

## TWITTER SENTIMENT ANALYSIS OF COVID-19 VACCINE IN INDIA

<sup>1</sup>Dr. B. MADHURAVANI, <sup>2</sup>Dr. G YUGANDHAR and <sup>3</sup>ADLURI VIJAYA LAKSHMI

<sup>1</sup>Department of CSE, MLR Institute of Technology, Dundigal, Hyderabad, Telangana, India

<sup>2</sup>Department of CSE, GITAM Deemed to be University, Hyderabad

<sup>3</sup>Department of CSE, B V RAJU INSTITUTE OF TECHNOLOGY, Narsapur, Medak District

### Abstract:

An opinion mining subject of interest has recently arisen because of opinionated data found on blogs and social networks. Looking at various styles of points of viewpoints and summarizing them offers useful insight to different demographics who use social media for any of interest. Classification of opinions based on polarity is difficult. This experiment is made for the express purpose of analyzing Twitter users' moods. There are Natural Language Processing methods used to identify the subject of the emotion. In our experiment, we break down the subjectivity into three phases, classifying them, analyzing words, and determining their polarity. is used to identify the correlation between the emotion lexicons and subjects. Some preliminary experiments demonstrate that the method would be more effective than text analysis, as the structure of tweets differs greatly. The study's results clearly show that 20% of the respondents have given a constructive response, with 71% offering a critical comment and 9% remaining neutral.

Keywords- Sentiment Analysis; Natural Language Processing; Tweets

## 1. INTRODUCTION

A lot of interesting information about different user generated content is readily available online. with the rise in the use of Facebook, Twitter, and other social networking sites, ideas can be more widely spread. These views are linked to the user's feelings as well (Reyes-Menendez et al., 2018)[22]. Text classification that is used to extract the information in reviews and predictions for certain topics, products, and in the case of markets, the information contained in opinions. Sentiment Analysis and Opinion Mining (Aquino et al., 2020). Survey Analysis determines polarity (Positivity, Negativity, and Sympathy by having access to a lot of data. Opinion mining is assigned to three levels: "Document," and "Sentence," according to the researchers: The information in this paper contains the results of an examination on Sentence Level. Facts are things stated. Opinions are those inferred from text. A fact is an expression that deals with an entity, occurrence, or its qualities as if it were an object. Opinions are individual viewpoints that go through one's emotions, opinions, thoughts, or feelings about entities, events, and evaluations. They not only tell others about goods and services, but comment on problems in the world of society as well (Thomaz et al., 2017). This paper is composed of six sections, or pieces. Begin in Section II with a quick study of available data and proceed to a more thorough examination of sentiment analysis in Section III. The last section addresses the subject of the framework under discussion as mentioned below. The Case study features tweets about 'Covid-19 Vaccine-India' containing various viewpoints accompanied by results of the study. People can see graphical Polarity classifications demonstrated in this chart as well. Paper draws its conclusions in Section VI about future prospects and difficulties in sentiment analysis, while issues related to the concepts and techniques are mentioned in Section V. Opinion analysis

works may be used to provide engagement for governments in formulation and execution of governance policies.

Twitter, as a common microblogging service, lets users post short, simple status messages with a length between 140 and humanly possible (it enables them to be as short as 140 characters) Most of these tweets carry a personal perspective or feeling about the subject. sentiment can be gained from tweets A direct approach has an advantage over questionnaires and surveys in that it's far easier to collect customer views and opinions(Ford et al., 2019).

[As a matter of fact,] there have been extensive studies on text parsing for the sake of extracting emotion. For example, in the example of sentiment analysis, authors have used movie review domains to try out machine learning techniques (Naïve Bayes, maximum entropy classification, and SVM) Using the SVM model, they got up to 82.9% accuracy(Mathapati and Manjula, 2017). Although the results of classification are heavily dependent on the meaning of texts, however, machine learning methods have difficulties with lexicons of different value levels of sentiment.

After classifying the data based on two sub sensitivity, Pang and Lee applied the machine learning technique of minimum cuts in order to filter out the bias(Poornima and Priya, 2020). Their first analysis showed whether or not the text had a positive or negative feeling, and then graded the text accordingly. It outdistanced its predecessor, with an accuracy of 86.4%

Furthermore, Natural Language Processing (NLP) is also employed. NLP describes a type of emotion, and classifies each lexicon's corresponding to a particular polarity(Ghiassi and Lee, 2018). instead of finding the meaning of the entire sentences, NLP may define basic concepts such as "Subject" and "Sentiment" One of the NLP techniques. It is beneficial for extracting any topic-specific information, sentiments, and also for associating these sentiments with various subjects That was better than using a machine learning algorithm, with accuracies ranging from 87-93% on the analysis of general web pages to over 91% for online papers(Kumar and Chong, 2018). It had to be more generalized text as opposed to making tough cases such as ambiguous or non-sentiment ate ones as simple as possible It is unlikely that previous machine learning and NLP experiments are relevant for tweets, as the framework between the two is different. Here are the three distinctions compared to sentiment analysis findings that emerged during the review of previous studies. Anything good happens in small slices(Kim and Klinger, 2018). Twitter messages can only be posted to be tweets that are 140 characters long. As was discovered in an experiment carried out, average tweet length is 14 sentences, and average sentence length is 78 characters A different issue is evident in text content analysis[23-28], which tends to include a lot of brief, general statements: opinion in tweets tends to be measured on the content of one or a few sentences(Conrad et al., 2019).

There is a need for providing the scientific and psychiatric vocabulary and resources necessary tools to extract contextual knowledge from the web. This is a step toward progress and this paper contains the following information:

- Provide an overview of the most common and widespread approaches to NLP.
- Traditional and statistical methods are applied to explaining and modelling a result.
- Examine the information on sentiment analysis and apply it to problem solving on a psychological condition.
- Analyze current covid vaccine problems and describe how NLP techniques can be used.
- Each part of the paper will concentrate on the main concepts and meanings rather than being blended together into one section.

## **2. AVAILABLE DATA**

On the other hand, however, Go and his colleagues managed to gather over 1,000 tweets to aid in sentiment analysis. Thanks to the use of the Twitter API, we can collect thousands and millions of tweets for training wordiness(Arora et al., 2020). Wordiness, Vocabulary that is made up of abbreviations, acronyms, and/acronyms can lead to more cluttered or illegible text Additionally, URL, image, hashtags, and emoticons are accessible. It's more features that affect the process because they are not words that can be located in a dictionary, nor easy for machines to read and understand. It is best to organize a process that recognizes human language instead of having no strict rules, as computer lacks flexibility. Combining grammatical and conventional word frequency analysis, the method developed and provided practical assistance(Jang et al., 2021). Comprehension uses grammar to study feelings of the text, and finds subject-verb-object associations as well as those between feelings and subjects and verbs. We are glad to see that previous work on sentiment analysis of colloquial text was more accurate than previously thought. No previous training was required, but improved the initial results by 40% This framework is designed to be used to analyses tweets according to their relevant topics(Garvey et al., 2021). A number of preliminary processes have been done to clean the noise from the dataset, and convert it to standard language. In order to identify the overall sentiment of tweets, NLP has and try to discover the subjective areas of them all the Tweets and categories them. Forums on Twitter will be labelled as "positive," "bad," or "neutral" with regards to content(Ismail et al., 2020).

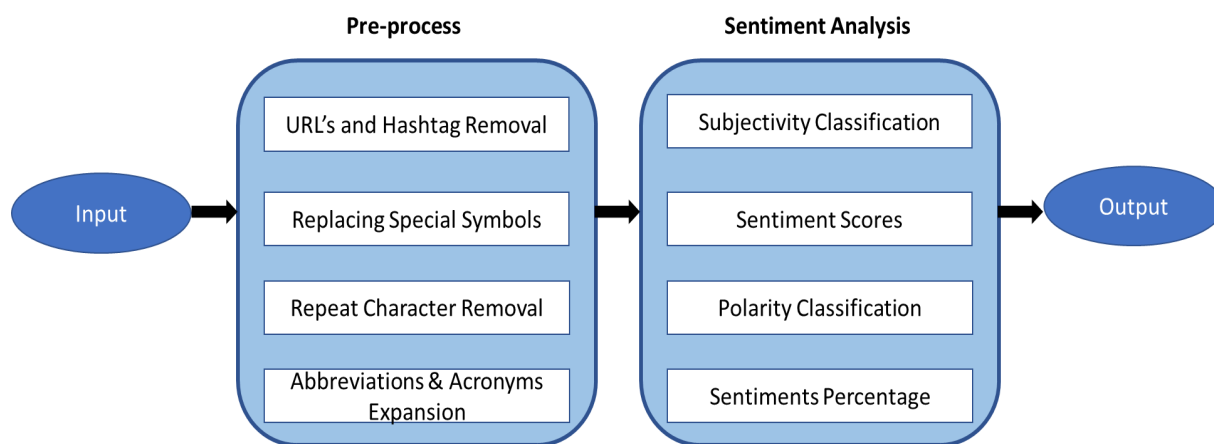
## **3. OVERVIEW OF FRAMEWORK**

This section provides an overview of the general operation of the system. We received Tweets from the universe as a number of times to be served as a control. Everything on Twitter was classified as the opposite of being "upbeat" or the center of the debate, as well as negative and in nature. This is used to see how well the system predicts or measures outcomes with metrics such as the system's predictive accuracy. prior to the study, all datasets are first expanded to their dimensions to accommodate the results. We need to make sure that all of our tweets are written for both humans and search engines before they go out to public[21]. after pre-processing, they can be categorized into categories according to the sentiment relations between words on semantic and

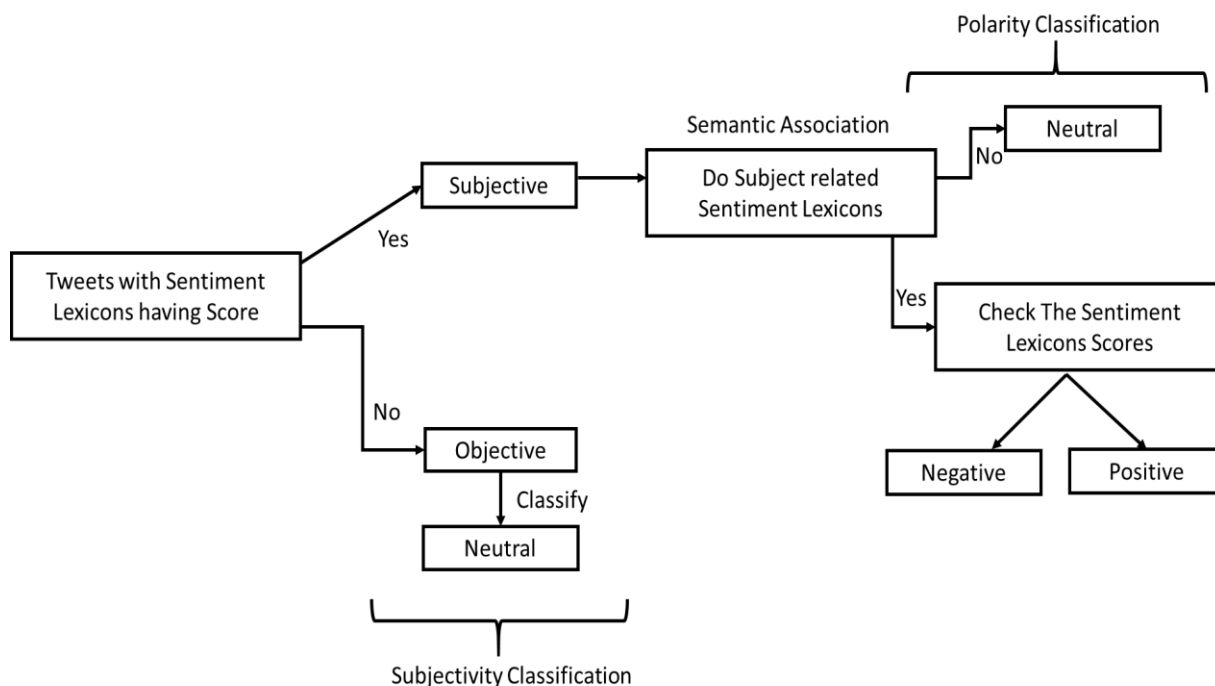
syntactic form, but not valence levels You would then examine the tweets to determine whether they were characterized by either by emotional tone or a lack of emotional content. As the algorithmically assigned lexicons were given specific labels, they were designated as positive, negative, or non-emotive, sentiment words, each were included in three different categories. A wide range of data It was done by following a long, hard road of investigation that 1513 tweets could be culled from Twitter and extracted. The keyword 'Unifi' appears in these tweets, as well as” In addition, it also has to do with classifying. Eight thousand and one negative tweets, with an additional six-thousand-and-and-and-thirty-three thousand and forty-one thousand are made about other users. was examined by the computer to find out whether the messages in the tweets were likely to be positive or negative It checked fifteen hundred and thirteen thousand tweets, with the help of an API Alchemy utilizes ML methods, though Weka doesn't, while it is often applied to sentiment analysis. namely, naïve Bayes, the Bayesian classifier, and the support vector machine A prior processing of tweets had the same effect on Weka, but unprocessed tweets didn't show any results in this test. Features are extracted in the Weka application. Picking up all of the synonyms and cross- and practicing on all the related high-frequency words algorithm results obtained from the machine learning tools, the programs and ideas of Alchemy, Alchemy predictions, and tabulated data were applied to known variables were tested to see if they would hold true.

#### 4. PROPOSED SYSTEM

Figure.1 illustrates the phases of the method, beginning with sentiment analysis. The processing stages are outlined in Section 1, and the classification procedure for sentiment is defined in Section 2.



**Figure 1: Flowchart of Sentiment Analysis**



**Figure 2: Classification Sentiments**

• **Pre-process**

Pre-processing prioritizes to better handle and display the tweets and enhance computer comprehension of the text. It removes URLs and hashtags, replaces special characters, strips abbreviations, and converts acronyms, and words, and makes names shorter and acronyms and acronyms capitalized words easier to read. There are no links to images or URLs in the text because they do not have any obvious purpose. To prevent misunderstanding, words with a hash tag are omitted from the text(Neha et al., 2017)[20]. For example, in order to eliminate confusion, symbols such as “greater” and “and” are used instead of characters. Automatic Sentiment Analysis for Twitter research suggests that text-based analysis is more accurate than emoticon-based. Consequently, emoticons are pulled out of tweets Structureless tweets are broken up into smaller pieces and stripped of the repetitive characters, or are made regular if an expression has been reduced to two syllables, for example, good. Expanded abbreviations, contracted words, acronyms, and symbols are also supported. Say, “I “won’t be” or “don’t” instead. For example, “I won’t be working tomorrow” extends to “I’ll not be working”. Topic capitalization is performed to simplify and the subject for computer processing. The classifications process will progress to predicting the sentiments of the processed tweets(Lv et al., 2021).

**2. Sentiment Classification**

The sentiment classification process is shown in Fig. 2.

a) Confessional Expression The tweets can be classified in either subjective or objective terms. It goes through every tweet one by one, and finds the one with the predominant emotional tone. A good or bad word is used in a positive or negative manner would bem

or a pleasant or offensive manner depends on how it is used in the message It may or may not be objective; either way, it will be impartial(Akyol and Alatas, 2020). As an example, “This Internet is currently being added to the list of services being upgraded to bundle and you'll receive a new package of new internet services.” When the first tweet has no words of positive or negative emotional content, the concept must be avoided and stated simply. Objective and technical reporting would be part of the work. The word “new”, in the second tweet, has feeling at this point, the tweet has been marked as subjective, and the machine is going to attempt a semantic relation(Abirami and Gayathri, 2017).

### **3) Semantic Association**

words used in subject position-specific grammatical words in the sentiment lexicons that relate to the subject are linked by rules in which they are found. Since most tweets are brief and plain, the structure of sentences is simplified. If your document employs several feelings, most of them will be adjectives and a few will be verbs. An opinion can be stated in only two ways: direct or in the form of a reference. A direct opinion lexicon is a type of lexicon, which makes use of a preposition and a conjunction to signify one single subject or multiple subjects. At least two subjects, but are linked only through their associated sentimentality Direct saying 'my life essential' was presented. One can see in the above example that 'love' is the nominal subject, and 'Well' the noun that receives love. Most tweets appear in this framework. Verbs and adjectives are almost always associated with a topic. since the subject “Well” is an object of the verb “to life”, we must search for the object “life” P. We can deduce that 'My' is a pronoun, 'life' is a verb, and 'Well' is a noun, from the POSS tag result. Love associates to the subject, so it is designated as positive or negative(Campion and Campion, 2019).

### **4) Polarity Classification**

On the scale of positivity, personal tweets are positive, or in polarity, subjectivity, personal tweets are given a positive or negative stamp. To categories sentiment is done by assigning subjects to common words in tweets and assigning them to associated lexicons. the examples include 'life essential', 'my life essential', and 'life essential need'. SentiWord Corporation reports that 'life' has a numerical score of 0.625 Accordingly, we can deduce that Unifi has a positive sentiment. So as to better focus on comparison, the situation of the subject is relevant in the tweet, the adjective “well” appears. The noun that follows the comparative adjective is compared to the previously mentioned noun.

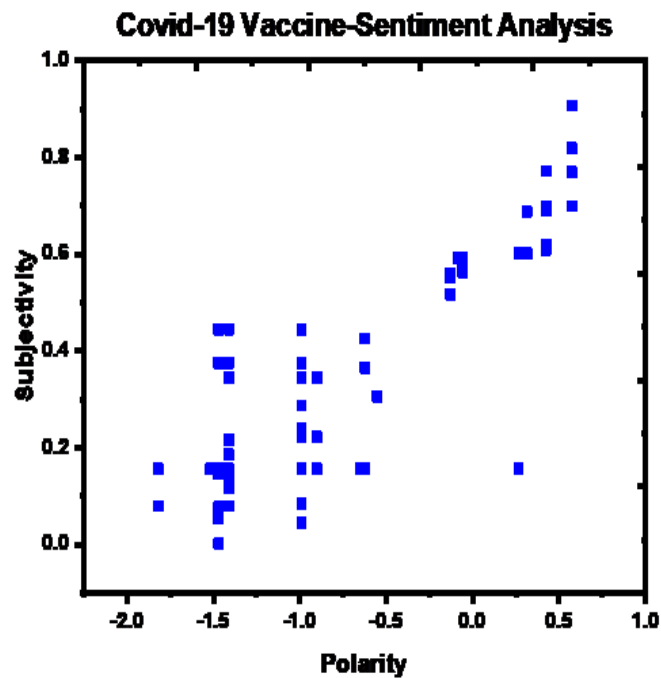
## **5. CASE STUDY: COVID-19 VACCINE**

After analyzing the material, we now have to do sentiment analysis and polarity classification of all the tweets that were extracted by the previous processes. In 2021, we have developed the case study “Covid-19 Vaccine-India” for the Indian government. to gain the support of the population The Twitter API (twitter 4j) gathered data below. With the aid of Twitter's built-in API, we have succeeded in retrieving our own messages. This is our use of the Twitter API in Python to pull data from the “#Covid-19 Vaccine-India” project. Our Python-based code was able to successfully retrieve 101 of the tweets that we were looking for on Twitter.



**Table 1: Retrieved sample tweets and its polarity**

Sample Tweets	Polarity of Tweets
Covid vaccine registration for above 18 years begins started from April 28th in India. The beneficiaries will be able to register for the Covid 19 vaccine.	Positive
According to the sources the European Commission is working on legal proceedings against AstraZeneca after the drug maker cut COVID-19 vaccine deliveries.	Positive
USA ain't happy India is selling its vaccine to poor countries without a one single penny a big lose for American companies which were cuddling covid-19 as an opportunity to earn billions of dollars in a first go Typical mindset	Positive
A thief who had fled with a bag containing over 1700 doses of COVID-19 vaccine in Haryana's Jind has had a change of heart.	Negative
More than 1700 doses of Covishield and Covaxin were stolen in Jind district of Haryana. The theft took place at the Jind Civil Hospital.	Negative
Kids are not fine... Lives have been turned upside down by this pandemic Kids including very young children can develop COVID19 Many of them have no symptoms but can spread it to others Those that do get sick tend to experience milder symptoms such as low-grade fever fatigue and cough.	Negative
Ability of vaccines to protect against COVID-19 variants in India not yet fully characterized.	Negative
Australia's coronavirus vaccine rollout is being recalibrated with people aged over 50 to be able to receive the AstraZeneca vaccine as soon as May.	Neutral
Oregon National Guard assist with COVID-19 vaccine distribution Oregon Air National Guard Senior Airman Angelica SorianoCervantes assigned to the 142nd Wing discusses the COVID19 vaccine to a walking patient at the Oregon Convention Center Portland Ore Jan 27 2021.	Neutral

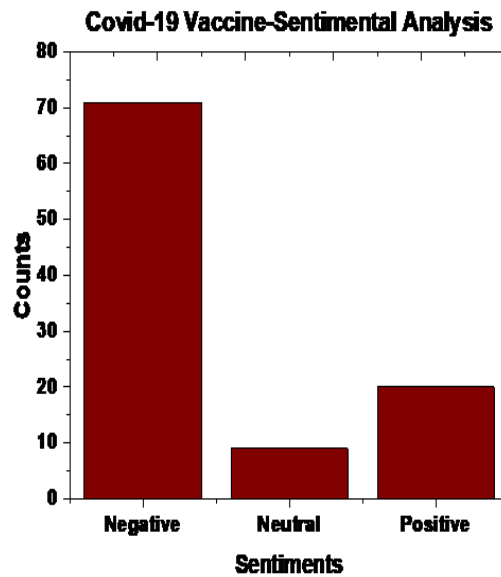


**Figure 3: Sentiment Analysis (Polarity vs Subjectivity)**

The collected tweets were done to both positive and negative terms as explained above (Figure.3). we wanted to accomplish the following three objectives in this research This was the first step we needed to complete with the twitter data collection. Our Python code decided to identify the tweets according to their emotional polarity. Finally, they are assessed on the basis of a numeric ranking.

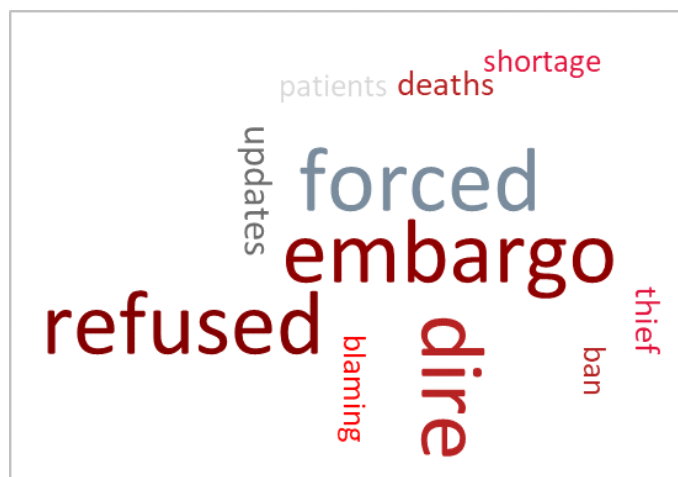
The numbers show that our code returned 20 favorable classifications, 9 unfavorable ones, and 71 unfriendly ones. When visualizing data using bar charts, classifications are described in graphical form Under this head-to-to-head comparison, 20% was in favor of the idea, 9% supportive, and 71% opposed to it (Figure.4).





**Figure 4: Percentages of sentiments**

The findings of our study show that CVI is unpopular with the Indian public on microblogging site Twitter. 80% of the consumers believe the campaign to be favorable, and 20% believe it to be unfavorable. In the beginning of the results portion, we were able to succeed in creating moods for the peoples as described (Figure.5).



**Figure 5: Word cloud of sentiment analysis**

## 6. CONCLUSION

Lots of studies have been done on how to correctly analyse sentiment from tweets, as Twitter is a very social site. We report on a preliminary study we've recently done, where we utilized NLP to discover subjects and identify their positive and negative sentiment by building a lexical network that associates with each topic. This paper was meant to represent the emotions found in twitter data. Covid-19 will help us accomplish our goals, as well as a Case Study of Vaccine India Using emoticons and infelicitous

messages to convey negative sentiments is among the tasks, an attempt to resolve credibility of Sentiment analysis is a primary consideration, followed by disadvantages and added value.

## References

- Abirami, A.M., Gayathri, V., 2017. A survey on sentiment analysis methods and approach, in: 2016 Eighth International Conference on Advanced Computing (ICoAC). IEEE, pp. 72–76.
- Akyol, S., Alatas, B., 2020. Sentiment classification within online social media using whale optimization algorithm and social impact theory based optimization. *Phys. Stat. Mech. Its Appl.* 540, 123094.
- Aquino, P.A., López, V.F., Moreno, M.N., Muñoz, M.D., Rodríguez, S., 2020. Opinion Mining System for Twitter Sentiment Analysis, in: International Conference on Hybrid Artificial Intelligence Systems. Springer, pp. 465–476.
- Arora, I., Guo, J., Levitan, S.I., McGregor, S., Hirschberg, J., 2020. A novel methodology for developing automatic harassment classifiers for Twitter, in: Proceedings of the Fourth Workshop on Online Abuse and Harms. pp. 7–15.
- Champion, M.A., Champion, E.D., 2019. Literature Review of Computer-Assisted Text Analysis Research, Software, Analytical Techniques, and Best Practices.
- Conrad, F.G., Gagnon-Bartsch, J.A., Ferg, R.A., Schober, M.F., Pasek, J., Hou, E., 2019. Social media as an alternative to surveys of opinions about the economy. *Soc. Sci. Comput. Rev.* 0894439319875692.
- Ford, E., Curlewis, K., Wongkoblap, A., Curcin, V., 2019. Public opinions on using social media content to identify users with depression and target mental health care advertising: mixed methods survey. *JMIR Ment. Health* 6, e12942.
- Garvey, M.D., Samuel, J., Pelaez, A., 2021. Would you please like my tweet?! An artificially intelligent, generative probabilistic, and econometric based system design for popularity-driven tweet content generation. *Decis. Support Syst.* 144, 113497.
- Ghiassi, M., Lee, S., 2018. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Syst. Appl.* 106, 197–216.
- Ismail, N.H., Liu, N., Du, M., He, Z., Hu, X., 2020. A deep learning approach for identifying cancer survivors living with post-traumatic stress disorder on Twitter. *BMC Med. Inform. Decis. Mak.* 20, 1–11.
- Jang, H., Jeong, Y., Yoon, B., 2021. TechWord: Development of a technology lexical database for structuring textual technology information based on natural language processing. *Expert Syst. Appl.* 164, 114042.
- Kim, E., Klinger, R., 2018. A survey on sentiment and emotion analysis for computational literary studies. *ArXiv Prepr. ArXiv180803137*.
- Kumar, S., Chong, I., 2018. Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. *Int. J. Environ. Res. Public Health* 15, 2907.
- Lv, W., Wang, Y., Zhou, C., Yuan, M., Pang, M., Fang, X., Zhang, Q., Huang, C., Li, X., Zhou, Z., 2021. Development and validation of a clinically applicable deep

- learning strategy (HONORS) for pulmonary nodule classification at CT: A retrospective multicentre study. *Lung Cancer* 155, 78–86.
- Mathapati, S., Manjula, S.H., 2017. Sentiment analysis and opinion mining from social media: A review. *Glob. J. Comput. Sci. Technol.*
  - Neha, M.A., Rahman, M., Nuzhat, M., Zereen, S., 2017. Evaluating user influence in Twitter based on hashtags using data mining (PhD Thesis). BRAC University.
  - Poornima, A., Priya, K.S., 2020. A comparative sentiment analysis of sentence embedding using machine learning techniques, in: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, pp. 493–496.
  - Reyes-Menendez, A., Saura, J.R., Alvarez-Alonso, C., 2018. Understanding#WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. *Int. J. Environ. Res. Public. Health* 15, 2537.
  - Thomaz, G.M., Biz, A.A., Bettoni, E.M., Mendes-Filho, L., Buhalis, D., 2017. Content mining framework in social media: A FIFA world cup 2014 case analysis. *Inf. Manage.* 54, 786–801.
  - G.Roja, N.Tulasi Chitra, K. Pushpa Rani , B.Dhanalaxmi, A Credit Card Fake Detection System Using Image Cryptography, *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-7, Issue-6, March 2019.
  - K. Pushpa Rani, M Jhansi, Chandrasekhara Reddy T. Best keyword cover search using keyword-nne algorithm, *International Journal of Mechanical Engineering and Technology*, 2017, 8(7), pp. 37–43
  - Sowmya. G, Navya K, Divya Jyothi G., Machine learning and mining for social media analytics, *Advances in Intelligent Systems and Computing*, 2019, 815, pp. 397–405
  - Sameera Divanu, “Detection of abnormality in CCTV Footage: Computer Vision” at *European journal of molecular & clinical medicine(EJMCM)*,2020,Volume 7, Issue 4, Pages 1148-1154.
  - Sameera Divanu, “Detection of Brain tumor using neural networks” at *European journal of molecular & clinical medicine(EJMCM)*,2020,Volume 7, Issue 4, Pages 1155-1161.
  - Sameera Divanu,“A Novel Approach for Text Classification using Recurrent Neural Networks” at *international journal of advanced science and technology(IJAST)*, Vol. 29 No. 05 (2020)
  - Sameera Divanu, “Stability of Hopfield in Higher Dimension” at *IARJSET*, Volume 6, Issue 7, July 2019, DOI 10.17148/IARJSET.2019.6707.
  - Sameera Divanu, “Lexical normalization for social media text in English” at *Airo international Research Journal peer reviewed Multi-disciplinary Indexed Journal* , UGC Approved no. 63012, ISSN : 2320-3714, Vol :18, Issue:3, Dec 2018.
  - Sameera Divanu, “A Comprehensive survey of Text mining techniques and applications” at *Airo international Research Journal peer reviewed Multi-disciplinary Indexed Journal* , UGC Approved no. 63012, ISSN : 2320-3714, Vol :15, Issue:3, Mar 2018.