Communication

Twitter Big Data as a Resource for Exoskeleton Research: A Large-Scale Dataset of about 140,000 Tweets from 2017–2022 and 100 Research Questions

Nirmalya Thakur

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

Abstract: The exoskeleton technology has been rapidly advancing in the recent past due to its multitude of applications and diverse use cases in assisted living, military, healthcare, firefighting, and industry 4.0. The exoskeleton market is projected to increase by multiple times its current value within the next two years. Therefore, it is crucial to study the degree and trends of user interest, views, opinions, perspectives, attitudes, acceptance, feedback, engagement, buying behavior, and satisfaction, towards exoskeletons, for which the availability of Big Data of conversations about exoskeletons is necessary. The Internet of Everything style of today's living, characterized by people spending more time on the internet than ever before, with a specific focus on social media platforms, holds the potential for the development of such a dataset by the mining of relevant social media conversations. Twitter, one such social media platform, is highly popular amongst all age groups, where the topics found in the conversation paradigms include emerging technologies such as exoskeletons. To address this research challenge, this work makes two scientific contributions to this field. First, it presents an open-access dataset of about 140,000 Tweets about exoskeletons that were posted in a 5-year period from 21 May 2017 to 21 May 2022. Second, based on a comprehensive review of the recent works in the fields of Big Data, Natural Language Processing, Information Retrieval, Data Mining, Pattern Recognition, and Artificial Intelligence that may be applied to relevant Twitter data for advancing research, innovation, and discovery in the field of exoskeleton research, a total of 100 Research Questions are presented for researchers to study, analyze, evaluate, ideate, and investigate based on this dataset.

Keywords: Exoskeleton; Twitter; Tweets; big data; social media; data mining; dataset; data science; natural language processing; information retrieval

1. Introduction

A robotic exoskeleton may be broadly defined as a wearable electromechanical device developed with the primary objective of augmenting the physical performance, stamina, and abilities of the person who wears it [1]. Depending on the specific application for which the exoskeleton would be used, they differ in design, functionality, operation, and necessary maintenance [2]. Exoskeletons can be broadly classified as upper limb exoskeletons [3] and lower limb exoskeletons [4]. In the last few years, exoskeletons have been developed for specific body parts or joints. These include exoskeletons for the knee [5], shoulder [6], elbow [7], ankle [8], waist [9], hip [10], neck [11], spine [12], wrist [13], and index finger [14]. The exoskeleton technology has been rapidly advancing [15] in the recent past on account of its multitude of applications and use cases. Some of the applications of exoskeletons [16–18] include (1) assisted living: helping older adults as well as people with various forms of disabilities to perform their daily routine tasks independently; (2) military: for augmenting productivity and lowering fatigue; (3) healthcare: to improve the quality of life of people who have lost one or more of their arms or legs or suffer paralysis in the same; (4) firefighting: to assist firefighters in climbing faster, as well as for helping them to

lift and carry heavy equipment; (5) industry 4.0: to increase labor productivity, to assist in the transportation of heavy machinery, and to help workers in labor-intensive tasks. Given these diverse applications of exoskeletons and the projected increase in the same, the exoskeleton market is growing rapidly. It was USD 200 million in 2017 [19], and it is estimated to become USD 1.3 billion by the end of 2024 [20].

With the advent and rapid adoption of the Internet of Everything lifestyle [21], in recent times, people's everyday lives involve interacting with computers and a myriad of technology-based gadgets and devices in multiple ways while using various forms of Internet-based services and applications, a lot more than ever before [22,23]. In this Internet of Everything era, the use of social media platforms has skyrocketed in the recent past [24]. Social media platforms have evolved and transformed into "virtual" spaces via which people from all demographics communicate with each other to form their social support systems [25] and develop "online" interpersonal relationships [26]. Such "virtual" spaces consist of conversations on diverse topics such as emerging technologies, news, current events, politics, family, relationships, and career opportunities [27]. When such conversations are related to a specific technology, the study of the same can indicate the degree and trends of user interest, views, opinions, perspectives, attitudes, and feedback towards that specific technology [28]. Such inferences from conversations about a technology can serve a wide range of use cases such as the indication of the market potential [29], sales prediction [30], consumer engagement [31], technology acceptance [32], buying behavior [33], and customer satisfaction [34], just to name a few.

Twitter, one such social media platform, used by people of almost all age groups [35], has been rapidly gaining popularity in all parts of the world and is currently the second most visited social media platform [36]. At present, there are about 192 million daily active users on Twitter, and approximately 500 million tweets are posted on Twitter every day [37]. Therefore, mining and studying this Big Data of conversations from Twitter has been of significant interest to the research community. In the last few years, there have been several works in the fields of Big Data, Data Mining, and Natural Language Processing that focused on the development of datasets of Twitter conversations related to different topics, technologies, events, diseases, viruses, etc., such as—movies [38], COVID-19 [39], elections [40], toxic behavior amongst adolescents [41], music [42], natural hazards [43], personality traits [44], civil unrest [45], drug safety [46], climate change [47], hate speech [48], migration patterns [49], conspiracy theories [50], and Inflammatory Bowel Disease [51], just to name a few. Recent studies [52–54] have shown that sharing such data helps in the advancement of research, improves the quality of innovation, supports better investigation, and helps to avoid redundant efforts.

In view of the rapid advances in exoskeleton research in the recent past [15], its applicability for a wide range of use cases for diverse users with diverse needs [3–14], projected increase of the booming exoskeleton market to become USD 1.3 billion by the end of 2024 [20], and limitations in prior works in this field [3–14] that did not focus on mining social media conversations about exoskeletons, which could have served as an information resource for studying user interest, views, opinions, perspectives, attitudes, acceptance, and feedback towards exoskeletons, it is crucial to develop a dataset of conversations related to exoskeletons. The work presented in this paper aims to address this research challenge by exploring the intersections of Big Data, Data Mining, Natural Language Processing, Information Retrieval, Internet of Everything, and their interrelated disciplines. It makes the following scientific contributions:

1. It presents a large-scale open-access Twitter dataset of 138,584 Tweets (including original Tweets, retweets, and replies) about exoskeletons posted on Twitter for a period of 5-years from 21 May 2017 to 21 May 2022. The dataset is available at https://dx.doi.org/10.21227/r5mv-ax79.

2. Based on a comprehensive review of 108 emerging works in these fields, this paper discusses multiple interdisciplinary applications of this dataset and presents a list

of 100 research questions for researchers to study, analyze, evaluate, ideate, and investigate based on this dataset.

The rest of this paper is organized as follows. The methodology that was used to develop this dataset is presented in Section 2. Section 3 presents the results. Section 4 discusses multiple interdisciplinary applications of this dataset and presents 100 research questions for researchers to investigate. Section 5 concludes the paper by summarizing the scientific contributions of this research and discussing the scope for future work.

2. Methodology

This section describes the methodology that was followed for the development of this dataset, which is publicly available at https://dx.doi.org/10.21227/r5mv-ax79. The dataset contains the Tweet IDs of 138,584 Tweets about exoskeletons that were posted over a 5-year period from 21 May 2017 to 21 May 2022. Here, 21 May 2022 is the most recent date of data collection at the time of submission of this paper, and 21 May 2017 is the earliest date for which Tweets could be mined when the process of data collection was started. As a central focus of this work involves developing a Twitter dataset; therefore, the privacy policy, developer agreement, and guidelines for content redistribution of Twitter [55,56] were thoroughly studied. The privacy policy of Twitter [55] states *"Twitter is public and Tweets are immediately viewable and searchable by anyone around the world"*. To add, the Twitter developer agreement [56] defines Tweets as *"public data"*. The guidelines for Twitter content redistribution [56] 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"*. 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) is in compliance with the privacy policy, developer agreement, and content redistribution guidelines of Twitter (at the time of writing of this paper).

The Tweets were collected by using the Search Twitter *operator* [57] in RapidMiner [58] and by using the Advanced Search feature of the Twitter API. The Advanced Search feature of the Twitter API is available to a user when they are logged in to twitter.com. It allows the user to tailor search results to specific date ranges, people and more. Specifically, the Advanced Search feature of the Twitter API allows providing several inputs which include specifications to search for Tweets containing all specified keywords in any position, Tweets containing an exact phrase(s), Tweets containing any of the specified keywords, Tweets excluding specific keywords, Tweets with a specific hashtag, and Tweets in a specific language. It also allows searching for Tweets from a specific account, Tweets sent as replies to a specific account, and Tweets that mention a specific account. This makes it easier to find specific Tweets posted during specific date ranges based on the values of one or more of these inputs. 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. Every application developed in RapidMiner is known as a *process*, and every RapidMiner *process* comprises one or more *operators*, which represent different operational features of the application. The Search Twitter *operator* works by building a connection with the Twitter API while following the rate limits for accessing Twitter data as per Twitter's policy [59]. This ensures that the Search Twitter *operator* functions by complying with the Twitter API search policies [60]. To use the Search Twitter *operator*, a *process* was developed in the RapidMiner studio that comprised this *operator*. The Search Twitter *operator* requires mandatory input from the user for the query *field*. Here, query represents a keyword or a set of keywords or phrases based on which relevant Tweets (Tweets containing that keyword or keywords or phrases) would be filtered and returned in the form of results.

Therefore, to determine the input for this query field, the previous works [1–19] in the field of exoskeleton technology were studied to determine the most common phrases or set of keywords that are used to refer to the underlining exoskeleton systems. A list of phrases and keywords can be found in these works, which include *"Exoskeleton-Type Systems"*, *"Upper-Limb Exoskeleton"*, *"Indego Explorer Lower-Limb Exoskeleton"*, *"Rehabilitation Exoskeleton"*, *"Powered Knee Exoskeleton"*, *"Exo4Work Shoulder Exoskeleton"*, *"Wearable Elbow Exoskeleton"*, *"Powered Ankle Exoskeleton"*, *"Wearable Waist Exoskeleton"*, *"Powered Hip Exoskeleton"*, *"Neck Supporting Exoskeleton"*, *"Elastic Spine Exoskeleton"*, *"Wrist Exoskeleton"*, *"Robotic Exoskeleton"*, *"Finger Exoskeleton"*, and *"Lower Limb Exoskeleton"*. As can be seen from this collection of phrases, the keyword "exoskeleton" always exists in a phrase that is being used to refer to a specific kind of exoskeleton system. Therefore, in the query field of the Search Twitter *operator*, only the keyword "exoskeleton" was entered. RapidMiner is not case-sensitive, so it ensured that the results would contain Tweets where any form of upper case or lower case combinations have been used to spell the word "exoskeleton" in a Tweet. The output of this *process* in RapidMiner that comprised the Search Twitter *operator* consisted of multiple attributes, which are mentioned in Table 1. This RapidMiner *process* was run multiple times, and the results were merged. Similarly, the keyword "exoskeleton" was used as an input for the Advanced Search feature of the Twitter API to search Tweets that contained this exact keyword. No other inputs other than this keyword were provided to the Advanced Search feature of the Twitter API. The results obtained from both these methodologies were merged and the duplicate Tweets were removed to develop this dataset.

Table 1. Description of the attributes from the results of the RapidMiner *process* that used the Search Twitter *operator*.

Attribute Name	Description
Row no.	Row number of the results
Id	ID of the Tweet
Created-At	Date and time when the Tweet was posted
From-User	Twitter username of the user who posted the Tweet
From-User-Id	Twitter User ID of the user who posted the Tweet
To-User	Twitter username of the user whose Tweet was replied to (if the Tweet was a reply) in the current Tweet
To-User-Id	Twitter user ID of the user whose Tweet was replied to (if the tweet was a reply) in the current Tweet
Language	Language of the tweet
Source	Source of the Tweet to determine if the Tweet was posted from an Android source, Twitter website, etc.
Text	Complete text of the Tweet, including embedded URLs
Geo-Location-Latitude	Geo-Location (Latitude) of the user posting the Tweet
Geo-Location-Longitude	Geo-Location (Longitude) of the user posting the Tweet
Retweet Count	Retweet count of the Tweet

To ensure compliance with the Privacy Policy, Developer Agreement, and Content Redistribution guidelines of Twitter [55,56] for complete data anonymization, and to comply with the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management [61], multiple data filters were introduced in the RapidMiner *process* to filter out all the attributes from the results, other than the **"Id"** attribute, that represents the unique Tweet ID for each Tweet. After running the *process*, each time, these Tweet IDs were exported as a .csv file and then converted to a .txt file to facilitate easy hydration of the Tweets (explained in Section 3). Similarly, for the results from the Advanced Search feature of the Twitter API, only the Tweet IDs were retained, and rest of the information related to these Tweets were filtered and removed before compiling the results with the results from the RapidMiner *process* and removing the duplicates. As the dataset contains close to 140,000 Tweet IDs, to ensure that the process of hydration of the Tweet IDs is easy and less time-consuming, the dataset is divided into seven sets or files (7 .txt files) based on the date ranges of the associated Tweets. An overview of these sets is shown in Table 2. It is worth mentioning here that neither Twitter API's standard search

feature (that was used via the Search Twitter *operator* in RapidMiner) nor the Advanced Search feature of the Twitter API return a complete index of all Tweets in a date range. So, it is possible that multiple Tweets about exoskeletons posted in the date range of 21 May 2017–21 May 2022, were not collected by these methodologies and are thus not a part of this dataset. To add, 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.

Table 2. Characteristics of the Dataset Files with the associated Tweet ID count and Date Range.

Filename	Number of Tweet IDs	Date Range of the Tweets
Exoskeleton_TweetIDs_Set1.txt	22,945	20 July 2021–21 May 2022
Exoskeleton_TweetIDs_Set2.txt	19,415	1 December 2020–19 July 2021
Exoskeleton_TweetIDs_Set3.txt	16,673	29 April 2020–30 November 2020
Exoskeleton_TweetIDs_Set4.txt	16,208	5 October 2019–28 April 2020
Exoskeleton_TweetIDs_Set5.txt	17,983	13 February 2019–4 October 2019
Exoskeleton_TweetIDs_Set6.txt	34,009	9 November 2017–12 February 2019
Exoskeleton_TweetIDs_Set7.txt	11,351	21 May 2017–8 November 2017

3. Results and Discussions

As mentioned in Section 2, this dataset contains Tweet IDs corresponding to 138,584 Tweets about exoskeletons posted on Twitter from 21 May 2017 to 21 May 2022. The complete information associated with a Tweet, such as the text of a Tweet, user name, user ID, timestamp, retweet count, etc., can be obtained from a Tweet ID by following a process known as hydration of Tweet ID [62]. A simple example of this process can be observed by visiting this website [63] and entering any Tweet ID from this dataset. The website would immediately generate the URL of the associated Tweet that can be studied. However, in a realistic scenario repeating this step 138,584 times would be a very difficult process. In view of the same and for processing Twitter datasets, researchers in this field have developed various tools that can hydrate Tweet IDs. Some of the most popular tools include—Hydrator [64], Social Media Mining Toolkit [65], and Twarc [66]. This section comprises two parts. In Section 3.1, the process of hydration in the context of using one of these tools—the Hydrator app, is explained, and a brief discussion about the results that would be obtained thereafter is also presented. Section 3.2 presents a statistical analysis of multimodal components of the Tweets of this dataset.

3.1. Hydrating the Dataset—Steps and Associated Results

In this context, results mean the complete information (text of the Tweet, user ID, user name, retweet count, language, Tweet URL, source, and other public information related to the Tweet) associated with each of the Tweet IDs in this dataset.

The following is the step-by-step process for using the Hydrator app to hydrate this dataset. For the purpose of this discussion, only one file from this dataset, "Exoskeleton_TweetIDs_Set2.txt", is being used.

1. Download and install the desktop version of the Hydrator app from this website [67]. The version that was used for this work is v0.30.0.
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 add a new dataset comprising only Tweet IDs (Figure 1). Browse and select the file—"Exoskeleton_TweetIDs_Set2.txt" available on local storage.
4. If the file upload is successful, the Hydrator app will show the total number of Tweet IDs present in the file. For this file—"Exoskeleton_TweetIDs_Set2.txt", the app would show the number of Tweet IDs as 19,415.
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 (Figure 2). 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 may choose to save. 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, user name, retweet count, language, Tweet URL, source, and other public information related to the Tweet.
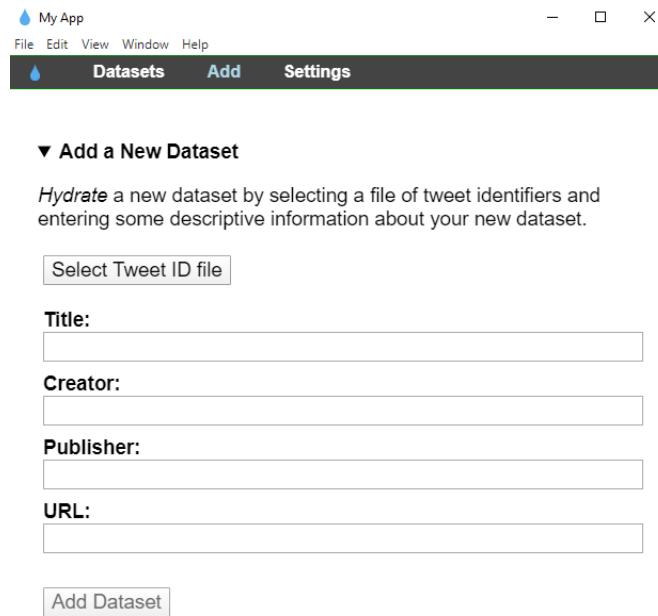

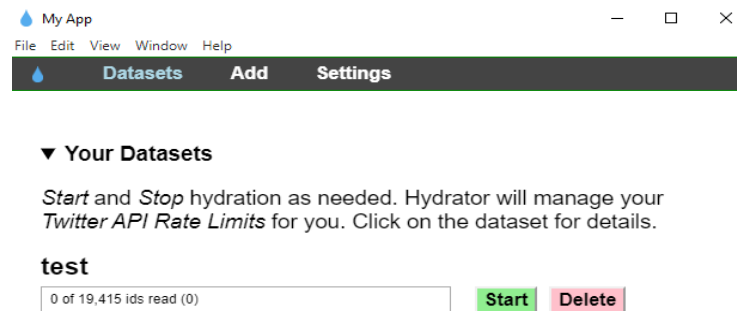**Figure 1.** Screenshot from the Hydrator app for the dataset upload step.

Figure 2. Screenshot from the Hydrator app for the dataset processing step.

Even though the above steps are discussed in the context of using the "Exoskeleton_TweetIDs_Set2.txt" file from the dataset, the same steps can be repeated for hydrating all the other dataset files. However, it is to be noted that Hydrator functions by connecting with the Twitter API and works in compliance with the Twitter API rate limits [59]. Therefore, when the app reaches the 15-min or 3-hour or any similar usage limit, hydrating any additional dataset files might not be possible in that 15-min or 3-hour or a similar usage criteria/window.

3.2. Statistical Analysis of the Dataset

This section presents the findings of a comprehensive statistical analysis of this dataset which was performed using RapidMiner [58]. RapidMiner has an in-built statistical analysis function that can be used for studying multimodal characteristics of datasets. The characteristics of these Tweets that were studied include language, embedded URLs, retweet

count, favorite count, length of the Tweets, dates when these Tweets were posted, and the number of users who posted these Tweets. This analysis was performed on 9 July 2022 (the most recent date at the time of resubmission of this paper after the first round of review). Prior to performing this analysis, the Tweet IDs needed to be hydrated, and the Hydrator app [64] was used to hydrate this dataset (Section 3.1). The Hydrator app reported that 7% of the Tweets were deleted at the time of Hydration. Therefore, this analysis is presented from the remaining set of Tweets. However, it is worth mentioning here that these specific characteristics could be studied only for those Tweets which presented the relevant data. For instance, a few Tweets were observed to be just images and videos of different types of exoskeletons and retweets of the same. For such Tweets, characteristics such as language analysis, embedded URL identification, and computing the length of the Tweet did not return any results. A total of 49 different languages were detected in the set of Tweets in this dataset. The languages were represented as per the ISO 639 codes [68] in the results that were obtained in RapidMiner. ISO 639 [68] is a standardized nomenclature used to classify languages. Each language is assigned a two-letter (639-1) and three-letter (639-2 and 639-3) lowercase abbreviation, amended in later versions of the nomenclature. These language codes were converted to the actual languages for a seamless understanding of the results. The results are presented in Table 3, along with the exact count and the fraction of Tweets posted in each language. Here, the values in the Fraction column represent values on a scale of 0 to 1. As can be seen from Table 3, English was the most common language that was used for tweeting about exoskeletons from 21 May 2017 to 21 May 2022.

Table 3. Analysis of the different languages represented in the Tweets present in this dataset.

Language	Language Code	Absolute Count	Fraction
English	en	71,585	0.904296308
Danish	da	1085	0.013706244
Tagalog	tl	926	0.011697679
Spanish	es	774	0.009777542
Indonesian	in	419	0.00529301
Japanese	ja	288	0.003638155
French	fr	267	0.003372873
German	de	206	0.002602292
Hungarian	hu	197	0.002488599
Czech	cs	192	0.002425437
Catalan	ca	183	0.002311744
Romanian	ro	181	0.002286479
Turkish	tr	119	0.001503265
Estonian	et	117	0.001478001
Portuguese	pt	110	0.001389573
Finnish	fi	105	0.001326411
Basque	eu	96	0.001212718
Dutch	nl	89	0.001124291
Slovenian	sl	84	0.001061129
Russian	ru	65	8.21E-04
Thai	th	50	6.32E-04
Haitian	ht	43	5.43E-04
Italian	it	36	4.55E-04
Arabic	ar	31	3.92E-04
Lithuanian	lt	19	2.40E-04
Swedish	sv	18	2.27E-04
Polish	pl	16	2.02E-04
Papiamentu	qst	13	1.64E-04
Korean	ko	11	1.39E-04
Kunstsprachen	art	8	1.01E-04
Chinese	zh	8	1.01E-04
Greek	el	6	7.58E-05
Vietnamese	vi	6	7.58E-05

Table 3. *Cont.*

Language	Language Code	Absolute Count	Fraction
Hindi	hi	5	6.32E-05
Icelandic	is	5	6.32E-05
Norwegian	no	4	5.05E-05
Persian	fa	3	3.79E-05
Hebrew	iw	3	3.79E-05
Welsh	cy	2	2.53E-05
Latvian	lv	2	2.53E-05
Malayalam	ml	2	2.53E-05
Urdu	ur	2	2.53E-05
Bulgarian	bg	1	1.26E-05
Gujrati	gu	1	1.26E-05
Armenian	hy	1	1.26E-05
Georgian	ka	1	1.26E-05
Burmese	my	1	1.26E-05
Tamil	ta	1	1.26E-05
Ukrainian	uk	1	1.26E-05

The analysis of the URLs embedded in the Tweets is presented next. All the Tweets did not have an URL embedded in the Tweet text. Out of the Tweets that had an URL embedded in the text, a total of 23,831 different URLs were detected. To avoid presenting a table with 23,831 rows, the top 50 URLs in terms of the number of times of occurrence are represented in Table 4. As can be seen from Table 4, the top two embedded URLs pointed to the news published by BBC News, a business division of the British Broadcasting Corporation (BBC), about a mind-controlled exoskeleton.

Table 4. Analysis of the different URLs embedded in the Tweets present in this dataset.

URL	Absolute Count	Fraction
https://www.bbc.co.uk/news/health-49907356	268	0.008193464
https://www.bbc.com/news/health-49907356	201	0.006145098
http://smarturl.it/xu3k1p	58	0.001773212
https://ehikioya.com/product/wearable-chairless-chair/	54	0.001650922
http://bit.ly/2tkwtak	48	0.001467486
https://ift.tt/2raBxMq	47	0.001436913
https://ift.tt/2J7R2N4	46	0.001406341
https://www.theguardian.com/world/2019/oct/04/paralysed-man-walks-using-mind-controlled-exoskeleton?CMP=share_btn_tw	45	0.001375768
http://youtu.be/0z0i8GzfyXs?a	40	0.001222905
http://youtu.be/qTxxwLWsMoA?a	40	0.001222905
https://www.caradvice.com.au/674659/ford-expands-trial-of-exoskeleton-vest/	40	0.001222905
http://bit.ly/1fZbb79	39	0.001192332
https://www.bbc.co.uk/news/technology-44628872	38	0.00116176
http://www.bbc.co.uk/news/technology-44628872	36	0.001100615
http://youtu.be/OiAVTz5BbZQ?a	36	0.001100615
https://techcrunch.com/2018/03/29/roam-debuts-a-robotic-exoskeleton-for-skiers/	36	0.001100615
https://prescient.info/xLHKDINg/	34	0.001039469
https://spectrum.ieee.org/the-human-os/biomedical/devices/cyberdynes-medical-exoskeleton-strides-to-fda-approval	34	0.001039469
http://bit.ly/2eNowmn	33	0.001008897
https://cbsn.ws/31OufyB?fbclid=IwAR1t_TSzmgKxZ73ASQ5Uc9nWoQSJoQyS45bwieBCAwKL00nCHXAJeebJMrw	33	0.001008897
https://mashable.com/2018/02/06/ford-exoskeleton-factory-ekso/	33	0.001008897
https://www.bbc.co.uk/news/av/technology-47319932/exoskeleton-helps-people-with-paralysis-to-walk	33	0.001008897
http://toastradio.com/np/71C894A884189507	32	0.000978324
https://prescient.info/xdtmhojB/	31	0.000947751
https://twitter.com/BadMedicalTakes/status/1499867782934142977	31	0.000947751
http://ift.tt/2CluPrh	30	0.000917179
https://prescient.info/xhzMlhIZ/	30	0.000917179
https://www.engadget.com/2017/07/06/russian-exoskeleton-suit-turns-soldiers-into-stormtroopers/	30	0.000917179
https://youtu.be/zLWuHo63C8k	29	0.000886606
https://futurism.com/amazons-alexa-helps-this-exoskeleton-respond-to-spoken-instructions/	28	0.000856034
http://wknc.org/listen	27	0.000825461
https://tunein.com/radio/NOFM-s184668/	27	0.000825461
http://twinybots.ch https://www.forbes.com/sites/servicenow/2020/06/11/ai-is-the-brains-exoskeleton/	26	0.000794888

Table 4. *Cont.*

URL	Absolute Count	Fraction
https://ift.tt/2F4GCzy	26	0.000794888
https://www.bbc.com/news/technology-44628872	26	0.000794888
https://www.forbes.com/sites/servicenow/2020/06/11/ai-is-the-brains-exoskeleton/	26	0.000794888
https://www.reuters.com/lifestyle/father-builds-exoskeleton-help-wheelchair-bound-son-walk-2021-07-26/	26	0.000794888
http://ift.tt/2usZoYj	25	0.000764316
https://design-milk.com/morpheus-hotel-zaha-hadid-architects-worlds-first-high-rise-exoskeleton/	25	0.000764316
https://wp.me/p8RPw8-eF	25	0.000764316
https://www.theguardian.com/world/2019/oct/04/paralysed-man-walks-using-mind-controlled-exoskeleton	25	0.000764316
http://wp.me/p5QmfL-158	24	0.000733743
https://www.cnn.com/2019/10/04/health/paralyzed-man-robotic-suit-intl-scli/index.html	23	0.00070317
https://youtu.be/qTxxwLWsMoA	23	0.00070317
http://ift.tt/2uBXYJI	22	0.000672598
https://apple.news/ApW9EjskUS5Wij9-qUhlzEQ	22	0.000672598
https://futurism.com/ford-pilots-new-exoskeleton-lessen-worker-fatigue/	22	0.000672598
https://ift.tt/2sRBFkC	22	0.000672598
https://twitter.com/blaireerskine/status/1329844449426468865	22	0.000672598
https://scout.com/military/warrior/Article/Army-Tests-New-Super-Soldier-Exoskeleton-111085386	21	0.000642025

The analysis of the retweet count of all the Tweets showed that the highest and lowest number of retweets received by a Tweet were 2624 and 0, respectively. The average retweet count of each Tweet was computed to be 1.127. Upon studying the favorite count of all the Tweets, it was observed that, on average, each Tweet received a favorite count of 6.915. The highest favorite count of a Tweet was observed to be 90,740, while the lowest favorite count was observed to be 0 for multiple Tweets. Thereafter, the length of the Tweets was studied in terms of character counts. For those Tweets that had an embedded URL and users tagged, the length of the URL and length of the Twitter user ID(s) of the tagged users were added to the character count of any text present in those respective Tweets. Based on this approach, the maximum character count of a Tweet was observed to be 949. This specific Tweet had 43 users tagged in the Tweet, which attributed to its unusually high character count. The lowest character count for a Tweet was observed to be 2, and the average character count of all the Tweets was observed to be 128.054. Even though the dataset presents Tweets posted over a period of 5 years, it was observed that, on certain dates, a high number of Tweets about exoskeletons were recorded. Table 5 presents the list of 21 specific dates from this dataset when 100 or more Tweets were posted about exoskeletons over a 24-h period. As can be seen from Table 5, on 4 October 2019, the maximum number of Tweets about exoskeleton were posted in the last 5 years. It was also on the same date that BBC news published the news articles (top two URLs in Table 4) about the mind-controlled exoskeleton. This seems to indicate that might be a direct correlation between BBC news posting about this new exoskeleton and the number of Tweets about exoskeletons rapidly increasing over a 24-h period.

Table 5. List of specific dates from this dataset when more than 100 Tweets about exoskeletons were posted in a single day.

Date	Number of Tweets
4 October 2019	1014
30 March 2018	214
10 November 2017	203
24 February 2019	194
8 July 2018	188
7 October 2019	184
6 July 2017	162
6 October 2019	160
22 November 2019	151
29 March 2018	137
24 November 2017	113
29 November 2018	111
23 June 2017	109

Table 5. *Cont.*

Date	Number of Tweets
27 September 2017	109
14 December 2020	108
11 November 2017	107
25 February 2019	104
9 January 2019	104
23 November 2019	103
24 November 2019	101
28 November 2017	100

The analysis of User IDs showed that a total of 73,345 distinct User IDs are present. Or in other words, a total of 73,345 Twitter users accounted for all the Tweets about exoskeletons present in this dataset. The highest and lowest number of Tweets about exoskeletons posted by a single user was observed to be 3927 and 1, respectively.

4. Potential Applications and Research Directions

This comprehensive dataset that consists of Tweet IDs of 138,584 Tweets about exoskeletons that were posted over a 5-year period from 21 May 2017 to 21 May 2022 is expected to have multiple interdisciplinary applications related to the advancement of research, innovation, and discovery in the field of exoskeleton technology. Some of these applications could include interpretation and analysis of the degree and trends of user interests, perspectives, opinions, reviews, attitudes, feedback, engagement, acceptance, buying behavior, and satisfaction, related to different exoskeletons used by diverse users for different use cases. To further support the same, based on a comprehensive review of emerging works in the fields of Big Data, Data Mining, Natural Language Processing, Information Retrieval, Pattern Recognition, and Artificial Intelligence that may be applied to relevant Twitter data for advancing research, innovation, and discovery the field of exoskeleton research, a total of 100 Research Questions (RQs) are presented for researchers to study, analyze, evaluate, ideate, and investigate based on this dataset. This section is divided into two parts. In Section 4.1, this list of 100 RQs is presented. As can be seen from Section 4.1, each of these RQs focuses on different problem statements and/or research directions. Therefore, the underlining procedure for the software development and its implementation for investigating each of these RQs is going to be different. Furthermore, all the recent works (Section 4.1) that have been reviewed to develop these RQs involved the usage of different programming languages, such Python, R, C++, JavaScript, Java, Ruby, etc., and research platforms/tools, such as RapidMiner, Tableau, Visual Basic, .NET, etc., for the development of the associated software framework. So, preparing a common framework that can be directly applied to investigate all these RQs is not possible. However, to facilitate easier investigation and a starting point for studying some of these RQs, a step-by-step methodology in RapidMiner is described in Section 4.2.

4.1. List of 100 Research Questions Related to Exoskeletons

These RQs are presented as follows:

RQ1. Sentiment analysis [69] of these Tweets would help to identify the positive, negative, and neutral sentiments associated with these conversations about exoskeletons on Twitter.

RQ2. Deep learning may be used to identify the emotional state of the users in terms of the basic emotional responses - fear, anger, joy, sadness, disgust, and surprise [70], at the time of posting of these respective Tweets.

RQ3. Aspect-based sentiment analysis, along with tokenization and lemmatization of these Tweets [71], may be performed to identify the specific aspects or subject matters related to exoskeletons to which certain specific sentiments or emotional responses are associated.

RQ4. Studying the trends in Tweet counts [72] and the associated sentiments and emotional responses to detect any correlations between the two.

RQ5. Studying the word count of each of these Tweets [73] to determine if any correlation exists between the word count and the associated sentiments and emotional responses towards different exoskeleton products.

RQ6. Investigating whether the number of replies or retweets of a Tweet [74] posted by a company to share the news about a new kind of exoskeleton could have a correlation with the influence and follower metrics of the company's Twitter profile.

RQ7. Detecting popular Tweets [75] related to exoskeletons and studying the subject matters and aspects mentioned in those Tweets by performing tokenization and lemmatization.

RQ8. Detecting sarcasm [76] related to exoskeletons and studying for any correlation of sarcasm with the sentiment or emotional response associated with the respective Tweets.

RQ9. Identifying the commonly used hashtags associated with Tweets related to exoskeletons and detecting the sentiment related to these respective hashtags [77].

RQ10. Analyzing the Tweets posted by users of exoskeletons to detect the diverse user personas [78] and their associated perspectives, experiences, opinions, and feedback about exoskeletons.

RQ11. Studying trending discussions [79] on Twitter related to exoskeletons and using machine learning algorithms to detect these trends in real-time.

RQ12. Studying the trends in sentiments and emotional responses [80] associated with exoskeletons to track if there is any correlation of the same with the trends in exoskeleton sales or its market potential.

RQ13. Investigating for any interdependence between Tweets about specific exoskeleton products and sales [81] of those specific exoskeleton products.

RQ14. Performing topic modeling [82] of these respective Tweets to interpret the associated communication as news, recommendation, discussion, feedback, perspective, opinion, etc., related to different kinds of exoskeletons.

RQ15. Investigating retweeting patterns of Tweets [83] to determine the interest in certain topics related to exoskeletons expressed in the respective Tweets.

RQ16. Studying the Tweeting patterns and content of the Tweets by various companies or manufacturers of exoskeletons to understand their audience management methodologies which include targeting different audiences, concealing subjects, and maintaining authenticity [84].

RQ17. Detecting the Point of Interest (P.O.I.) of a Tweet [85], which presents high-level location information about a place, to understand the location-specific opinions, perspectives, or attitudes of the public towards exoskeleton technology.

RQ18. Developing a personalized Tweet recommendation system [86] that would present the latest developments in exoskeleton technology, including exoskeletons available for purchase to potential user groups, which may include the elderly, disabled, handicapped, etc.

RQ19. Performing a case study on different kinds of machine learning classifiers [87] to develop sentiment analysis approaches [88] to deduce the best machine learning classifier in terms of performance characteristics for sentiment analysis of Tweets related to exoskeletons.

RQ20. Performing semantic analysis of the content of each Tweet as per the methodology discussed in [89] to determine if any political leaders have influenced the sale or public opinion or perspectives towards a specific kind of exoskeleton.

RQ21. Analysis of Tweets to determine the emergence of exoskeletons [90] in different fields such as healthcare and medicine.

RQ22. Studying the mentions of exoskeleton companies in Tweets to determine the patterns of customer engagement [91] with each of these companies.

RQ23. Investigating Tweets to determine the global [92] and region-specific [93] reason/drive centered around the purchase or use of specific exoskeletons.

RQ24. Development of a Tweet ranking model to present important Tweets [94] to potential end-users or customers of exoskeletons based on their specific needs or interests.

RQ25. Determining how official accounts on Twitter play a role in the propagation and correction of online rumors related to exoskeletons in different geographic locations [95].

RQ26. Studying the semantics of the Tweets to determine how user diversities such as gender differences [96,97] may impact the tweeting patterns and content of Tweets centered around exoskeletons, including their usage and needs.

RQ27. Performing content value analysis [98] of Tweets to filter the most relevant and least relevant Tweets [99] involving exoskeletons.

RQ28. Developing an approach to determine the audience size [100] of any potential Tweet related to a specific exoskeleton that might be helpful for the consumer outreach of exoskeleton companies and manufacturers.

RQ29. Application of the gratification theory [101] on these Tweets to deduce the factors that gratify users related to different use-cases of exoskeletons.

RQ30. Performing a study on the Tweets to investigate the role of news organizations, including regional media, local media, national media, and broadcast news agencies, in the dissemination of the latest developments [102] in the field of exoskeleton technology.

RQ31. Determining the occupation of potential end-users of exoskeletons from their Tweets [103], which may be helpful for companies and/or manufacturers to develop or improvise exoskeletons to better assist these end-users in their respective professions.

RQ32. Implementation of the TCV-Rank summarization technique for generating online summaries and historical summaries related to Tweets [104] about exoskeletons posted from different geographic regions.

RQ33. Implementation of the TURank (Twitter User Rank) algorithm [105] to find authoritative Twitter users who post Tweets related to exoskeletons.

RQ34. Studying the trends of entity linking [106] in Tweets about upcoming exoskeletons for different use-cases.

RQ35. Investigating the impact of following clusters of exoskeleton users or exoskeleton companies [107] on the ideologies of the Twitter user(s) over exoskeletons.

RQ36. Analysis of the Tweets centered around specific hashtags [108] related to exoskeletons to analyze the tweeting trends and replies related to these hashtags.

RQ37. Implementation of the HybridSeg approach [109] to find the optimal segmentation of Tweets related to exoskeletons for improving segmentation quality as well as for exploring applications of this approach for named entity recognition.

RQ38. Interpretation of the use of Twitter by companies or organizations in the exoskeleton industry to examine brand attributes (both product-related and non-product-related) and their relation to Twitter's key engagement features (Reply, Retweet, Favorite) [110].

RQ39. Development of an approach by application of the Latent Dirichlet Allocation (LDA) model as proposed in [111] to deduce the information credibility related to exoskeleton-based Tweets originating from different sources.

RQ40. Studying Tweets to detect and predict any potential conspiracy theories [112,113] related to emerging developments in the field of exoskeletons or any specific exoskeleton-based product.

RQ41. Implementation of the Tweet2Vec method [114] for learning Tweet embeddings using character-level CNN-LSTM encoder-decoder for efficient categorization of Tweets centered around exoskeleton technologies in general or related to any specific exoskeleton technology.

RQ42. Implementation of the Self-Exciting Point Process Model for Predicting Tweet Popularity (SEISMIC) model [115] to predict the popularity of Tweets related to exoskeletons.

RQ43. Studying the geographic diffusion patterns in terms of random, local, and information brokerage of the information contained in a specific Tweet [116] related to exoskeletons and their diverse use cases.

RQ44. Performing Tweet wikification [117] to identify different concepts mentioned in a Tweet to link these concepts to existing concepts about exoskeletons present in a knowledge base, such as Wikipedia.

RQ45. Detecting spam accounts [118] and social spam [119] on Twitter that may be the source of spam related to exoskeleton-based information expressed in Tweets.

RQ46. Development of an approach similar to the work in [120] for detection of complaints related to specific exoskeleton technologies.

RQ47. Predicting the cost [121,122] of planned and expected developments in existing exoskeletons based on drawing insights from the Tweets about these developments.

RQ48. Application of the approach proposed in [123] to detect the patterns of emojis present in information-based Tweets about exoskeletons for the analysis of the relationships between plain texts and emojis usage in such Tweets.

RQ49. Tracking and investigating the usage of multiple emojis expressed in the Tweets related to different use cases of exoskeletons for investigating the associated sentiment [124,125], performing Tweet classification [126], user verification [127], irony detection [128], and trust modeling [129].

RQ50. Detection of fake users [130] who post fake news [131] about exoskeletons on Twitter.

RQ51. Investigating the effect of tweeting about research papers [132] on exoskeletons on the downloads and citations of these respective papers.

RQ52. Determining the social identities [133] of diverse users of exoskeletons based on the content and context of their Tweets.

RQ53. Studying the relevance of a Tweet [134] about a specific exoskeleton based on the hyperlinked documents in the same.

RQ54. Investigating how exoskeleton companies and/or manufacturers use tagging [135] on Twitter for audience engagement and retention.

RQ55. Performing stance detection [136] towards exoskeletons by analyzing the Tweets posted by its users.

RQ56. Interpretation of satire [137] in the context of Tweets about new and upcoming exoskeleton technologies.

RQ57. Predicting the age of existing users or potential users of exoskeletons from their Tweets [138] to personalize the exoskeletons as per the age-specific needs.

RQ58. Investigating the selective attention over different entities expressed in any Tweet pertaining to exoskeletons, as per the methodology proposed in [139].

RQ59. Studying the paradigms of readability in Tweets posted by users of exoskeletons to interpret the degrees of engagement [140].

RQ60. Deducing the best time to Tweet [141] any information related to exoskeletons that might be helpful for the sales and marketing team of exoskeleton companies and/or manufacturers.

RQ61. Tracking repliers and retweeters of Tweets [142] about improvisations in existing exoskeletons posted by exoskeleton companies to detect degrees of intimacy with the target audience.

RQ62. Detecting the number of Tweets [143] related to exoskeletons from a geographic area that could be helpful in understanding the associated needs or public perceptions of a specific exoskeleton-based technology available or marketed in that area.

RQ63. Analyzing the multimodal factors that are associated with the retweet of any Tweet [144] communicating news about exoskeletons.

RQ64. Using the concept of knowledge graphs for Tweet summarization for effectiveness in obtaining useful information [145] related to exoskeleton technologies on Twitter.

RQ65. Recommendation of specific hashtags [146] related to exoskeletons to Twitter users who could be potential users of exoskeletons.

RQ66. Performing contextualization of Tweets [147] related to exoskeletons based on hashtag performance prediction and multi-document summarization.

RQ67. Assigning value to Tweets related to specific use cases of exoskeletons based on the approach proposed in [148] to compute the worth of the underlining Tweets.

RQ68. Deducing the number of followers of exoskeleton companies from their Tweets [149] to determine their customer base.

RQ69. Studying Tweets for the detection of suggestions and classifications of suggestions [150] related to existing and/or emerging technologies associated with exoskeletons.

RQ70. Studying Tweets to interpret any forms of discrimination [151] faced by existing or potential users of exoskeletons.

RQ71. Implementation of the iFACT framework [152] on Tweets associated with exoskeletons to identify, assess, and evaluate the underlying factual information mentioned in the Tweets.

RQ72. Implementation of the SEDTWik framework [153] for segment-based detection of any kinds of events from Tweets that focus on the use of exoskeletons by diverse user groups.

RQ73. Developing an approach as per [154] for followee recommendation to existing and/or potential users of exoskeletons based on topic extraction and sentiment analysis from the Tweets.

RQ74. Studying Tweets for detecting stress levels and reasons for stress [155] in current or potential users of exoskeletons.

RQ75. Detecting if any Tweet about exoskeletons posted by exoskeleton companies can be classified as a "regrettable" Tweet [156] so that these companies may delete the Tweet to reduce the chances of any potential damage to their reputation.

RQ76. Interpreting diverse activities related to use cases of exoskeletons by studying the associated Tweets [157] and mapping these activities on pleasure and arousal dimensions using cognitive computing principles.

RQ77. Studying Tweets posted by users of exoskeletons to monitor their mental health [158].

RQ78. Detecting deception (both positive and negative deception) from Tweets [159] about the use of exoskeletons by specific user groups.

RQ79. Tracking happiness associated with exoskeleton usage in different cities [160] based on studying Tweets related to exoskeletons originating from these cities.

RQ80. Identification of hate speech and abusive language in Tweets [161] made by unsatisfied customers of exoskeletons.

RQ81. Developing an approach as per [162] to filter out relevant Tweets comprising of latest breaking news in the context of exoskeletons.

RQ82. Extracting information from Tweets related to exoskeletons to interpret the multimodal forms of purchase intentions [163] in potential users.

RQ83. Inferring shared interests [164] related to exoskeletons based on studying the Tweets of both current and potential users of exoskeletons.

RQ84. Modeling public mood in different geographic regions [165] towards new advances in exoskeletons based on semantic analysis of the Tweets originating from these respective regions.

RQ85. Implementation of the Categorical Topic Model [166] for extracting categorical topics and emerging issues about exoskeletons from Tweets.

RQ86. Using classification approaches to deduce inundation levels [167] in the context of use case scenarios of different exoskeletons by different user groups.

RQ87. Studying the Tweets to interpret bias and degrees of the same [168] towards using exoskeletons by potential user groups.

RQ88. Detection, classification, and ranking of trending topics [169] related to conversations about exoskeletons on Twitter.

RQ89. Analyzing Tweets about exoskeletons posted by companies that sell one or more types of exoskeletons to study and predict the changes in stock prices of these companies based on their tweeting patterns [170,171].

RQ90. Performing user characterization [172] from the tweeting patterns of any potential user to develop user personas for personalization of exoskeletons.

RQ91. Studying Tweets posted by users of exoskeletons to detect and analyze their feedback and suggestions [173] for possible improvements in different types of exoskeletons.

RQ92. Application of the Similarity Learning Algorithm (SiLA) as proposed in [174] to identify popular Tweets related to current and emerging exoskeletons and their use cases.

RQ93. Implementation of the approach proposed in [175] to study patterns of Tweets related to exoskeletons to detect Tweets that represent "extreme behavior" on social media.

RQ94. Classifying Tweets about specific use cases of exoskeletons as "alarming" and "reassuring" [176] to investigate the views of different user groups.

RQ95. Performing semantic analysis of Tweets posted by new users of exoskeletons to detect instances of euphoria or delusion [177] in the context of the use cases mentioned in the underlining Tweets.

RQ96. Detecting obesity from Tweets [178] posted by users of exoskeletons and investigating any potential correlations between obesity and exoskeleton usage.

RQ97. Classifying potential user groups of exoskeletons into communities [179] based on studying their needs expressed in their Tweets for the development of specific exoskeletons to meet these community-based needs.

RQ98. Estimating demographic information of exoskeleton users from their Tweets [180] to interpret any variation of use cases based on user diversity.

RQ99. Studying the Tweets to deduce the perceptions [181] of exoskeleton users about different exoskeleton companies to interpret their buying behavior.

RQ100. Analyzing the Tweets to track misinformation and trends in the same [182] about upcoming or existing exoskeletons.

As can be seen from these RQs, the prior works that were reviewed [69–182] required Tweets (and in some cases, certain characteristics of Tweets such as retweet counts, hashtags used in the tweets, date, and time of the tweets, etc. that can be obtained upon Hydration) as a data source. In some of these works, the relevant Tweets were mined, and in some of the other works, instead of data mining, a dataset comprising relevant Tweets was used. In this dataset, the presence of about 140,000 Tweet IDs corresponding to the same number of Tweets upholds the fact that it can be directly used as a data source for investigating the above-mentioned research questions about exoskeletons and researchers would not have to spend time developing and implementing the methodology to collect relevant Tweets followed by preprocessing of the same.

4.2. Methodology as a Starting Point for Investigating Some of the RQs

This section describes a methodology that may be considered as starting point for investigating some of the RQs centered around studying the occurrence of a set of words in the Tweets about exoskeletons which could include names of exoskeleton companies, specific exoskeleton products, influential people who are users of exoskeletons, politicians or other popular personalities sharing positive or negative opinions about exoskeletons. This methodology is presented to be implemented in RapidMiner [58], as RapidMiner was utilized both for the development of this dataset as well as for the statistical analysis of the same. It is also worth mentioning here that following this methodology in RapidMiner as a starting point is just one out of the many ways by which such RQs may be investigated. This methodology would also require modifications if any of the other RQs are being investigated. Furthermore, this methodology cannot be directly applied for investigating such RQs if RapidMiner is not being used as the application development platform. For the development of this methodology, a few extensions would have to be installed in RapidMiner Studio. These include Text Processing 9.3 (or a higher version if available) by RapidMiner [183], Aylien Text Analysis 0.2.0 (or a higher version if available) by Aylien

Inc. [184], and String Matching 1.0.0 (or a higher version if available) by Aptus Data Labs [185]. These extensions are available for free in the RapidMiner Marketplace. The steps to be followed are mentioned next:

1. Hydrate all the Tweet IDs and merge the results into a single .csv file. Import this .csv file as a "New Dataset" into RapidMiner Studio.
2. Develop a "Bag of Words" model in RapidMiner Studio, which would act as a collection of keywords and/or phrases of interest related to exoskeletons. These could include names of exoskeleton companies, specific exoskeleton products, influential people who are users of exoskeletons, politicians, celebrities, or other popular personalities sharing positive or negative opinions about exoskeletons.
3. Use the "Select Attribute" *operator* to select the text of the Tweets from the dataset. Identify and remove stop words from the Tweet texts.
4. Use the "Data Filters" in RapidMiner to filter out unwanted attributes from the dataset to have only the Tweets and other essential information needed for this study. Here unwanted attributes refer to non-essential details associated with each of the Tweets, such as the default profile image of the users, description mentioned on each user's profile, follower count of the users, location information of all the users, screenname of all the users, and verified status of all the user accounts, which would be obtained upon hydration using the Hydrator app.
5. Using the "Read Document" *operator*, set up a path to a file on the local system that contains a set of keywords or phrases that would be used for checking similarity and/or occurrence (Step 2). It is recommended that this file is a .txt file.
6. Implement the Levenshtein distance algorithm [186] using the "Fuzzy matching" *operator*.
7. Provide the output from the *operator* in Step 4 as the source and the "Read Document" *operator* (Step 5) as the grounds for comparison to generate similarity scores based on the string-comparison of each Tweet.
8. Enable the advanced parameters of the "Fuzzy matching" *operator* to define the threshold value. This can be any user-defined value, and only those Tweets that have a similarity (indicating occurrence) greater than the threshold would be retained in the results.
9. Integrate all the above *operators* and develop a RapidMiner process and set up a Twitter connection in RapidMiner Studio.
10. Run the process to compute the results after defining specific parameters in Step 2 and Step 8.

As mentioned earlier in this Section, this step-by-step process would serve as a starting point for investigating certain RQs only if RapidMiner is used as the application development platform. If RapidMiner is not used as the application development platform, then the following are a few resources such as libraries and/or packages in different programming languages such as Python, R, Java, C++, JavaScript, Scala, Rust, Clojure, and Ruby, which may be used to develop a similar methodology for investigating one or more of these RQs.

- Python: Natural Language Toolkit [187], Spacy [188], TextBlob [189], Stanford Core NLP [190], PyNLPI [191], SciPY [192], Scikit-Learn [193], Keras [194], PyTorch [195], and Pandas [196].
- R: koRpus [197], OpenNLP [198], Quanteda [199], RWeka [200], Spacyr [201], Stringr [202], Text2vec [203], and Text Mining Package [204].
- Java: Apache OpenNLP [205], Apache UIMA [206], General Architecture for Text Engineering [207], LingPipe [208], Machine Learning for Language Toolkit (Mallet) [209], Natural Language Processing for JVM languages [210], and Apache Lucene [211].
- C++: MIT Information Extraction [212], MeTa [213], CRF++ [214], Colibri Core [215], InsNet [216], and Libfolia [217].
- JavaScript: Twitter Text [218], Knwl.js [219], Poplar [220], nlp.js [221], and Node Question Answering [222].
- Scala: Saul [223], ATR4S [224], Word2Vec-Sala [225], Epic [226], and TM [227].
- Rust: whatlang [228], Snips-NLU [229], and Rust-BERT [230].

- Clojure: Clojure-OpenNLP [231], Inflections-CLJ [232], and Postagga [233].
- Ruby: MonkeyLearn-Ruby [234], DialogFlow Ruby Client [235], FastText-Ruby [236], Ruby-WordNet [237], Ruby-Fann [238], TensorFlow.Rb [239], and the Ruby Language Toolkit [240].

5. Conclusions

The exoskeleton technology has been rapidly advancing in the last few years on account of its multitude of applications and diverse use cases in assisted living, military, healthcare, firefighting, and industry 4.0. The exoskeleton market is projected to increase by multiple times of its current value within the next two years. Therefore, it is crucial to study the trends of user interest, views, opinions, perspectives, attitudes, acceptance, feedback, buying behavior, and satisfaction, towards exoskeletons, for which the availability of Big Data of conversations about exoskeletons is necessary. In today's Internet of Everything era, the use of social media platforms has skyrocketed in the recent past as social media platforms provide a sense of "community" where people develop "virtual" relationships and converse on diverse topics, which includes emerging technologies such as exoskeletons. Twitter, popular amongst users of all age groups, is the second most visited social media platform, and its popularity has constantly been increasing in the last few years. Researchers from different disciplines have worked on developing datasets by mining this Big Data of Twitter to record, study, interpret, and analyze conversations on Twitter related to different emerging technologies, topics, applications, matters of global concern, diseases, viruses, events, disasters, and so on, which shows the immense potential, relevance, importance, and applicability of mining of Twitter Big Data.

Even though there have been several advances in the field of exoskeleton research in the last decade and a half, no prior work in this field or in the field of social media research has focused on the development of a dataset of conversations on Twitter related to exoskeletons. The work presented in this paper addresses this research challenge. It presents an open-access dataset of 138,584 Tweet IDs corresponding to 138,584 Tweets about exoskeletons posted on Twitter for a period of 5-years from 21 May 2017 to 21 May 2022. In addition to this, based on a comprehensive review of 108 recent works in the fields of Big Data, Natural Language Processing, Information Retrieval, Data Mining, Pattern Recognition, and Artificial Intelligence, that may be applied to relevant Twitter data for advancing the field of exoskeleton research, this paper presents 100 Research Questions for researchers to investigate, analyze, ideate, and explore by using this dataset. Future works would involve updating this dataset to incorporate more recent Tweets, as well as investigating these research questions and developing new ones to advance research and development in this field.

References

1.   Olar, M.-L.; Leba, M.; Risteiu, M. Exoskeleton—Wearable Devices. Literature Review. *MATEC Web Conf.* **2021**, *342*, 5005. [CrossRef]

2.   Yang, C.-J.; Zhang, J.-F.; Chen, Y.; Dong, Y.-M.; Zhang, Y. A Review of Exoskeleton-Type Systems and Their Key Technologies. *Proc. Inst. Mech. Eng. Part C* **2008**, *222*, 1599–1612. [CrossRef]

3.   Palazzi, E.; Luzi, L.; Dimo, E.; Calanca, A. An Affordable Upper-Limb Exoskeleton Concept for Rehabilitation Applications. *Technologies* **2022**, *10*, 22. [CrossRef]

4.   Laubscher, C.A.; Goo, A.; Farris, R.J.; Sawicki, J.T. Hybrid Impedance-Sliding Mode Switching Control of the Indego Explorer Lower-Limb Exoskeleton in Able-Bodied Walking. *J. Intell. Robot. Syst.* **2022**, *104*, 76. [CrossRef]

5.   Sarkisian, S.V.; Ishmael, M.K.; Lenzi, T. Self-Aligning Mechanism Improves Comfort and Performance with a Powered Knee Exoskeleton. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 629–640. [CrossRef] [PubMed]

6.   van der Have, A.; Rossini, M.; Rodriguez-Guerrero, C.; Van Rossom, S.; Jonkers, I. The Exo4Work Shoulder Exoskeleton Effectively Reduces Muscle and Joint Loading during Simulated Occupational Tasks above Shoulder Height. *Appl. Ergon.* **2022**, *103*, 103800. [CrossRef] [PubMed]

7.   Zahedi, A.; Wang, Y.; Martinez-Hernandez, U.; Zhang, D. A Wearable Elbow Exoskeleton for Tremor Suppression Equipped with Rotational Semi-Active Actuator. *Mech. Syst. Signal Process.* **2021**, *157*, 107674. [CrossRef]

8.   Peng, X.; Acosta-Sojo, Y.; Wu, M.I.; Stirling, L. Actuation Timing Perception of a Powered Ankle Exoskeleton and Its Associated Ankle Angle Changes during Walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 869–877. [CrossRef] [PubMed]

9.   Liu, H.; Zeng, B.; Liu, X.; Zhu, X.; Song, H. Detection of Human Lifting State Based on Long Short-Term Memory for Wearable Waist Exoskeleton. In *Lecture Notes in Electrical Engineering*; Springer Singapore: Singapore, 2022; pp. 301–310.

10. Ishmael, M.K.; Archangeli, D.; Lenzi, T. A Powered Hip Exoskeleton with High Torque Density for Walking, Running, and Stair Ascent. *IEEE ASME Trans. Mechatron.* **2022**, 1–12. [CrossRef]

11. Garosi, E.; Mazloumi, A.; Jafari, A.H.; Keihani, A.; Shamsipour, M.; Kordi, R.; Kazemi, Z. Design and Ergonomic Assessment of a Passive Head/Neck Supporting Exoskeleton for Overhead Work Use. *Appl. Ergon.* **2022**, *101*, 103699. [CrossRef] [PubMed]

12. Song, J.; Zhu, A.; Tu, Y.; Zou, J. Multijoint Passive Elastic Spine Exoskeleton for Stoop Lifting Assistance. *Int. J. Adv. Robot. Syst.* **2021**, *18*, 172988142110620. [CrossRef]

13. Dragusanu, M.; Iqbal, M.Z.; Baldi, T.L.; Prattichizzo, D.; Malvezzi, M. Design, Development, and Control of a Hand/Wrist Exoskeleton for Rehabilitation and Training. *IEEE Trans. Robot.* **2022**, *38*, 1472–1488. [CrossRef]

14. Athira; Oommen, R.M. Advancements in Robotic Exoskeleton. In Proceedings of the 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kottayam, India, 21–22 December 2018; pp. 1–3.

15. Li, G.; Cheng, L.; Sun, N. Design, Manipulability Analysis and Optimization of an Index Finger Exoskeleton for Stroke Rehabilitation. *Mech. Mach. Theory* **2022**, *167*, 104526. [CrossRef]

16. Guntara, A.; Rahyussalim, A.J. The Uses of Lower Limb Exoskeleton, Functional Electrical Stimulation, and Future Improvements for Leg Paralysis Management—A Systematic Review. In Proceedings of the 5th International Symposium of Biomedical Engineering (ISBE) 2020, Depok, Indonesia, 28–29 July 2021.

17. Thamsuwan, O.; Milosavljevic, S.; Srinivasan, D.; Trask, C. Potential Exoskeleton Uses for Reducing Low Back Muscular Activity during Farm Tasks. *Am. J. Ind. Med.* **2020**, *63*, 1017–1028. [CrossRef] [PubMed]

18. Kumar, V.; Hote, Y.V.; Jain, S. Review of Exoskeleton: History, Design and Control. In Proceedings of the 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE), Noida, India, 10–11 October 2019; pp. 677–682.

19. Coren, M.J. Robot Exoskeletons are Finally Here, and They're Nothing Like the Suits from Iron Man. Available online: https://qz.com/971741/robot-exoskeletons-are-finally-here-and-theyre-nothing-like-the-suits-from-iron-man/ (accessed on 26 May 2022).

20. Global Market Insights; Inc Exoskeleton Market Worth $3.4bn by 2024: Global Market Insights, Inc. Available online: https://www.globenewswire.com/en/news-release/2017/08/30/1104254/0/en/Exoskeleton-Market-worth-3-4bn-by-2024-Global-Market-Insights-Inc.html (accessed on 26 May 2022).

21. da Costa, V.C.F.; Oliveira, L.; de Souza, J. Internet of Everything (IoE) Taxonomies: A Survey and a Novel Knowledge-Based Taxonomy. *Sensors* **2021**, *21*, 568. [CrossRef] [PubMed]

22. Radetić-Paić, M.; Boljunčić, V. The Causes of I.C.T. Use Which Increase Time Spent on the Internet by Secondary School Students and Affect Exposure to Bullying from Other Students. *Econ. Res.* **2021**, *35*, 2859–2867. [CrossRef]

23. Pan, Y.-C.; Chiu, Y.-C.; Lin, Y.-H. Systematic Review and Meta-Analysis of Epidemiology of Internet Addiction. *Neurosci. Biobehav. Rev.* **2020**, *118*, 612–622. [CrossRef] [PubMed]

24. Boulianne, S. Social Media Use and Participation: A Meta-Analysis of Current Research. *Inf. Commun. Soc.* **2015**, *18*, 524–538. [CrossRef]

25. Gruzd, A.; Haythornthwaite, C. Enabling Community through Social Media. *J. Med. Internet Res.* **2013**, *15*, e248. [CrossRef] [PubMed]

26. Shepherd, A.; Sanders, C.; Doyle, M.; Shaw, J. Using Social Media for Support and Feedback by Mental Health Service Users: Thematic Analysis of a Twitter Conversation. *BMC Psychiatry* **2015**, *15*, 29. [CrossRef]

27. Kavada, A. Social Media as Conversation: A Manifesto. *Soc. Media Soc.* **2015**, *1*, 205630511558079. [CrossRef]

28. Goldberg, S.C. The Promise and Pitfalls of Online' Conversations'. *Roy Inst. Philos. Suppl.* **2021**, *89*, 177–193. [CrossRef]
29. Ramnarain, Y.; Govender, K. Social Media Browsing and Consumer Behaviour: Exploring the Youth Market. *Afr. J. Bus. Manag.* **2013**, *7*, 1885–1893. [CrossRef]
30. Awan, M.J.; Rahim, M.S.M.; Nobanee, H.; Munawar, A.; Yasin, A.; Azlanmz, A.M.Z. Social Media and Stock Market Prediction: A Big Data Approach. *Comput. Mater. Contin.* **2021**, *67*, 2569–2583. [CrossRef]
31. Pezzuti, T.; Leonhardt, J.M.; Warren, C. Certainty in Language Increases Consumer Engagement on Social Media. *J. Interact. Mark.* **2021**, *53*, 32–46. [CrossRef]
32. Wang, L.; Lee, J.H. The Impact of K-Beauty Social Media Influencers, Sponsorship, and Product Exposure on Consumer Acceptance of New Products. *Fash. Text.* **2021**, *8*, 15. [CrossRef]
33. Varghese, M.S.; Agrawal, M.M. Impact of Social Media on Consumer Buying Behavior. *Saudi J. Bus. Manag. Stud.* **2021**, *6*, 51–55. [CrossRef]
34. Majeed, M.; Asare, C.; Fatawu, A.; Abubakari, A. An Analysis of the Effects of Customer Satisfaction and Engagement on Social Media on Repurchase Intention in the Hospitality Industry. *Cogent Bus. Manag.* **2022**, *9*, 2028331. [CrossRef]
35. 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, Virtual Event, 7–9 July 2021*; SCITEPRESS—Science and Technology Publications: Setubal, Portugal, 2021.
36. Gruzd, A.; Wellman, B.; Takhteyev, Y. Imagining Twitter as an Imagined Community. *Am. Behav. Sci.* **2011**, *55*, 1294–1318. [CrossRef]
37. Aslam, S. Twitter by the Numbers (2022): Stats, Demographics & Fun Facts. 2022. Available online: https://www.omnicoreagency.com/ (accessed on 27 May 2022).
38. Dooms, S.; De Pessemier, T.; Martens, L. MovieTweetings: 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 A.C.M. Conference on Recommender Systems (RecSys 2013), Hong Kong, 12–16 October 2013.
39. Banda, J.M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Chowell, G. A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration. *Epidemiologia* **2020**, *2*, 315–324. [CrossRef]
40. Chen, E.; Deb, A.; Ferrara, E. #Election2020: The First Public Twitter Dataset on the 2020 U.S. Presidential Election. *J. Comput. Soc. Sci.* **2022**, *5*, 1–18. [CrossRef]
41. 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 *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2020; pp. 427–439, ISBN 9783030609740.
42. 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.
43. Meng, L.; Dong, Z.S. Natural Hazards Twitter Dataset. *arXiv* **2020**, arXiv:2004.14456. [CrossRef]
44. Salem, M.S.; Ismail, S.S.; Aref, M. Personality Traits for Egyptian Twitter Users Dataset. In Proceedings of the 2019 8th International Conference on Software and Information Engineering, Cairo, Egypt, 9–12 April 2019.
45. 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.
46. Tekumalla, R.; Banda, J.M. A Large-Scale Twitter Dataset for Drug Safety Applications Mined from Publicly Existing Resources. *arXiv* **2020**, arXiv:2003.13900. [CrossRef]
47. Effrosynidis, D.; Karasakalidis, A.I.; Sylaios, G.; Arampatzis, A. The Climate Change Twitter Dataset. *Expert Syst. Appl.* **2022**, *204*, 117541. [CrossRef]
48. Febriana, T.; Budiarto, A. Twitter Dataset for Hate Speech and Cyberbullying Detection in Indonesian Language. In Proceedings of the 2019 International Conference on Information Management and Technology (ICIMTech), Denpasar, Indonesia, 19–20 August 2019; Volume 1, pp. 379–382.
49. 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; pp. 146–149.
50. Schroeder, D.; Schaal, F.; Filkukova, P.; Pogorelov, K.; Langguth, J. WICO Graph: A Labeled Dataset of Twitter Subgraphs Based on Conspiracy Theory and 5G-Corona Misinformation Tweets. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence, Virtual Event, 7–9 July 2021*; SCITEPRESS—Science and Technology Publications: Setubal, Portugal, 2021.
51. Stemmer, M.; Parmet, Y.; Ravid, G. What Are IBD Patients Talking about on Twitter? In *I.C.T. for Health, Accessibility and Wellbeing*; Springer International Publishing: Cham, Switzerland, 2021; pp. 206–220, ISBN 9783030942083.
52. Warren, E. Strengthening Research through Data Sharing. *N. Engl. J. Med.* **2016**, *375*, 401–403. [CrossRef]
53. Fecher, B.; Friesike, S.; Hebing, M. What Drives Academic Data Sharing? *PLoS ONE* **2015**, *10*, e0118053. [CrossRef]
54. Logan, J.A.R.; Hart, S.A.; Schatschneider, C. Data Sharing in Education Science. *AERA Open* **2021**, *7*, 233285842110064. [CrossRef]
55. Privacy Policy. Available online: https://twitter.com/en/privacy/previous/version_15 (accessed on 27 May 2022).

56. Developer Agreement and Policy. Available online: https://developer.twitter.com/en/developer-terms/agreement-and-policy (accessed on 27 May 2022).

57. RapidMiner GmbH Search Twitter—RapidMiner Documentation. Available online: https://docs.rapidminer.com/latest/studio/operators/data_access/applications/twitter/search_twitter.html (accessed on 27 May 2022).

58. 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.

59. Rate Limits: Standard v1.1. Available online: https://developer.twitter.com/en/docs/twitter-api/v1/rate-limits (accessed on 27 May 2022).

60. Using Standard Search. Available online: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/guides/standard-operators (accessed on 27 May 2022).

61. 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]

62. Lamsal, R. Hydrating Tweet I.D.s. Available online: https://theneuralblog.com/hydrating-tweet-ids/ (accessed on 27 May 2022).

63. Bramus. Accessing a Tweet Using Only Its ID (and Without the Twitter API). Available online: https://www.bram.us/2017/11/22/accessing-a-tweet-using-only-its-id-and-without-the-twitter-api/ (accessed on 27 May 2022).

64. Hydrator. Available online: https://github.com/DocNow/hydrator (accessed on 27 May 2022).

65. Tekumalla, R.; Banda, J.M. Social Media Mining Toolkit (SMMT). *Genom. Inform.* **2020**, *18*, e16. [CrossRef]

66. Twarc. Available online: https://github.com/docnow/twarc (accessed on 27 May 2022).

67. Hydrator Versions. Available online: https://github.com/docnow/hydrator/releases (accessed on 27 May 2022).

68. ISO 639. Available online: https://www.iso.org/iso-639-language-codes.html (accessed on 10 July 2022).

69. Carvalho, J.; Plastino, A. On the Evaluation and Combination of State-of-the-Art Features in Twitter Sentiment Analysis. *Artif. Intell. Rev.* **2021**, *54*, 1887–1936. [CrossRef]

70. Gu, S.; Wang, F.; Patel, N.P.; Bourgeois, J.A.; Huang, J.H. A Model for Basic Emotions Using Observations of Behavior in Drosophila. *Front. Psychol.* **2019**, *10*, 781. [CrossRef]

71. Do, H.H.; Prasad, P.W.C.; Maag, A.; Alsadoon, A. Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Syst. Appl.* **2019**, *118*, 272–299. [CrossRef]

72. Asur, S.; Huberman, B.A.; Szabo, G.; Wang, C. Trends in Social Media: Persistence and Decay. *SSRN Electron. J.* **2011**, *5*, 434–437. [CrossRef]

73. Fouad, M.M.; Mahany, A.; Aljohani, N.; Abbasi, R.A.; Hassan, S.-U. ArWordVec: Efficient Word Embedding Models for Arabic Tweets. *Soft Comput.* **2020**, *24*, 8061–8068. [CrossRef]

74. Chen, G.M. Tweet This: A Uses and Gratifications Perspective on How Active Twitter Use Gratifies a Need to Connect with Others. *Comput. Hum. Behav.* **2011**, *27*, 755–762. [CrossRef]

75. Hong, L.; Dan, O.; Davison, B.D. Predicting Popular Messages in Twitter. In Proceedings of the 20th International Conference Companion on World Wide Web—W.W.W., Hyderabad, India, 28 March–1 April 2011.

76. Rajadesingan, A.; Zafarani, R.; Liu, H. Sarcasm Detection on Twitter: A Behavioral Modeling Approach. In Proceedings of the Eighth A.C.M. International Conference on Web Search and Data Mining—WSDM '15, Shanghai, China, 2–6 February 2015.

77. Wang, X.; Wei, F.; Liu, X.; Zhou, M.; Zhang, M. Topic Sentiment Analysis in Twitter: A Graph-Based Hashtag Sentiment Classification Approach. In Proceedings of the 20th A.C.M. International Conference on Information and Knowledge Management—CIKM '11, Glasgow, UK, 24–28 October 2011.

78. Li, J.; Galley, M.; Brockett, C.; Spithourakis, G.P.; Gao, J.; Dolan, B. A Persona-Based Neural Conversation Model. *arXiv* **2016**, arXiv:1603.06155. [CrossRef]

79. Aiello, L.M.; Petkos, G.; Martin, C.; Corney, D.; Papadopoulos, S.; Skraba, R.; Goker, A.; Kompatsiaris, I.; Jaimes, A. Sensing Trending Topics in Twitter. *IEEE Trans. Multimed.* **2013**, *15*, 1268–1282. [CrossRef]

80. Lee, K.; Palsetia, D.; Narayanan, R.; Patwary, M.M.A.; Agrawal, A.; Choudhary, A. Twitter Trending Topic Classification. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011; pp. 251–258.

81. Dijkman, R.; Ipeirotis, P.; Aertsen, F.; van Helden, R. Using Twitter to Predict Sales: A Case Study. *arXiv* **2015**, arXiv:1503.04599. [CrossRef]

82. 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, Phoenix, AZ, USA, 12–13 February 2016.

83. 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; pp. 1–10.

84. Marwick, A.E.; Boyd, D. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media Soc.* **2011**, *13*, 114–133. [CrossRef]

85. Li, W.; Serdyukov, P.; de Vries, A.P.; Eickhoff, C.; Larson, M. The Where in the Tweet. In Proceedings of the 20th A.C.M. International Conference on Information and Knowledge Management–CIKM '11, Glasgow, UK, 24–28 October 2011.

86.   Chen, K.; Chen, T.; Zheng, G.; Jin, O.; Yao, E.; Yu, Y. Collaborative Personalized Tweet Recommendation. In Proceedings of the 35th international ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR '12, Portland, OR, USA, 12–16 August 2012.

87.   Ray, S. A Quick Review of Machine Learning Algorithms. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 35–39.

88.   da Silva, N.F.F.; Hruschka, E.R.; Hruschka, E.R., Jr. Tweet Sentiment Analysis with Classifier Ensembles. *Decis. Support Syst.* **2014**, *66*, 170–179. [CrossRef]

89.   Kreis, R. The "Tweet Politics" of President Trump. *J. Lang. Politics* **2017**, *16*, 607–618. [CrossRef]

90.   Myslín, M.; Zhu, S.-H.; Chapman, W.; Conway, M. Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products. *J. Med. Internet Res.* **2013**, *15*, e174. [CrossRef]

91.   Wigley, S.; Lewis, B.K. Rules of Engagement: Practice What You Tweet. *Public Relat. Rev.* **2012**, *38*, 165–167. [CrossRef]

92.   Liu, I.L.B.; Cheung, C.M.K.; Lee, M.K.O. Understanding Twitter Usage: What Drive People Continue to Tweet. In Proceedings of the Pacific Asia Conference on Information Systems, PACIS 2010, Taipei, Taiwan, 9–12 July 2010.

93.   Cheng, Z.; Caverlee, J.; Lee, K. You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users. In Proceedings of the 19th A.C.M. international conference on Information and knowledge management—CIKM '10, Toronto, ON, Canada, 26–30 October 2010.

94.   Uysal, I.; Croft, W.B. User Oriented Tweet Ranking: A Filtering Approach to Microblogs. In Proceedings of the 20th A.C.M. International Conference on Information and Knowledge Management—CIKM '11, Glasgow, UK, 24–28 October 2011.

95.   Andrews, C.A.; Fichet, E.S.; Ding, Y.; Spiro, E.S.; Starbird, K. Keeping up with the Tweet-Dashians: The Impact of 'official-Accounts on Online Rumoring. In Proceedings of the 19th A.C.M. Conference on Computer-Supported Cooperative Work & Social Computing—CSCW '16, San Francisco, CA, USA, 27 February–2 March 2016.

96.   Pujazon-Zazik, M.; Park, M.J. To Tweet, or Not to Tweet: Gender Differences and Potential Positive and Negative Health Outcomes of Adolescents' Social Internet Use. *Am. J. Mens. Health* **2010**, *4*, 77–85. [CrossRef] [PubMed]

97.   Merler, M.; Cao, L.; Smith, J.R. You Are What You Tweet . . . pic! Gender Prediction Based on Semantic Analysis of Social Media Images. In Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 29 June–3 July 2015; pp. 1–6.

98.   André, P.; Bernstein, M.; Luther, K. Who Gives a Tweet? Evaluating Microblog Content Value. In Proceedings of the A.C.M. 2012 Conference on Computer Supported Cooperative Work–CSCW '12, Seattle, WA, USA, 11–15 February 2012.

99.   Tao, K.; Abel, F.; Hauff, C.; Houben, G.-J. What Makes a Tweet Relevant for a Topic? Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.309.8507&rep=rep1&type=pdf (accessed on 27 May 2022).

100. Kupavskii, A.; Umnov, A.; Gusev, G.; Serdyukov, P. Predicting the Audience Size of a Tweet. *ICWSM* **2013**, *7*, 693–696.

101. Han, S.; Min, J.; Lee, H. Antecedents of Social Presence and Gratification of Social Connection Needs in S.N.S.: A Study of Twitter Users and Their Mobile and Non-Mobile Usage. *Int. J. Inf. Manag.* **2015**, *35*, 459–471. [CrossRef]

102. Armstrong, C.L.; Gao, F. Now Tweet This: How News Organizations Use Twitter. *Electron. News* **2010**, *4*, 218–235. [CrossRef]

103. Hu, T.; Xiao, H.; Nguyen, T.-V.T.; Luo, J. What the Language You Tweet Says about Your Occupation. *arXiv* **2017**, arXiv:1701.06233. [CrossRef]

104. Shou, L.; Wang, Z.; Chen, K.; Chen, G. Sumblr: Continuous Summarization of Evolving Tweet Streams. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 28 July–1 August 2013.

105. Yamaguchi, Y.; Takahashi, T.; Amagasa, T.; Kitagawa, H. TURank: Twitter User Ranking Based on User-Tweet Graph Analysis. In *Web Information Systems Engineering–WISE 2010*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 240–253, ISBN 9783642176159.

106. Guo, S.; Chang, M.-W.; Kıcıman, E. To Link or Not to Link? A Study on End-to-End Tweet Entity Linking. Available online: https://aclanthology.org/N13-1122.pdf (accessed on 27 May 2022).

107. Himelboim, I.; McCreery, S.; Smith, M. Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *J. Comput. Mediat. Commun.* **2013**, *18*, 40–60. [CrossRef]

108. Bruns, A. How Long is a Tweet? Mapping Dynamic Conversation Networks Ontwitterusing Gawk and Gephi. *Inf. Commun. Soc.* **2012**, *15*, 1323–1351. [CrossRef]

109. Li, C.; Sun, A.; Weng, J.; He, Q. Tweet Segmentation and Its Application to Named Entity Recognition. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 558–570. [CrossRef]

110. Parganas, P.; Anagnostopoulos, C.; Chadwick, S. ' You'Ll Never Tweet Alone': Managing Sports Brands through Social Media. *J. Brand Manag.* **2015**, *22*, 551–568. [CrossRef]

111. 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–W.W.W. '15 Companion, Florence, Italy, 18–22 May 2015.

112. Stephens, M. A Geospatial Infodemic: Mapping Twitter Conspiracy Theories of COVID-19. *Dialogues Hum. Geogr.* **2020**, *10*, 276–281. [CrossRef]

113. Fong, A.; Roozenbeek, J.; Goldwert, D.; Rathje, S.; van der Linden, S. The Language of Conspiracy: A Psychological Analysis of Speech Used by Conspiracy Theorists and Their Followers on Twitter. *Group Process. Intergroup Relat.* **2021**, *24*, 606–623. [CrossRef]

114. Vosoughi, S.; Vijayaraghavan, P.; Roy, D. Tweet2Vec: Learning Tweet Embeddings Using Character-Level CNN-LSTM Encoder-Decoder. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, Pisa, Italy, 17–21 July 2016.

115. Zhao, Q.; Erdogdu, M.A.; He, H.Y.; Rajaraman, A.; Leskovec, J. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '15, Sydney, Australia, 10–13 August 2015.

116. van Liere, D. How Far Does a Tweet Travel? Information Brokers in the Twitterverse. In Proceedings of the International Workshop on Modeling Social Media—M.S.M. '10, Toronto, ON, Canada, 13 June 2010.

117. Huang, H.; Cao, Y.; Huang, X.; Ji, H.; Lin, C.-Y. Collective Tweet Wikification Based on Semi-Supervised Graph Regularization. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014.

118. Alom, Z.; Carminati, B.; Ferrari, E. Detecting Spam Accounts on Twitter. In Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, 28–31 August 2018; pp. 1191–1198.

119. Wang, B.; Zubiaga, A.; Liakata, M.; Procter, R. Making the Most of Tweet-Inherent Features for Social Spam Detection on Twitter. *arXiv* **2015**, arXiv:1503.07405. [CrossRef]

120. Purwarianti, A.; Andhika, A.; Wicaksono, A.F.; Afif, I.; Ferdian, F. InaNLP: Indonesia Natural Language Processing Toolkit, Case Study: Complaint Tweet Classification. In Proceedings of the 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA), Penang, Malaysia, 16–19 August 2016; pp. 1–5.

121. Pant, D.R.; Neupane, P.; Poudel, A.; Pokhrel, A.K.; Lama, B.K. Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. In Proceedings of the 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Kathmandu, Nepal, 25–27 October 2018; pp. 128–132.

122. Jain, A.; Tripathi, S.; Dwivedi, H.D.; Saxena, P. Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. In Proceedings of the 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2–4 August 2018; pp. 1–7.

123. Wu, C.; Wu, F.; Wu, S.; Huang, Y.; Xie, X. Tweet Emoji Prediction Using Hierarchical Model with Attention. In Proceedings of the 2018 A.C.M. International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 8–12 October 2018.

124. Tomihira, T.; Otsuka, A.; Yamashita, A.; Satoh, T. What Does Your Tweet Emotion Mean? Neural Emoji Prediction for Sentiment Analysis. In Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services—iiWAS2018, Yogyakarta, Indonesia, 19–21 November 2018.

125. Bansal, B.; Srivastava, S. Lexicon-Based Twitter Sentiment Analysis for Vote Share Prediction Using Emoji and N-Gram Features. *Int. J. Web Based Communities* **2019**, *15*, 85. [CrossRef]

126. Singh, A.; Blanco, E.; Jin, W. Incorporating Emoji Descriptions Improves Tweet Classification. In Proceedings of the 2019 Conference of the North, Minneapolis, MN, USA, 2–7 June 2019; pp. 2096–2101.

127. Suman, C.; Saha, S.; Bhattacharyya, P.; Chaudhari, R.S. Emoji Helps! A Multi-Modal Siamese Architecture for Tweet User Verification. *Cognit. Comput.* **2021**, *13*, 261–276. [CrossRef]

128. Reyes, A.; Rosso, P.; Veale, T. A Multidimensional Approach for Detecting Irony in Twitter. *Lang. Resour. Eval.* **2013**, *47*, 239–268. [CrossRef]

129. Mendoza, M.; Poblete, B.; Castillo, C. Twitter under Crisis: Can We Trust What We RT? In Proceedings of the First Workshop on Social Media Analytics–SOMA '10, Washington, DC, USA, 25 July 2010.

130. Ersahin, B.; Aktas, O.; Kilinc, D.; Akyol, C. Twitter Fake Account Detection. In Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK), London, UK, 5–8 October 2017; pp. 388–392.

131. Saez-Trumper, D. Fake Tweet Buster: A Webtool to Identify Users Promoting Fake News Ontwitter. In Proceedings of the 25th A.C.M. Conference on Hypertext and Social Media, Santiago de Chile, Chile, 1–4 September 2014.

132. Tonia, T.; Van Oyen, H.; Berger, A.; Schindler, C.; Künzli, N. If I Tweet Will You Cite? The Effect of Social Media Exposure of Articles on Downloads and Citations. *Int. J. Public Health* **2016**, *61*, 513–520. [CrossRef] [PubMed]

133. Huang, B.; Carley, K.M. Discover Your Social Identity from What You Tweet: A Content Based Approach. In *Lecture Notes in Social Networks*; Springer International Publishing: Cham, Switzerland, 2020; pp. 23–37, ISBN 9783030426989.

134. McCreadie, R.; Macdonald, C. Relevance in Microblogs: Enhancing Tweet Retrieval Using Hyperlinked Documents. In Proceedings of the 10th Conference on Open Research Areas in Information Retrieval, Lisbon, Portugal, 15–17 May 2013; Le Centre De Hautes Etudes Internationales D'informatique Documentaire: Paris, France, 2013; pp. 189–196.

135. Haugh, B.R.; Watkins, B. Tag Me, Tweet Me If You Want to Reach Me: An Investigation into How Sports Fans Use Social Media. *Int. J. Sport Commun.* **2016**, *9*, 278–293. [CrossRef]

136. Darwish, K.; Stefanov, P.; Aupetit, M.; Nakov, P. Unsupervised User Stance Detection on Twitter. *arXiv* **2019**, arXiv:1904.02000. [CrossRef]

137. Salas-Zárate, M.D.P.; Paredes-Valverde, M.A.; Rodriguez-García, M.Á.; Valencia-García, R.; Alor-Hernández, G. Automatic Detection of Satire in Twitter: A Psycholinguistic-Based Approach. *Knowl. Based Syst.* **2017**, *128*, 20–33. [CrossRef]

138. Pandya, A.; Oussalah, M.; Monachesi, P.; Kostakos, P. On the Use of Distributed Semantics of Tweet Metadata for User Age Prediction. *Future Gener. Comput. Syst.* **2020**, *102*, 437–452. [CrossRef]

139. Ran, C.; Shen, W.; Wang, J. An Attention Factor Graph Model for Tweet Entity Linking. In Proceedings of the 2018 World Wide Web Conference on World Wide Web—W.W.W. '18, Lyon, France, 23–27 April 2018.

140. Davis, S.W.; Horváth, C.; Gretry, A.; Belei, N. Say What? How the Interplay of Tweet Readability and Brand Hedonism Affects Consumer Engagement. *J. Bus. Res.* **2019**, *100*, 150–164. [CrossRef]

141. Al Abdullatif, A.M.; Alsoghayer, R.A.; AlMajhad, E.M. An Algorithm to Find the Best Time to Tweet. In Proceedings of the International Conference on Computer Vision and Image Analysis Applications, Sousse, Tunisia, 18–20 January 2015; pp. 1–13.

142. Yuan, N.J.; Zhong, Y.; Zhang, F.; Xie, X.; Lin, C.-Y.; Rui, Y. Who Will Reply to/Retweet This Tweet? The Dynamics of Intimacy from Online Social Interactions. In Proceedings of the Ninth A.C.M. International Conference on Web Search and Data Mining, San Francisco, CA, USA, 22–25 February 2016.

143. Wei, H.; Zhou, H.; Sankaranarayanan, J.; Sengupta, S.; Samet, H. Residual Convolutional LSTM for Tweet Count Prediction. In Proceedings of the Companion of the The Web Conference 2018 on The Web Conference W.W.W. '18, Lyon, France, 23–27 April 2018.

144. Lee, M.; Kim, H.; Kim, O. Why Do People Retweet a Tweet? Altruistic, Egoistic, and Reciprocity Motivations for Retweeting. *Psychologia* **2015**, *58*, 189–201. [CrossRef]

145. Kim, T.-Y.; Kim, J.; Lee, J.; Lee, J.-H. A Tweet Summarization Method Based on a Keyword Graph. In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication—ICUIMC '14, Siem Reap, Cambodia, 9–11 January 2014.

146. Jeon, M.; Jun, S.; Hwang, E. Hashtag Recommendation Based on User Tweet and Hashtag Classification on Twitter. In *Web-Age Information Management*; Springer International Publishing: Cham, Switzerland, 2014; pp. 325–336, ISBN 9783319115375.

147. Deveaud, R.; Boudin, F. Effective Tweet Contextualization with Hashtags Performance Prediction and Multi-Document Summarization. In Proceedings of the Initiative for the Evaluation of XML Retrieval (INEX), Valencia, Spain, 23–26 September 2013.

148. Yan, J.L.S.; Kaziunas, E. What is a Tweet Worth? Measuring the Value of Social Media for an Academic Institution. In Proceedings of the 2012 iConference on—iConference '12, Agadir, Morocco, 5–6 November 2012.

149. Klotz, C.; Ross, A.; Clark, E.; Martell, C. Tweet!—And I Can Tell How Many Followers You Have. In *Advances in Intelligent Systems and Computing*; Springer International Publishing: Cham, Switzerland, 2014; pp. 245–253, ISBN 9783319065373.

150. Dong, L.; Wei, F.; Duan, Y.; Liu, X.; Zhou, M.; Xu, K. The Automated Acquisition of Suggestions from Tweets. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, Washington, DC, USA, 14–18 July 2013.

151. Yuan, S.; Wu, X.; Xiang, Y. A Two Phase Deep Learning Model for Identifying Discrimination from Tweets. Available online: http://www.csce.uark.edu/~{}xintaowu/publ/edbt16p.pdf (accessed on 27 May 2022).

152. Lim, W.Y.; Lee, M.L.; Hsu, W. IFACT: An Interactive Framework to Assess Claims from Tweets. In Proceedings of the 2017 A.C.M. on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017.

153. Morabia, K.; Murthy, N.L.B.; Malapati, A.; Samant, S. SEDTWik: Segmentation-Based Event Detection from Tweets Using Wikipedia. In Proceedings of the 2019 Conference of the North Association for Computational Linguistics, Stroudsburg, PA, USA, 2–7 June 2019; pp. 77–85.

154. Yamamoto, Y.; Kumamoto, T.; Nadamoto, A. Followee Recommendation Based on Topic Extraction and Sentiment Analysis from Tweets. In Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services, Brussels, Belgium, 11–13 December 2015.

155. Kvtkn, P.; Ramakrishnudu, T. A Novel Method for Detecting Psychological Stress at Tweet Level Using Neighborhood Tweets. *J. King Saud Univ. Comput. Inf. Sci.* **2021**. *Epub ahead of print*. [CrossRef]

156. Zhou, L.; Wang, W.; Chen, K. Identifying Regrettable Messages from Tweets. In Proceedings of the 24th International Conference on World Wide Web—W.W.W. '15 Companion, Florence, Italy, 18–22 May 2015.

157. Jussila, J.; Madhala, P. Cognitive Computing Approaches for Human Activity Recognition from Tweets—A Case Study of Twitter Marketing Campaign. In *Research & Innovation Forum 2019*; Springer International Publishing: Cham, Switzerland, 2019; pp. 153–170, ISBN 9783030308087.

158. McClellan, C.; Ali, M.M.; Mutter, R.; Kroutil, L.; Landwehr, J. Using Social Media to Monitor Mental Health Discussions—Evidence from Twitter. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 496–502. [CrossRef] [PubMed]

159. Alowibdi, J.S.; Buy, U.A.; Yu, P.S.; Ghani, S.; Mokbel, M. Deception Detection in Twitter. *Soc. Netw. Anal. Min.* **2015**, *5*, 1–16. [CrossRef]

160. Pauken, B.; Pradyumn, M.; Tabrizi, N. Tracking Happiness of Different U.S. Cities from Tweets. In *Big Data—BigData 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 140–148, ISBN 9783319943008.

161. Ibrohim, M.O.; Budi, I. Multi-Label Hate Speech and Abusive Language Detection in Indonesian Twitter. In Proceedings of the Third Workshop on Abusive Language Online, Florence, Italy, 1 August 2019; pp. 46–57.

162. Sankaranarayanan, J.; Samet, H.; Teitler, B.E.; Lieberman, M.D.; Sperling, J. TwitterStand: News in Tweets. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems—G.I.S. '09, Seattle WA, USA, 4–6 November 2009.

163. Haque, R.; Hasanuzzaman, M.; Ramadurai, A.; Way, A. Mining Purchase Intent in Twitter. *Comput. Sist.* **2019**, *23*, 871–881. [CrossRef]

164. Chaloulos, K. Inferring Shared Interests from Tweets. Available online: https://pub.tik.ee.ethz.ch/students/2011-FS/SA-2011-2 0.pdf (accessed on 27 May 2022).

165. Bollen, J.; Mao, H.; Pepe, A. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. *ICWSM* **2011**, *5*, 450–453.

166. 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.

167. Ilona, K.F.; Budi, I. Classification of Inundation Level Using Tweets in Indonesian Language. In Proceedings of the 2021 10th International Conference on Software and Computer Applications, Kuala Lumpur, Malaysia, 23–26 February 2021.

168. Tankard, E.; Flowers, C.; Li, J.; Rawat, D.B. Toward Bias Analysis Using Tweets and Natural Language Processing. In Proceedings of the 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2021; pp. 1–3.

169. Umakanth, N.; Santhi, S. Classification and Ranking of Trending Topics in Twitter Using Tweets Text. *J. Crit. Rev.* **2020**, *7*, 895–899. [CrossRef]

170. Garcia-Lopez, F.J.; Batyrshin, I.; Gelbukh, A. Analysis of Relationships between Tweets and Stock Market Trends. *J. Intell. Fuzzy Syst.* **2018**, *34*, 3337–3347. [CrossRef]

171. Liew, J.K.-S. Do Tweet Sentiments Still Predict the Stock Market? *SSRN Electron. J.* **2016**, 1–16. [CrossRef]

172. 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, Taiwan, 14–16 December 2018.

173. Balusamy, B.; Murali, T.; Thangavelu, A.; Krishna, P.V. A Multi-Level Text Classifier for Feedback Analysis Using Tweets to Enhance Product Performance. *Int. J. Electron. Mark. Retail.* **2015**, *6*, 315. [CrossRef]

174. Ahmed, H.; Razzaq, M.A.; Qamar, A.M. Prediction of Popular Tweets Using Similarity Learning. In Proceedings of the 2013 IEEE 9th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 9–10 December 2013; pp. 1–6.

175. Sharif, W.; Mumtaz, S.; Shafiq, Z.; Riaz, O.; Ali, T.; Husnain, M.; Choi, G.S. An Empirical Approach for Extreme Behavior Identification through Tweets Using Machine Learning. *Appl. Sci.* **2019**, *9*, 3723. [CrossRef]

176. Vemprala, N.; Akello, P.; Valecha, R.; Rao, H.R. An Exploratory Analysis of Alarming and Reassuring Messages in Twitterverse during the Coronavirus Epidemic. In Proceedings of the AMCIS 2020, Virtual Conference, 15–17 August 2020.

177. Akpojivi, U. Euphoria and Delusion of Digital Activism: Case Study of #ZumaMustFall. In *Advances in Social Networking and Online Communities*; I.G.I. Global: Hershey, PA, USA, 2018; pp. 179–202, ISBN 9781522528548.

178. Anwar, M.; Yuan, Z. Linking Obesity and Tweets. In *Smart Health*; Springer International Publishing: Cham, Switzerland, 2016; pp. 254–266, ISBN 9783319291741.

179. Silva, W.; Santana, Á.; Lobato, F.; Pinheiro, M. A Methodology for Community Detection in Twitter. In Proceedings of the International Conference on Web Intelligence—W.I. '17, Leipzig, Germany, 23—26 August 2017.

180. Sloan, L.; Morgan, J.; Housley, W.; Williams, M.; Edwards, A.; Burnap, P.; Rana, O. Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter. *Sociol. Res. Online* **2013**, *18*, 74–84. [CrossRef]

181. Culotta, A.; Cutler, J. Mining Brand Perceptions from Twitter Social Networks. *Mark. Sci.* **2016**, *35*, 343–362. [CrossRef]

182. Jain, S.; Sharma, V.; Kaushal, R. Towards Automated Real-Time Detection of Misinformation on Twitter. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21–24 September 2016; pp. 2015–2020.

183. Text Processing Extenstion of RapidMiner. Available online: https://marketplace.rapidminer.com/UpdateServer/faces/product_ details.xhtml?productId=rmx_text (accessed on 11 July 2022).

184. Text Analysis by AYLIEN. Available online: https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml? productId=rmx_com.aylien.textapi.rapidminer (accessed on 11 July 2022).

185. String Matching Extenstion of RapidMiner. Available online: https://marketplace.rapidminer.com/UpdateServer/faces/ product_details.xhtml?productId=rmx_string_matching (accessed on 11 July 2022).

186. Levenshtein, V.I. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Sov. Phys. Dokl.* **1966**, *10*, 707.

187. Natural Language Toolkit. Available online: https://sourceforge.net/projects/nltk/ (accessed on 10 July 2022).

188. SpaCy. Industrial-Strength Natural Language Processing in Python. Available online: https://spacy.io/ (accessed on 10 July 2022).

189. TextBlob: Simplified Text Processing—TextBlob 0.16.0 Documentation. Available online: https://textblob.readthedocs.io/en/dev/ (accessed on 10 July 2022).

190. Overview. Available online: https://stanfordnlp.github.io/CoreNLP/index.html (accessed on 10 July 2022).

191. PyNLPl. Available online: https://pypi.org/project/PyNLPl/ (accessed on 10 July 2022).

192. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]

193. Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-Learn: Machine Learning without Learning the Machinery. *GetMob. Mob. Comput. Commun.* **2015**, *19*, 29–33. [CrossRef]

194. Keras: The Python Deep Learning API. Available online: https://keras.io/ (accessed on 10 July 2022).

195. PyTorch. Available online: https://pytorch.org/ (accessed on 10 July 2022).

196. Pandas. Available online: https://pandas.pydata.org/ (accessed on 10 July 2022).

197. Michalke, M. Text Analysis with Emphasis on POS Tagging, Readability, and Lexical Diversity [R Package KoRpus Version 0.13-8]. Available online: https://cran.r-project.org/web/packages/koRpus/index.html (accessed on 10 July 2022).

198. Hornik, K. Apache OpenNLP Tools Interface [R Package OpenNLP Version 0.2-7]. Available online: https://cran.r-project.org/web/packages/openNLP/index.html (accessed on 10 July 2022).

199. Quantitative Analysis of Textual Data. Available online: https://quanteda.io/ (accessed on 10 July 2022).

200. Hornik, K. R/Weka Interface [R Package RWeka Version 0.4-44]. Available online: https://cran.r-project.org/web/packages/RWeka/index.html (accessed on 10 July 2022).

201. Wrapper to the SpaCy NLP Library. Available online: https://spacyr.quanteda.io/ (accessed on 10 July 2022).

202. Wickham, H. Simple, Consistent Wrappers for Common String Operations [R Package Stringr Version 1.4.0]. Available online: https://cran.r-project.org/web/packages/stringr/index.html (accessed on 10 July 2022).

203. Selivanov, D. Text2vec: Fast Vectorization, Topic Modeling, Distances and GloVe Word Embeddings in R. Available online: https://github.com/dselivanov/text2vec (accessed on 10 July 2022).

204. Text Mining Package [R Package Tm Version 0.7-8]. Available online: https://cran.r-project.org/web/packages/tm/index.html (accessed on 10 July 2022).

205. Apache OpenNLP. Available online: https://opennlp.apache.org/ (accessed on 10 July 2022).

206. Apache UIMA—Apache UIMA. Available online: https://uima.apache.org/ (accessed on 11 July 2022).

207. Wikipedia Contributors General Architecture for Text Engineering. Available online: https://en.wikipedia.org/w/index.php?title=General_Architecture_for_Text_Engineering&oldid=1065938586 (accessed on 11 July 2022).

208. LingPipe Home. Available online: http://www.alias-i.com/lingpipe/ (accessed on 10 July 2022).

209. Mallet: MAchine Learning for LanguagE Toolkit. Available online: https://mimno.github.io/Mallet/ (accessed on 10 July 2022).

210. NLP4J by Emorynlp. Available online: https://emorynlp.github.io/nlp4j/ (accessed on 10 July 2022).

211. Welcome to Apache Lucene. Available online: https://lucene.apache.org/ (accessed on 10 July 2022).

212. Emms, S. MITIE: MIT Information Extraction. Available online: https://www.linuxlinks.com/mitie-mit-information-extraction/ (accessed on 11 July 2022).

213. Emms, S. MeTA—Modern C++ Data Sciences Toolkit. Available online: https://www.linuxlinks.com/meta-modern-c-plus-plus-data-sciences-toolkit/ (accessed on 11 July 2022).

214. Emms, S. CRF++: Yet Another CRF Toolkit. Available online: https://www.linuxlinks.com/crf-yet-another-crf-toolkit/ (accessed on 11 July 2022).

215. van Gompel, M. Colibri-Core. Available online: https://github.com/proycon/colibri-core (accessed on 11 July 2022).

216. Wang, C. InsNet. Available online: https://github.com/chncwang/InsNet (accessed on 11 July 2022).

217. Libfolia: FoLiA Library for C++. Available online: https://github.com/LanguageMachines/libfolia (accessed on 11 July 2022).

218. Twitter-Text. Available online: https://github.com/twitter/twitter-text (accessed on 11 July 2022).

219. Moore, B. Knwl.Js. Available online: https://github.com/benhmoore/Knwl.js (accessed on 11 July 2022).

220. Poplar. Available online: https://github.com/synyi/poplar (accessed on 11 July 2022).

221. Nlp.Js. Available online: https://github.com/axa-group/nlp.js (accessed on 11 July 2022).

222. Node-Question-Answering. Available online: https://github.com/huggingface/node-question-answering (accessed on 11 July 2022).

223. Saul. Available online: https://github.com/CogComp/saul (accessed on 11 July 2022).

224. Astrakhantsev, N. ATR4S: Toolkit with State-of-the-Art Automatic Terms Recognition Methods in Scala. *arXiv* **2016**, arXiv:1611.07804. [CrossRef]

225. Stanton, A. Word2vec-Scala: Scala Port of the Word2vec Toolkit. Available online: https://github.com/Refefer/word2vec-scala (accessed on 11 July 2022).

226. Hall, D. Epic. Available online: https://github.com/dlwh/epic (accessed on 11 July 2022).

227. Tm: Regularized Multilingual Probabilistic Semantic Analysis Scala Implementation. Available online: https://github.com/ispras/tm (accessed on 11 July 2022).

228. Potapov, S. Whatlang-Rs. Available online: https://github.com/greyblake/whatlang-rs (accessed on 11 July 2022).

229. Snips-Nlu-Rs: Snips NLU Rust Implementation. Available online: https://github.com/snipsco/snips-nlu-rs (accessed on 11 July 2022).

230. Rust-Bert. Available online: https://github.com/guillaume-be/rust-bert (accessed on 11 July 2022).

231. Hinman, L. Clojure-Opennlp. Available online: https://github.com/dakrone/clojure-opennlp (accessed on 11 July 2022).

232. Inflections-Clj. Available online: https://github.com/r0man/inflections-clj (accessed on 11 July 2022).

233. Postagga: A Library to Parse Natural Language in Pure Clojure and ClojureScript. Available online: https://github.com/turbopape/postagga (accessed on 11 July 2022).

234. Monkeylearn-Ruby. Available online: https://github.com/monkeylearn/monkeylearn-ruby (accessed on 11 July 2022).

235. Dialogflow-Ruby-Client: Ruby SDK for Dialogflow. Available online: https://github.com/dialogflow/dialogflow-ruby-client (accessed on 11 July 2022).

236. Kane, A. FastText-Ruby: Efficient Text Classification and Representation Learning for Ruby. Available online: https://github.com/ankane/fastText-ruby (accessed on 11 July 2022).

237. Granger, M. Ruby-Wordnet. Available online: https://github.com/ged/ruby-wordnet (accessed on 11 July 2022).

238. Ruby-Fann. Available online: https://github.com/tangledpath/ruby-fann (accessed on 11 July 2022).

239. Tensorflow.Rb: Tensorflow for Ruby. Available online: https://github.com/somaticio/tensorflow.rb (accessed on 11 July 2022).
240. Wailes, C. RLTK: The Ruby Language Toolkit. Available online: https://github.com/chriswailes/RLTK (accessed on 11 July 2022).