Grenze International Journal of Engineering and Technology, June Issue

Twitter Data Analysis using Natural Language Processing and Recurrent Neural Network

Sachin Deshmukh[1], M. S. Ali[2] Amol P. Bhagat[3] and Vikram Kakde[4]
[1-2]Dept. of Computer Science and Engineering, Prof Ram Meghe College of Engg. & Mgmt., Badnera, Amravati, India
Email: {sachindeshmukh24, softalis}@gmail.com
[3]Dept. of Information Technology, Prof Ram Meghe College of Engg. & Mgmt., Badnera, Amravati, India
Email: amol.bhagat84@gmail.com
ORCID: 0000-0003-2674-6604
[4]Department of Electronics and Telecommunication Engineering, Prof Ram Meghe College of Engg. & Mgmt., Badnera, Amravati, India
Email: vikramkakade.160@gmail.com

***Abstract*—The essence of this work presented is the conception and construction of a unique model. This model harnesses the power of Natural Language Processing (NLP) and Recurrent Neural Networks (RNN) to decipher the sentiments expressed across various social media platforms. The ultimate goal is to illuminate the emotional landscape of the digital world. In the work presented in this paper the twitter social media posts/messages are retrieved having a particular keyword. Many posts are made using identical message formats which users are unaware about. For effective retrieving of the post identifying the relevant information is required. Many social media post have data which are irrelevant to analyze. For effective analysis removing this irrelevant data is necessary. Once the posts are processed, tokens and keywords are identified. Keywords are important aspects which are used further to identify the sentiment. The processed data is then stored in database and split into two category for training and testing purpose. The polarity and subjectivity of the data is then calculated for analyzing the sentiments using NLP and RNN. The proposed NLP-RNN approach is evaluated using accuracy, loss, and evaluation time required.**

***Index Terms*— data analysis, deep learning, natural language processing, recurrent neural network, sentiment analysis.**

I. INTRODUCTION

Deep Learning (DL), a subset of Artificial Intelligence (AI), is the exploration of computer algorithms that enhance autonomously through experience. Such algorithms architect a mathematical framework using a subset of data, referred to as training data, enabling them to make informed predictions or choices without the need for explicit programming. DL is intimately linked with computational statistics, which emphasizes making predictions using computers. For straightforward tasks assigned to computers, it’s feasible to program algorithms instructing the machine how to carry out all necessary steps to solve the problem at hand; in this case, the computer doesn’t need to learn anything. However, for more complex tasks, manually creating the required

Grenze ID: 01.GIJET.10.2.524_1

algorithms can be a daunting task for humans. In practice, it might be simpler to help the machine develop its own algorithm, rather than having human programmers specify every necessary step. Deep learning finds its applications in a plethora of fields such as agriculture, banking, computer vision, medical diagnostics, natural language processing, and many more. It's fascinating how these algorithms can learn and adapt over time, much like a human brain, but in a more systematic and unbiased manner. The potential of deep learning is vast and we're just beginning to scratch the surface.

Unravelling the essence of text to uncover valuable insights is a remarkable application of machine learning, known as Natural Language Processing (NLP). By employing text analysis, biometrics, and NLP, we can discern, extract, and scrutinize emotional states and subjective information, a comprehensive process referred to as Sentiment Analysis. This technique finds its utility in comprehending the sentiments of customers or users expressed in materials like reviews, survey responses, social media posts, and healthcare documents. These applications span across diverse domains from marketing to customer service to clinical medicine. The field of opinion mining and sentiment analysis has seen a swift expansion, with the objective to probe into the sentiments or text present on various social media platforms through machine-learning techniques, sentiment analysis, subjectivity analysis, or polarity calculations. Even with the widespread application of numerous advanced learning methodologies and instruments for sentiment analysis, particularly in the context of elections, there remains an urgent demand for a more innovative strategy. The integration of a dual-method strategy, which includes a sentiment analysis tool driven by machine learning, is crucial for improving the practicality of these systems. This fusion of techniques promises to revolutionize the way we understand and interpret sentiments in the digital world.

In 2019, the digital denizens of the world dedicated approximately 2.5 hours daily to social networks and messaging, as per the Social Media Trends 2019 report by Global Web Index. Platforms like Twitter, Facebook, TikTok, Instagram, and others are experiencing a surge in popularity, thanks to their open nature and the freedom they offer for content sharing. In India, the escalating usage can be largely credited to the plummeting cost of data. On average, Indian netizens spent 2.4 hours on social media, mirroring the global average. However, the COVID-19 pandemic and the ensuing lockdown saw Indians turning to social media to while away their time. In the initial week of the lockdown, Indians devoted more than four hours daily to social media, marking an 87% increase from the preceding week. Prior to the lockdown, the average daily social media usage was 160 minutes, which leaped to 280 minutes during the first week of lockdown, as revealed by a survey. The study additionally underscored that 75% of individuals were dedicating more time to platforms like Facebook, Twitter, and WhatsApp in comparison to the previous week. In the throes of the coronavirus pandemic, individuals were predominantly engaging with news content and maintaining connections with loved ones via social media platforms. Internet browsing witnessed a 72% surge during the first week of lockdown. The distribution of this trend is represented in Figure 1.

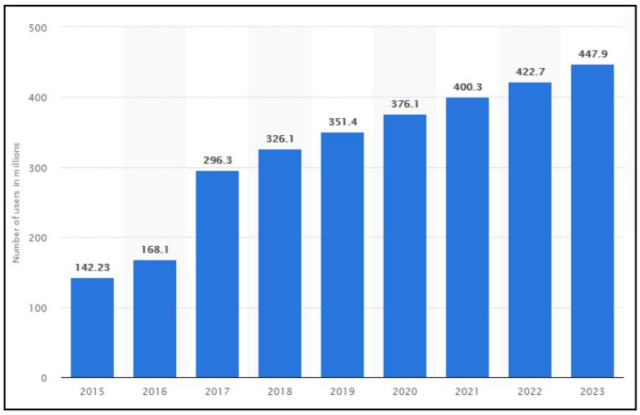


Figure 1 Number of users in India using Social Media (in millions)

In the recent lockdown, social media emerged as a real-time conduit for disseminating vital information related to diseases. Research indicated that the earliest accounts of the H1N1 influenza epidemic were initially recorded on social media platforms. This led to government bodies like the Centres for Disease Control and Prevention (CDC) leveraging social media to alert the public about imminent infections such as the Zika and Ebola outbreaks. The ever-evolving landscape of social media has established itself as a crucial communication channel during crises. Social media platforms are being increasingly adopted by public health departments and individuals alike to communicate and share information during public health emergencies. It’s a testament to the power of digital platforms in fostering communication and promoting awareness during critical times.

II. Related Work

A multilingual dataset encompassing the coronavirus (COVID-19) Twitter discourse, which has been in continuous accumulation since January 22, 2020, has been unveiled by [1]. This dataset remains accessible to the research fraternity for further investigation and discovery. The dataset contain the tweets of people around the globe who have tweeted using the hashtags related to recent pandemics. This dataset has been used by many researchers to build the models which can track misinformation and rumours or can be used to get the reason behind the fear or anxiety among the users. They maintain a dynamic GitHub account that features weekly updates on the dataset procured from Twitter. As of now, it boasts over 123 million tweets, with English constituting 60% of the content. Infodemiology [2], a term coined by Gunther Eysenbach in the early 2000s, represents a novel scientific research domain. It concentrates on scouring the Internet for health-related content contributed by users, all with the paramount objective of enhancing public health..

The application of Infodemiology studies by [3] extends to network analysis, harmoniously paired with content analysis. The researchers have delved into the exploration of information dissemination networks and the trends of disseminating news related to COVID-19 on Twitter, particularly within the Korean landscape. They collected the data since February 29, 2020 and have around 43,832 tweets in their sample. This study has helped to understand how to apply the twitter insights to draft the policies related to the health sector. Nine G7 world leader’s tweets during recent pandemic situation are studied in [4]. All tweets belong to period 17 November 2019 to 17 March 2020. The collective following of these leaders’ Twitter accounts amasses to a staggering 85.7 million users. This study encapsulates the effectiveness of social platforms as a swift conduit for disseminating public health information to citizens. An archive of more than 20 million tweets, created in the early phases of the COVID-19 pandemic from January 28 to April 9, 2020, were gathered using ‘wuhan’, ‘corona’, ‘nCov’, and ‘covid’ as search terms [5]. The global sentiment reverberating around the COVID-19 pandemic was meticulously analysed.

Four primal sentiments—dread, rage, grief, and happiness—were scrutinized on Twitter, particularly in relation to the COVID-19 situation. It was observed that the public’s emotional state underwent a significant shift from fear to anger as the pandemic unfolded, with traces of sadness and joy also emerging. An unguided ML methodology was utilized by [6] on Twitter datasets to identify self-reported indications linked to COVID-19, along with experiences related to testing, and so on. Using the Biterm Topic Model (BTM), an unsupervised machine learning method, the tweets were analysed. The tweets were grouped into topics based on the similarity of the words they contained, which reflected discussions on symptoms, testing, and recovery. The tweets in these topics were then extracted, manually labelled for content analysis, and examined for their statistical and geographic features. The authors collected 4,492,954 tweets with words related to COVID-19 symptoms. After applying BTM to find relevant topics and removing duplicate tweets, the authors identified 3465 (<1%) tweets. These tweets contained user-generated conversations about experiences that users linked to possible COVID-19 symptoms and other disease experiences.

The role of Twitter and other social media platforms in spreading medical information and misinformation amid the current pandemic is discussed in [7]. It is posited that the unrestricted and unreviewed flow of medical information shared on social media could potentially be detrimental to the general populace. The analysis and classification of social media users’ sentiments (e.g., positive, negative) towards an object, situation, or system are very popular among researchers. However, debates about the underlying socioeconomic factors affecting such sentiments are rarely addressed. In [8], data from Twitter and census data, covering socioeconomic factors like income, family size, and employment status, are combined to build a binary logic model.. This model is employed to examine people’s sentiments about the reopening of the US economy. It was observed that individuals with lower income and less education expressed more happiness about the economy reopening than those who were well-educated and employed. It is recommended that government agencies take such studies into account while allocating resources to the populace and deciding on policy options to ameliorate the situation.

Tagging a tweet to an associated sub topic is demonstrated in [9] . An unguided topic grouping method, known as Latent Dirichlet Allocation (LDA), is employed to enhance the comprehension of the underlying subtopics. LDA is a probabilistic generative model which learns a multinomial distribution of latent topics in a given document and words in a given topic. Every tweet is tagged to one of the ten sub topic identified using LDA and then further used emotion intensity.

Many reports around the globe have claimed that domestic violence has increased during pandemic lockdown. Tweets are analyzed in [10] and ML technique LDA is applied to identify the salient themes, topics from the tweets. Nine distinct themes have been unearthed pertaining to domestic violence in the background of the COVID-19 pandemic. A thorough analysis on 301,606 tweets about family violence and COVID-19, covering the period from April 12 to July 16, 2020. On the onset of current pandemic situation there was rumour that nicotine present in cigarettes and vaping is effective on Corona virus. This motivated [11] to study the sentiments of the smokers during the pandemic over the period of time. An analysis was conducted on tweets about COVID-19 and smoking, from both casual and tobacco industry-related accounts, were analysed. The compilation of tweets, which amounted to 33,890, spanned from January 1 to May 1, 2020.

The stringent lockdown has disrupted the supply chain and the availability of goods needed by consumers. The influence of COVID-19 on supply chain decisions has been examined by [12] through the analysis of Twitter data from NASDAQ 100 companies. This analysis has led to the emergence of themes concerning the challenges encountered by these firms and the strategies they are implementing, as revealed by text analytics tools. In [13], the sentiments of Indians post-lockdown announcements have been taken into account. Twitter, the social media applications, served as the medium for this analysis. The study involved scrutinizing tweets to understand the Indian populace's feelings about the lockdown. Using two prominent hashtags: #IndiaLockdown and #IndiafightsCorona, tweets were collected from March 25th to March 28th, 2020. The analysis comprised 24,000 tweets in total.

Social media platforms serve as a bountiful reservoir where the public's responses to societal issues can be examined. Given the profound impact of COVID-19 on people's existence, it is vital to grasp how people react to public health actions and recognize their fears. The reactions and concerns of people regarding COVID-19 have been scrutinized by [14] in North America, with a particular emphasis on Canada. Using techniques such as topic modeling and aspect-based sentiment analysis, tweets about COVID-19 are examined and the outcomes are explained in cooperation with public health experts. These insights can be useful for public health agencies when designing policies for new actions.

Textual analytics, in tandem with machine learning, is harnessed by [15] to categorize tweets using a classification algorithm. This process employs the Naïve Bayes and logistic regression methods for classification. For short tweets, the Naïve Bayes method has a classification accuracy of 91%, while the logistic regression method reaches 74%. However, both methods show relatively lower performance for longer tweets. The implementation of lockdowns in numerous countries to restrain the transmission of the coronavirus resulted in severe mental health issues among the populace due to prolonged strict social isolation. Therefore, monitoring the mental health of individuals during various stages and topics of the ongoing pandemic became crucial for effective policymaking. In [16], these problems are tackled by examining social media platforms, which act as channels for individuals to convey and exchange their personal views and emotions. More than 13 million tweets about COVID-19 were gathered and a new framework was suggested. This framework can be utilized to analyze the sentiment and topic dynamics stemming from COVID-19 in social media posts. Initial efforts have been undertaken to analyze user behavior, sentiment analysis, content recommendations, and more [17-21].

III. Natural Language Processing Integrated in Recurrent Neural Network for Twitter Data Analysis

Recurrent Neural Networks (RNNs) represent a class of machine learning algorithms that are particularly adept at handling intricate data types such as text, temporal data, financial records, speech, audio, video, and more. RNNs excel in tackling problems where the significance lies more in the array than in the individual objects. At their core, RNNs are a type of neural network that transforms a certain operation into a repetitive loop. This loop typically involves an iteration over two concepts: matrix multiplication and the addition or amalgamation of nonlinear functions. When it comes to text-related tasks, RNNs demonstrate proficiency in several areas, including: Sequence tagging; Natural Language Processing (NLP) text classification; and NLP text rendering. These capabilities make RNNs a powerful tool in the realm of machine learning and text analysis.

RNNs excel in a variety of tasks, including time series prediction and array predictions for images or tables. The media has been abuzz with high-profile and contentious reports about advancements in text generation, particularly with respect to OpenAI's GPT-2 algorithm. In many instances, the text generated by these algorithms is virtually indistinguishable from human-written text. One of the key features of RNNs is their internal memory, which allows previous inputs to shape future predictions. This makes it easier and more accurate to predict the next word if the preceding word is known. NLP, a sub branch of computer science and AI, is dedicated to processing and generating natural language information.

While there exists research beyond the realm of machine learning, the majority of Natural Language Processing (NLP) now relies on language models crafted by machine learning. The Gateway Repetitive Unit, occasionally referred to as the Gated Repetitive Network, is a compact neural network that emerges at the output of each iteration. This network comprises three layers, including the recovery layer of the RNNs, the reset gate, and the update gate. The update gate functions as a forgotten and entered gate. The amalgamation of these two gates operates similarly to the trio of forget, entry, and exit gates in Long Short Term Memory (LSTM).

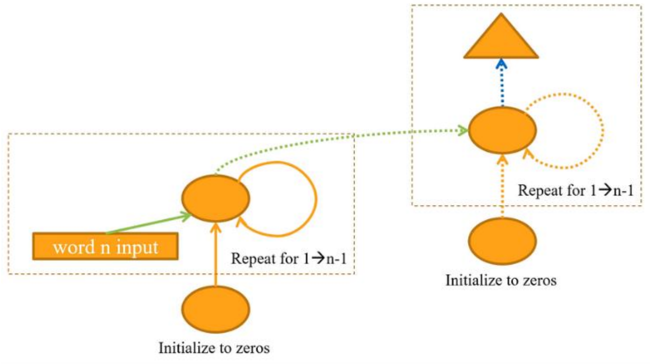


Figure 2 Proposed RNN with GRU for NLP

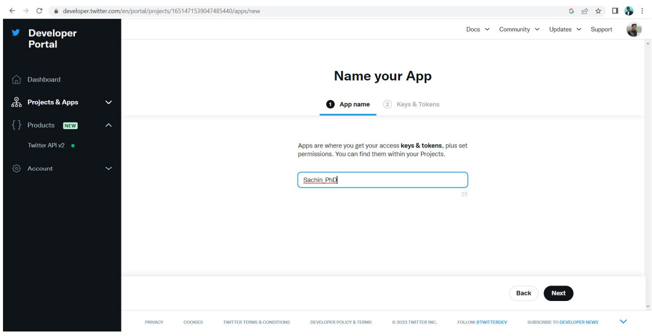


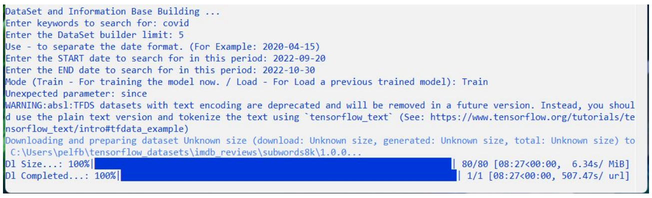
Figure 3 Twitter Developer Portal for generating Consumer API Key, Consumer API Secret, Access Token and Access Token Secret

In contrast to LSTM, the Gated Recurrent Unit (GRU) integrates the cell state and the latent state, whereas in LSTM, these states are distinct. In the realm of text classification, the prediction network is tasked with determining the category or group to which the text belongs. One of its most prevalent applications is discerning whether the sentiment of a text is positive or negative. If a Recurrent Neural Network (RNN) is trained to predict text from a specific domain, as previously described in this document for RNNs, its ability to replicate it for the distribution of articles of the same name is nearly flawless. The productive "head" of the web is detached, leaving behind the "arm" of the web. Subsequently, the weight at the back can be frozen. A new classification head can be attached to the backend and trained to predict the desired classification. Gradually unloading weights in layers is an effective strategy for training speed. The process begins with the weights of the last two layers, then proceeds to the weights of the last three layers, and finally increases the total weight of each layer. The proposed approach is depicted in Figure 2.

IV. Experimental Evaluations

The proposed approach is designed and developed using Anaconda Python distribution and Twitter Developer API. The proposed approach is evaluated on Windows 11, An Intel Core i3 processor from the 11th generation, coupled with 8GB of RAM and a 256GB SSD, powers the system. The Consumer API Key, Consumer API Secret, Access Token, and Access Token Secret are all generated via the Twitter Developer Login. This is accomplished by creating an app on the Twitter Developer portal, as depicted in Figure 3. The python modules textblob, tweepy, tensorflow, and tensorflow-dataset are utilized to carry out the implementation and evaluation of the proposed approach. To implement the proposed approach the python script is written "dlsa_sma.py" indicating Deep Learning based Sentiment Analysis for Social Media Applications. This can be executed in

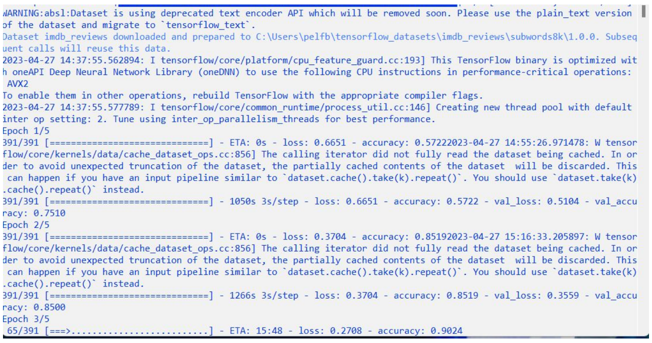
Anaconda prompt as “python dlsa_sma.py”. Once executed it starts downloading the tweets for the entered keyword and the time period specified as shown in figure 4.



```
DataSet and Information Base Building ...
Enter keywords to search for: covid
Enter the DataSet builder limit: 5
Use - to separate the date format. (For Example: 2020-04-15)
Enter the START date to search for in this period: 2022-09-20
Enter the END date to search for in this period: 2022-10-30
Mode (Train - For training the model now. / Load - For Load a previous trained model): Train
Unexpected parameter: since
WARNING:absl:TFDS datasets with text encoding are deprecated and will be removed in a future version. Instead, you shoul
d use the plain text version and tokenize the text using `tensorflow_text` (See: https://www.tensorflow.org/tutorials/te
nsorflow_text/intro#tfdata_example)
Downloading and preparing dataset Unknown size (download: Unknown size, generated: Unknown size, total: Unknown size) to
 C:\Users\pelfb\tensorflow_datasets\imdb_reviews\subwords8k\1.0.0...
Dl Size...: 100%|                                        | 80/80 [08:27<00:00,  6.34s/ MiB]
Dl Completed...: 100%|                                   | 1/1 [08:27<00:00, 507.47s/ url]
```

Figure 4 Downloading Tweets for Training the Proposed Model

The suggested model undergoes training for a predetermined number of epochs. Key metrics such as training time, accuracy, loss, validation accuracy, and validation loss are determined for the proposed approach, as exemplified in Figure 5. The evaluation outcomes for the proposed approach are presented in Table 1. Tweets are classified based on sentiment into seven different classes: positive, mildly positive, very positive, negative, mildly negative, very negative, and neutral.



```
WARNING:absl:Dataset is using deprecated text encoder API which will be removed soon. Please use the plain_text version
of the dataset and migrate to `tensorflow_text`.
Dataset imdb_reviews downloaded and prepared to C:\Users\pelfb\tensorflow_datasets\imdb_reviews\subwords8k\1.0.0. Subseq
uent calls will reuse this data.
2023-04-27 14:37:55.562894: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized wit
h oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critical operations:
 AVX2
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-04-27 14:37:55.577789: I tensorflow/core/common_runtime/process_util.cc:146] Creating new thread pool with default
inter op setting: 2. Tune using inter_op_parallelism_threads for best performance.
Epoch 1/5
391/391 [==============================] - ETA: 0s - loss: 0.6651 - accuracy: 0.57222023-04-27 14:55:26.971478: W tensor
flow/core/kernels/data/cache_dataset_ops.cc:856] The calling iterator did not fully read the dataset being cached. In or
der to avoid unexpected truncation of the dataset, the partially cached contents of the dataset  will be discarded. This
 can happen if you have an input pipeline similar to `dataset.cache().take(k).repeat()`. You should use `dataset.take(k)
.cache().repeat()` instead.
391/391 [==============================] - 1050s 3s/step - loss: 0.6651 - accuracy: 0.5722 - val_loss: 0.5104 - val_accu
racy: 0.7510
Epoch 2/5
391/391 [==============================] - ETA: 0s - loss: 0.3704 - accuracy: 0.85192023-04-27 15:16:33.205897: W tensor
flow/core/kernels/data/cache_dataset_ops.cc:856] The calling iterator did not fully read the dataset being cached. In or
der to avoid unexpected truncation of the dataset, the partially cached contents of the dataset  will be discarded. This
 can happen if you have an input pipeline similar to `dataset.cache().take(k).repeat()`. You should use `dataset.take(k)
.cache().repeat()` instead.
391/391 [==============================] - 1266s 3s/step - loss: 0.3704 - accuracy: 0.8519 - val_loss: 0.3559 - val_accu
racy: 0.8500
Epoch 3/5
 65/391 [===>..........................] - ETA: 15:48 - loss: 0.2708 - accuracy: 0.9024
```

Figure 5 Training time, accuracy, loss, validation accuracy and validation accuracy of the proposed approach

V. Conclusion

This paper describes the design and development of a model for sentiment analysis on Twitter data, leveraging the power of NLP and RNNs. The proposed model is put to the test using the Anaconda Python distribution and the Twitter Developer API. Tweets are categorized based on sentiment into seven distinct classes: positive, mildly positive, very positive, negative, mildly negative, very negative, and neutral. The model achieves a peak accuracy of 94.49% and a validation accuracy of 86.09%. Looking ahead, the proposed methodology can be expanded to encompass other social platforms such as Facebook, WhatsApp, Instagram, etc. The use of hashtags and mentions in conjunction with Twitter data can yield filtered results, providing a more refined analysis.

TABLE I. Performance Evaluation Outcomes of the Proposed Approach

Epoch	Time (S)	Loss %	Accuracy %	Validation Loss %	Validation Accuracy %
1	1050	66.51	57.22	51.04	75.10
2	1266	37.04	85.19	35.59	85
3	1040	25.99	90.92	34.20	85.83
4	1018	21.18	93.21	35.73	86.09
5	1001	17.82	94.49	40	85.99

References

[1] Eason, E. Chen, K. Lerman, and E. Ferrara, ―Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set, JMIR Public Health Surveill., vol. 6, no. 2, p. e19273, May 2020, doi: 10.2196/19273.

[2] Infodemiology, Wikipedia. Jul. 21, 2020, Accessed: Sep. 12, 2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title= Infodemiology&oldid=96880651
[3] H. W. Park, S. Park, and M. Chong, ―Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea, J. Med. Internet Res., vol. 22, no. 5, p. e18897, 2020, doi: 10.2196/18897.
[4] S. R. Rufai and C. Bunce, ―World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis, J. Public Health, vol. 42, no. 3, pp. 510–516, Aug. 2020, doi: 10.1093/pubmed/fdaa049.
[5] M. O. Lwin et al., ―Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends, JMIR Public Health Surveill., vol. 6, no. 2, p. e19447, 2020, doi: 10.2196/19447.
[6] T. Mackey et al., ―Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study, JMIR Public Health Surveill., vol. 6, no. 2, p. e19509, 2020, doi: 10.2196/19509.
[7] H. Rosenberg, S. Syed, and S. Rezaie, ―The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic,‖ Can. J. Emerg. Med., vol. 22, no. 4, pp. 418–421, Jul. 2020, doi: 10.1017/cem.2020.361.
[8] M. M. Rahman, G. G. M. N. Ali, X. J. Li, K. C. Paul, and P. H. J. Chong, ―Twitter and Census Data Analytics to Explore Socioeconomic Factors for Post-COVID-19 Reopening Sentiment, Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3639551, Jun. 2020. Accessed: Sep. 12, 2022. [Online]. Available: https://papers.ssrn.com/abstract=3639551.
[9] R. K. Gupta, A. Vishwanath, and Y. Yang, ―COVID-19 Twitter Dataset with Latent Topics, Sentiments and Emotions Attributes, ArXiv200706954 Cs, Sep. 2020, Accessed: Sep. 12, 2022. [Online]. Available: http://arxiv.org/abs/2007.06954.
[10] J. Xue, J. Chen, C. Chen, R. Hu, and T. Zhu, ―Abusers indoors and coronavirus outside: an examination of public discourse about COVID-19 and family violence on Twitter using machine learning, medRxiv, p. 2020.08.13.20167452, Aug. 2020, doi: 10.1101/2020. 08.13.20167452.
[11] M. Kamiński, A. Muth, and P. Bogdański, ―Smoking, Vaping, and Tobacco Industry During COVID-19 Pandemic: Twitter Data Analysis, Cyberpsychology Behav. Soc. Netw., Jul. 2020, doi: 10.1089/cyber.2020.0384.
[12] A. Sharma, A. Adhikary, and S. B. Borah, ―Covid-19′s impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data, J. Bus. Res., vol. 117, pp. 443–449, Sep. 2020, doi: 10.1016/j.jbusres.2020.05.035.
[13] G. Barkur, Vibha, and G. B. Kamath, ―Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India, Asian J. Psychiatry, vol. 51, p. 102089, Jun. 2020, doi: 10.1016/j.ajp.2020.102089.
[14] H. Jang, E. Rempel, G. Carenini, and N. Janjua, ―Exploratory Analysis of COVID-19 Related Tweets in North America to Inform Public Health Institutes, ArXiv200702452 Cs, Jul. 2020, Accessed: Sep. 13, 2020. [Online]. Available: http://arxiv.org/abs/2007.02452.
[15] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, ―COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification, Information, vol. 11, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/info11060314.
[16] H. Yin, S. Yang, and J. Li, ―Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media, ArXiv200702304 Cs, Jul. 2020, Accessed: Sep. 13, 2020. [Online]. Available: http://arxiv.org/abs/2007.02304
[17] A. S. Ujjainkar, A. D. Chepe and A. P. Bhagat, "Behavioural Analysis of Students and Faculties," *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*, San Salvador, El Salvador, 2018, pp. 1-8, doi: 10.1109/RICE.2018.8509089.
[18] A. P. Bhagat, K. A. Dongre and P. A. Khodke, "Cut-based classification for user behavioral analysis on social websites," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Greater Noida, India, 2015, pp. 53-59, doi: 10.1109/ICGCIoT.2015.7380427.
[19] P. S. Badase, G. P. Deshbhratar and A. P. Bhagat, "Classification and analysis of clustering algorithms for large datasets," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2015, pp. 1-5, doi: 10.1109/ICIIECS.2015.7193191.
[20] N. G. Meshram and A. P. Bhagat, "Connotative features based affective movie recommendation system," International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 2014, pp. 1-7, doi: 10.1109/ICICES.2014.7033834.
[21] S. A. Khan, P. A. Khodke and A. P. Bhagat, "Neuro Fuzzy based user queries categorization," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2015, pp. 1-7, doi: 10.1109/ICIIECS.2015. 7193133