.* **2020**, *10*, 276–281. [[CrossRef](#)]
124. Wu, C.; Wu, F.; Wu, S.; Huang, Y.; Xie, X. Tweet Emoji Prediction Using Hierarchical Model with Attention. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 8–12 October 2018; ACM: New York, NY, USA, 2018.
125. Mccreadie, R.; Macdonald, C. Relevance in Microblogs: Enhancing Tweet Retrieval Using Hyperlinked Documents. Available online: [http://terrierteam.dcs.gla.ac.uk/publications/oair2013\\_McCreadie.pdf](http://terrierteam.dcs.gla.ac.uk/publications/oair2013_McCreadie.pdf) (accessed on 7 June 2022).
126. Salas-Zárate, M.D.P.; Paredes-Valverde, M.A.; Rodríguez-García, M.; Valencia-García, R.; Alor-Hernández, G. Automatic detection of satire in Twitter: A psycholinguistic-based approach. *Knowl.-Based Syst.* **2017**, *128*, 20–33. [[CrossRef](#)]
127. Alowibdi, J.S.; Buy, U.A.; Yu, P.S.; Ghani, S.; Mokbel, M. Deception detection in Twitter. *Soc. Netw. Anal. Min.* **2015**, *5*, 32. [[CrossRef](#)]
128. Zheng, L.; Han, K. Extracting Categorical Topics from Tweets Using Topic Model. In *Information Retrieval Technology*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 86–96. ISBN 9783642450679.
129. Zahra, K.; Azam, F.; Butt, W.H.; Ilyas, F. A Framework for User Characterization Based on Tweets Using Machine Learning Algorithms. In Proceedings of the 2018 VII International Conference on Network, Communication and Computing—ICNCC 2018, Taipei City, Taiwan, 14–16 December 2018; ACM Press: New York, NY, USA, 2018.
130. Sloan, L.; Morgan, J.; Housley, W.; Williams, M.; Edwards, A.; Burnap, P.; Rana, O. Knowing the Tweepers: Deriving Sociologically Relevant Demographics from Twitter. *Sociol. Res. Online* **2013**, *18*, 74–84. [[CrossRef](#)]
131. Hydrator. Available online: <https://github.com/DocNow/hydrator/releases> (accessed on 1 August 2022).