

# Sentimental Analysis of Tweets During COVID-19 Pandemic: BERT Algorithm



Gurkirat Kaur, Munish Saini, and Amit Chhabra

**Abstract** As the coronavirus (COVID-19) grows its impact from China, expanding its catchment into surrounding regions and other countries, increased national and international measures are being taken to contain the outbreak. This perspective paper is written to capture and analyze the various mental state health issues being perceived via emotional analysis of Twitter data during the COVID-19 virus outbreak from a single nation further spread of to the whole world. A data-driven approach with higher accuracy as here can be very useful for a proactive response from the government and citizens. In the proposed work, tweets during the COVID situation have been collected and their sentiments are explored using BERT (Bidirectional Encoder Representation from Transformer) algorithm. BERT is the algorithm that takes text as input, and the trained basis on the epochs (number of passes performed). The performance parameters are computed such as accuracy, precision, recall, and F-measure. Further, the proposed approach is compared with other existing algorithms such as Naïve Bayes (NB), support vector machine (SVM), and logistic regression (LR). The performance measures indicate that the BERT algorithm outperforms all other existing algorithms with an accuracy of 86.7% as compared to 67.3%, 63.4%, and 61.2% with Naïve Bayes, support vector machine, and logistic regression, respectively. The government and other medical health agencies can use the outcomes of this paper for implementing and taking preventative measures to maintain the good mental and physical health of medical staff.

**Keywords** COVID · Twitter sentiment analysis · Novel coronavirus · BERT · Prediction algorithm · Metrics

---

G. Kaur (✉) · M. Saini · A. Chhabra  
Guru Nanak Dev University, Amritsar, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022  
N. Marriwala et al. (eds.), *Emergent Converging Technologies and Biomedical Systems*,  
Lecture Notes in Electrical Engineering 841,  
[https://doi.org/10.1007/978-981-16-8774-7\\_11](https://doi.org/10.1007/978-981-16-8774-7_11)

117

# 1 Introduction

The health officials are the most prone to any devastating circumstance like a pandemic outbreak as a large number of patients are to be treated in a limited capacity with time constraints as well [7]. The effect of the pandemic not only causes physical pain and sufferings to the normal people of every society, but it has a huge mental impact on the persons who are held answerable for their cure and remedies during any highly evolving bad or even worse situations related to the health of *homo sapiens* [5]. The doctors, nurses, and even higher authority health officials, all have a huge responsibility on their shoulders to save people from the chaos and limit the worst outcomes of any pandemic like the COVID-19 virus. Every human is affected whether directly or indirectly during the sickness effective on a global level as shown in Fig. 1.

The state of mind of a normal person gets troubled and utterly disturbed when so much suffering is seen with their naked eyes. Such pain and remorse emotions are key factors in demotivating the person who is going to find a proper remedy to cure the suffering of their breed. The emotion of hope and faith gets lost somewhere during bad times that are affecting the whole world altogether [8]. Further, outcomes can be new ways of treatment and state-of-the-art new methods to be used to minimize the effects like localizing the spread as much as possible by quarantining affected citizens under proper surveillance.

People nowadays, express most of their emotions by sharing their views on social media Web sites like Facebook, Twitter, and other popular social media platforms by writing or even sharing images with captions along with them [15]. The popular trends are mentioned using a *hashtag*, through which the current hot topics can be tagged and further views are shared [11]. Tracking the emotions and sentiments of people during a pandemic will have mixed feelings of expressions due to widespread effects as in the whole world confirmed cases shown in Fig. 2.

In this paper, the focus is given on public opinion analysis via tweet data collected from tweepy API [2] in the English language to capture the various emotions and feelings expressed and then limiting the exploration toward the health-related text data.

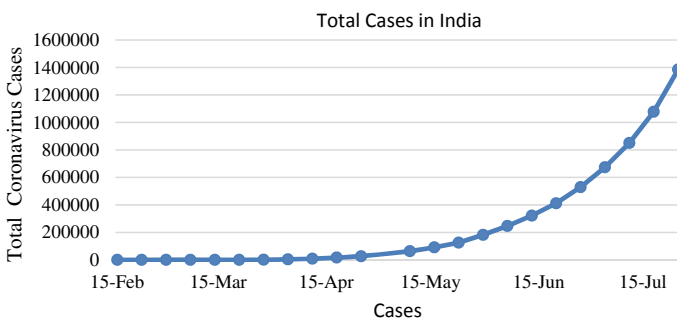
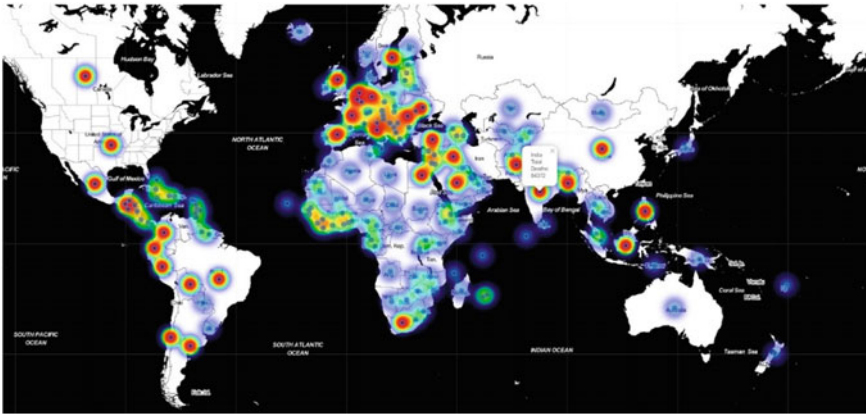


Fig. 1 Total cases in India during this pandemic [20]



**Fig. 2** Geospatial analysis of confirmed corona cases [3]

For performing this task, Twitter [9] is preferred because it is a famous social media platform having a plethora of information to be tapped and analyzed to preprocess as well as refine to gain the inside knowledge from the tweets. The tweets generally convey individual perspectives or feelings toward the subject referenced in the tweets. Sentimental analysis using natural language processing is one of the techniques that help to extract the feeling of the user in any situation [4]. Twitter provides an easier platform to retrieve user perspectives and feelings [18].

The objective of the research is to crawl through the various bundle of diverse sentiments articulated using Twitter, a social media platform. The crawled data is further refined and tokenized for accuracy in capturing the opinions and views of the people regarding health all over the world. Tweets with various health synonymous hashtags along with text data are used to analyze the psychological condition of various health officials.

Exploratory data analysis (EDA) is performed on tweets to check the performance of the model. It has achieved 82.3% accuracy. EDA is used to see the structure of the dataset collected. This step helps to expose patterns and relationships between the data. For further analyzing the data, natural language processing (NLP) and BERT methods are used, which brings out the resourceful information regarding peak levels of sentiments during the pandemic. Most of the authors have preferred machine learning methods to achieve high accuracy while analyzing the text data. Pang and Lee [17] have performed emotion classification using machine learning methods for the analysis of tweets. Besides applying machine learning strategies, natural language processing methods have been introduced. NLP helps to resolve uncertainty and add valuable information in the text. BERT is one of the best NLP techniques. BERT was designed by Google [19], which is a Fortune Five Hundred company. BERT is a trained model that works based on a transformer encoder. BERT set new state-of-the-art performance on various sentence classification and sentence-pair regression tasks. BERT uses cross-encoder [21].

This study will help the users, developers, researchers, academicians, and other stakeholders to comprehend as well as gain knowledge about the insights of the varied emotions running in the minds of the people around the globe. This further will help to get the state of mind of medical staff and other persons in any bad situation like pandemic, so that better measures can be taken beforehand.

In the rest of the paper, Sect. 2 places the work in the context of related work in this domain. Section 3 describes the research methodology. Section 3 discusses the methods, and a model for evaluating the popular tweets related to health are collected and used as part of the method process. Section 4 presents a discussion on the findings. We conclude the paper with some ideas for future work in Sect. 5.

## 2 Related Work

Much research interest has been directed toward sentiment analysis, particularly the challenges encountered in detecting word sentiment. The advent of technology has dramatically influenced human lives and their communities.

Sujatha et al. [21] have introduced a forecasting model to predict the COVID cases in India using the multilayer perceptron (MLP) method using the WEKA and Orange tools on COVID-19 Kaggle data and shown that their machine learning model is giving better results than their counterparts, that is, Linear Regression (LR) and vector auto-regression (VAR) methods. Mittal [13] researcher analyzes the current trend of COVID-19 based on certain criterion using “Exploratory Data Analysis.” Exploratory Data Analysis (EDA) is the way to explore the data to extract useful and actionable information from it. EDA is the revelatory step in any kind of analysis.

Majid et al. presented a diagnosis model for the novel coronavirus infection detection. Based on Bayesian optimization and deep learning mechanisms, the model is using the convolution neural network (CNN) layered architecture approach to assist the field specialists, radiologists, and physicians to make better decisions in diagnosing the novel COVID-19 virus in the patients in a faster and reliable manner.

Mardani et al. [12] have published a new way of using fuzzy logistics called hesitant fuzzy sets used in designing a novel framework to address as well as assess the key challenges faced in the digital health during the pandemic outburst of coronavirus globally. By combining the unique fuzzy approach with Stepwise Weight Assessment Ratio Analysis (SWARA) and Weighted Aggregated Sum Product Assessment (WASPAS) methods, the work is done to rank the life-threatening challenges being faced by the current digital technologies while controlling the COVID-19 situation.

Tuli et al. [22] proposed a novel method called Deep Bayes-Squeeze Net-based COVIDiagnosis-Net to classify the COVID-19 cases as the COVID-19 or normal (healthy). The model ensures an end-to-end learning schema that can directly learn discriminative features from the input chest CT X-ray images and eliminate handcrafted feature engine.

Devlin et al. [6] developed Bidirectional Encoder Representations from Transformers (BERT) at Google AI Language. BERT is “*designed to pretrain deep bidirectional representations from the unlabeled text by jointly conditioning on both left and right context in all layers.*” The state-of-the-art BERT is pretrained on two unsupervised tasks—masked language modeling and next sentence prediction, thus making it an effective technique for eight sentiment classification. BERT is known to have achieved exceptional results in eleven natural language understanding (NLU) tasks.

Wang et al. [23] investigated the influence of air temperature and relative humidity on the transmission of COVID-19 by calculating the “effective reproductive number”(R), and under the “Linear Regression” framework, they found out that a one-degree Celsius rise in temperature and one percent increase in the relative humidity lower R by 0.0225 and 0.0158, respectively, and indicates that arrival of summer and rainy season in the northern hemisphere can effectively reduce the transmission of COVID-19.

Inferring from the latest trend nowadays as per the hashtags captured using Twitter data (tweets), it can be concluded that the people are having a rich amount of intensifying emotions. The state of the mindset especially of the people who are going to defend all of us from the outspread virus needs to be considered. The captured Twitter data is based on health-related tags to further narrow down our Web crawling and limiting the tweets, making it more relevant to this research.

## 3 Experimental Setup and Analysis Methodology

### 3.1 Experimental Setup

In this study, tweets are collected related to mental health during the COVID situation. The experiment is performed on Jupyter Notebook using Python 3.3.0. Dataset is taken in CSV format, with 20 columns such as TweetPostedTime, Tweet ID, Tweet-Body, TweetHashtags, UserID, UserName, etc. Firstly, Exploratory Data Analysis has been performed after that text has been classified using BERT algorithm.

### 3.2 Experimental Setup

This section discusses the steps performed for analysis (see Fig. 3).

#### Data Collection

The data was collected using the Tweepy API [2] for Python in which all the latest tweets related to the COVID pandemic were captured (refer to Table 1). All the tweets have emotion and that was what needed for the research, and by applying sentiment

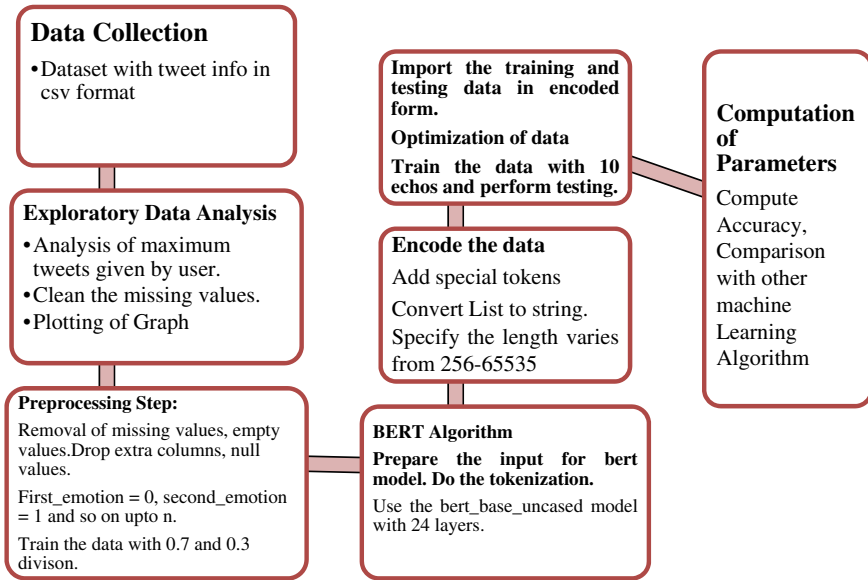


Fig. 3 Methodology of proposed work

Table 1 Attributes of the dataset

Attribute	Description
TweetPostedTime	It represents the timestamp of the collected tweet
TweetID	A Unique identifier associated with each tweet
TweetBody	It includes text mentioned in the tweet
TweetRetweetFlag	It is a flag for the indication of tweet retweet status. It is either true or false. The true flag represents the tweet is retweeted and false specifies it is not retweeted
TweetSource	It contains the location information for the tweet origin
TweetRetweetCount	It consists of the count of total retweets
TweetFavoriteCount	It stores the count of marked tweets
TweetHashtags	It represents the hashtags that are mentioned in the tweet
UserID	It specifies the unique user identifier
UserName	It represents the name of the user
UserLocation	The geographical location of the user
UserDescription	It contains the information related to the user
UserFollowersCount	It specifies the total number of followers

analysis, the captured tweets were grouped based on the emotions they possess. The usernames are kept anonymous focusing solely on the text in the tweet. The timestamp of the tweet is also considered to relate to the sentiments as time plays a vital role in understanding the mindset of the person, while they tweet [16]. The dataset that was considered, covered all the various other aspects of the tweet as well, such as time of the tweet, the description of the user, and the hashtags used in the tweet alongside description about the user posting the tweet. All these factors, when combined, provide more insight about the sentiments of the user while tweeting, and thus, we can further understand the state of mind of the user during the pandemic.

### Exploratory Data Analysis

Exploratory Data Analysis (EDA) [1] is the first step while starting the analysis of the dataset. This process mainly focuses on efficiently understanding the dataset. The process of EDA is graphically presented in Fig. 4.

Figure 5 represents the top 10 trending hashtags that were exploited during the rise of the pandemic, having #*covid* as the most repeatedly used. It further depicts the mental state of the users who were more concerned about the disastrous outbreak rather than any other topic during that time.



Fig. 4 General steps of exploratory data analysis

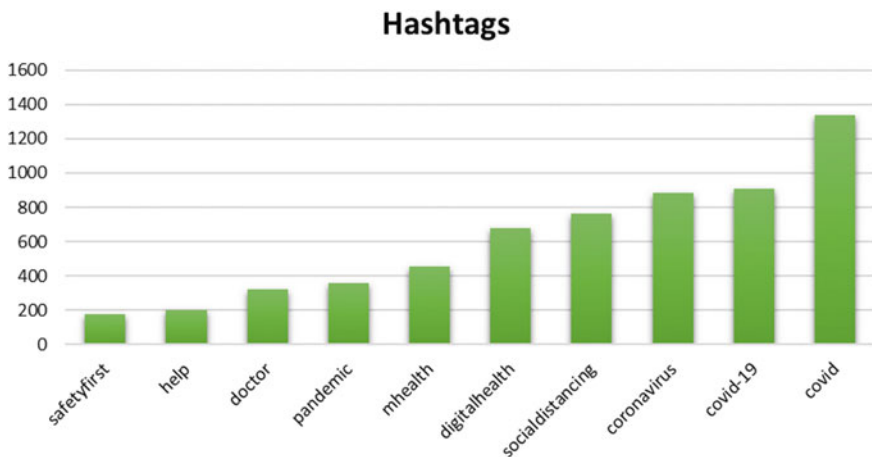
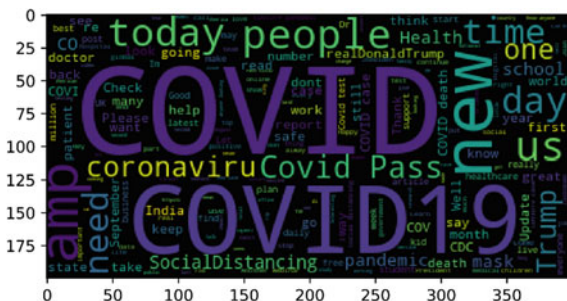


Fig. 5 Top 10 hashtags corresponding to their count

**Fig. 6** Word cloud of tweet body

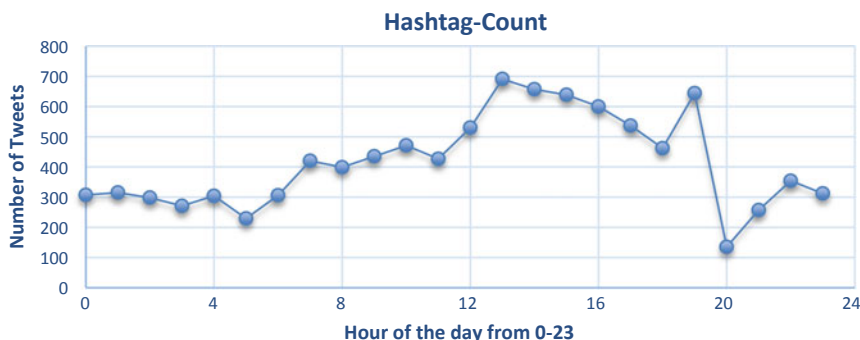


Moreover, analyzing the frequent hashtags is essential, but the Tweet\_Body has much more rather than just hashtags. The text posted in the tweets is having as much significance as a hashtag. Figure 6 shows the wordcloud of most frequent phrases used in the Tweet\_Body during the COVID outbreak, with the size of the text conveying the frequency of repetition of that word in the dataset.

This paper concentrates on the mental health analysis of a person, evident to reveal that the mood and emotions of a person get influenced by the time of the day as well. As shown in Fig. 7, the thoughts can be witnessed to be at their peak levels from 12:00 to 16:00 h of the day. Apparent to say, in this particular time, when people usually have their lunch, either alone or with close ones, primarily discuss and share their mental state of mind, making the data being presented in Fig. 7 much more obvious and appropriate at the same time.

### Data Preprocessing

The tweet data has been extracted based on various hashtags such as #doctors, #socialdistancing, #coronaheros, #digitalhealth, #mhealth, #healthinnovations, and #medtech. Using these hashtags, 10,000 tweets are extracted. These all are labeled based on their emotions such as ['fear', 'surprise', 'sad', 'happy', 'trust', 'anger']. In the preprocessing steps, data cleaning has been performed, and in this step, all null



**Fig. 7** Count of tweets corresponding to their hour



**Table 2** Labels of sentiments

Sentiment	Label
Fear	0
Pride	1
Sad	3
Disgust	6
Sad	4
Surprise	5
Happy	2
Anger	3

has been extracted. Each emotion has been labeled using the enumeration function as given in Table 2.

### Classification Using BERT Algorithm

BERT (Bidirectional Encoder Representations from Transformers) [24] model is used for unlabeled data to give bidirectional representation. This algorithm works from both the context, i.e., from left and right both. BERT is one of the efficient techniques to get high accuracy in natural language processing. The basic steps of the BERT algorithm are shown in Fig. 8.

To perform the text classification, we used a BERT model known as an advance supervised model. This component has used the emotion labeled data such as different emotions of doctors (sad, happy, and many more). These emotions are labeled further from 0 to n where n is the number of emotions. In our experiment, we have used the BERT of HuggingFace [24]. The following Fig. 9 shows the steps of the proposed algorithm.

### Computation of Parameters

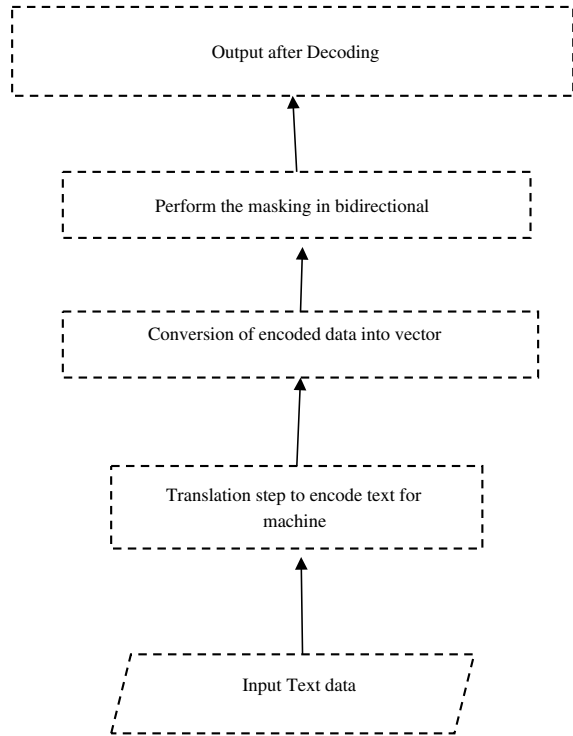
In this paper, the model is focused to compute the accuracy of positive emotions and negative emotions: positive emotions are considered happy and trust, and other emotions are considered in a negative impact [14]. The performance is measured on basis of true positive, true negative, false positive, and false negative. The following Fig. 10 shows the meaning of these terms.

### Comparison of BERT with Traditional Machine Learning Algorithms

In this section, the BERT algorithm results are compared with support vector machine (SVM), Naïve Bayes (NB), and logistic regression (LR). These machine learning algorithms are used because of their effectiveness and performance.

Logistic regression (LR) [25] is the algorithm used to find the data which is dependent, and this helps to find the relation between the independent variable. Here, this model will help to predict the emotions are sad or anger based on other tweet emotions used for training. This model is useful for linear as well as nonlinear data. The LR model returns the 1 if true emotion and 0 if false emotion predicted. Here,

**Fig. 8** Flow of BERT algorithm



the model is used for tweets, to find their probability of occurrence in this pandemic. Consider the tweets and their sentiments, the LR model will give the attributes of  $A(S|T)$  where  $S$  is the class of sentiments and  $T$  is tweets retrieved.

$$A(\text{class of sentiments} | \text{Tweets}) = A(\text{Tweets}) \quad (1)$$

It will return output in two values 1- for true predicted emotion values and 0 – for false predicted emotion values. In this,  $A(S|T)$   $S$  are sentiments, and  $T$  is text so sentiment range varies from 0 to  $n$ . Therefore, the equation will be used as an exponential function [10].

$$A(S|T) = \frac{1}{x} e^{w^T y} \quad (2)$$

where  $x$  is the normalizing factor. Naïve Bayes (NB) [9] is the algorithm that works on Bayes rule. This algorithm is used for text classification based on a supervised algorithm. In this paper, we computed the  $P(S)$  and  $P(T|S)$  using the equation.

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)} \quad (3)$$

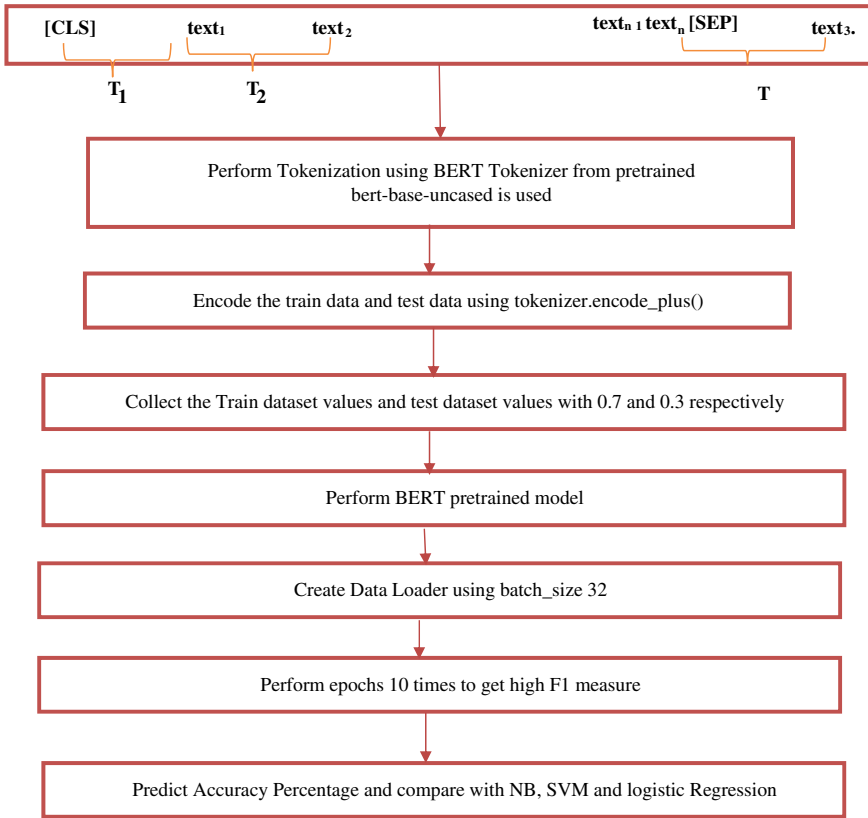


Fig. 9 Steps of proposed approach where T1, T2.... Tn is tweets

Predicted Emotion -ve	Predicted Emotion +ve	
False Negative	True Positive	Actual Emotion +ve
True Negative	False Positive	Actual Emotion -ve

Fig. 10 Confusion matrix

For estimating P(S), the relative frequency of each tweet has been targeted in trained data. Consider different tweets as T1, T2,..... Tn are attributes and computed using

$$P(T1, T2, \dots, Tn|S)P(S) = \prod_k P(Tk|S) \tag{4}$$

Support vector machine (SVM) is used for text classification. It focuses on more features, and estimate the discriminate function.

$$y(x) = w^T g(x) + b \quad (5)$$

where  $w$  is weight vector,  $g(x)$  is feature space, and  $b$  is biased. Here, due to a large number of tweets, the classification will be performed linearly separable.

## 4 Experimental Results and Discussion

After capturing the data and further assessing the acquired knowledge from the information gained by applying the BERT algorithm, we found these results which clearly explain the plethora of emotions running in the minds of people all around the world. To validate our proposed model, we have compared the proposed model with Naïve Bayes (NB), support vector machine (SVM), and logistic regression (LR). The parameters computed are accuracy, precision, recall, and F-measure.

The accuracy is defined as the total number of tweets and emotions are classified correctly. This parameter is to check the performance of a complete model that includes all the emotions.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{false positive} + \text{false negative}} \quad (6)$$

The precision is also known as specificity. It is the ratio of correctly classified tweets to all of the correctly predicted tweets.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{false positive}} \quad (7)$$

The recall is also known as sensitivity. Recall defines as the ratio of true positive and addition of true positive and false negative.

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{false negative}} \quad (8)$$

F-measure is also known as the F1-score. This term has a high predictive success rate. It can be calculated using the following formula:

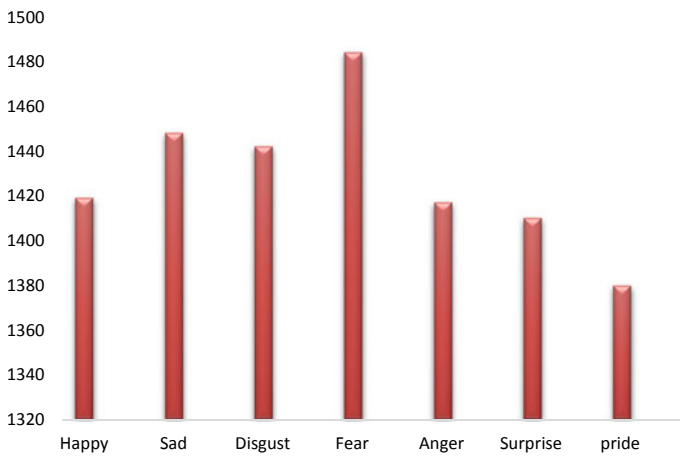
$$F - \text{measure} = 2 \left( \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \right) \quad (9)$$

The tweets contain various emotions, the count of emotions is shown in Table 3, and Fig. 11 shows that fear emotion is more prominent during this COVID pandemic.

The data is divided into train and test data. After that, it has been compared with other algorithms and resulted in Table 4, and Fig. 12 shows the value of performance metrics.

**Table 3** Dataset classification

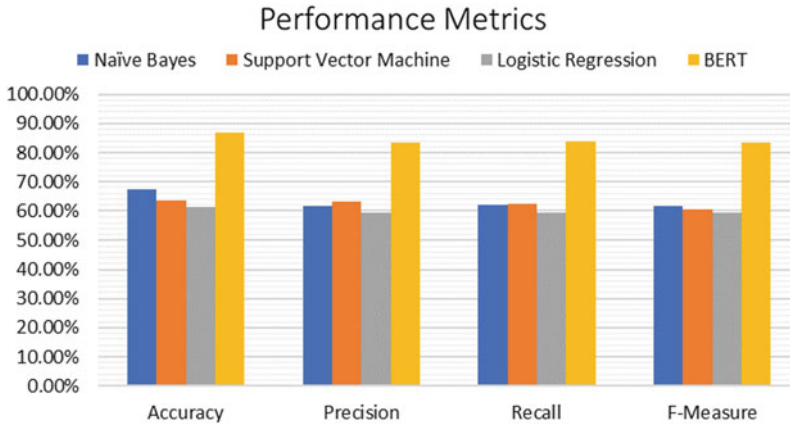
Sentiment	Count	Count of test and train data
Happy	1419	Test-420
		Train-999
Sad	1448	Test-438
		Train-1010
Disgust	1442	Test-398
		Train-1044
Fear	1484	Test-439
		Train-1045
Anger	1417	Test-430
		Train-987
Surprise	1410	Test-450
		Train-960
Pride	1380	Test-425
		Train- 955
Total	10,000	



**Fig. 11** Count of each emotion

**Table 4** Comparison of the proposed algorithm with Naïve Bayes, SVM, and logistic regression

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
Naïve Bayes	67.3	61.6	62.2	61.5
Support vector machine	63.4	63.0	62.3	60.3
Logistic regression	61.2	59.3	59.2	59.2
BERT (Proposed algorithm)	86.7	83.4	83.8	83.5



**Fig. 12** Performance metrics computation of all the algorithm

The findings from this research clearly show how bad a pandemic can affect the intensity of the emotional mindset in the human beings, creating fear and sadness as the most powerful ones in comparison to additional primary sentiments. The research further explicates how sentiments of normal people get so much affected during the time of the day and how popular social media platforms can help in capturing all their emotions through their expressional posts.

The result clearly states that the BERT algorithm outperforms the other major machine learning-based sentiment analysis algorithms in determining as well as predicting the actual emotions in text-based data. It is to be noted that other researchers can also make use of these findings to learn about a more realistic mindset of persons, which can easily get affected under various major circumstances.

## 5 Conclusion and Future Scope

In this research, the prediction of the states of minds of the people has been normalized as per the hashtags used to capture the text data. The Twitter live dataset is used to justify as well as forecast the various sentimental challenges to be seen during the novel coronavirus widespread. BERT is used as a proficient model to understand as well as determine the rush of sentiments and deep emotions of the people facing the struggle during the COVID outbreak including the health officials and the other experts related to the field of medical sciences.

The proposed model can predict the sentiment residing in the tweets posted by tweet users, which makes the model much more accurate in analyzing the tweet data fed to the model. The results further articulate the accuracy in using the BERT algorithm for sentimental analysis of tweets used as test dataset. BERT algorithm performed well, and it can handle various emotions like sadness, anger, and disgust.

Further, in the future, we can work more on other machine learning algorithms to deal with emotions others that positive, negative, and neutral.

## References

1. Alaparthi S, Mishra M (2020) Bidirectional encoder representations from transformers (BERT): a sentiment analysis odyssey. [arXiv:2007.01127](https://arxiv.org/abs/2007.01127)
2. Allen TT, Sui Z, Akbari K (2018) Exploratory text data analysis for quality hypothesis generation. *Qual Eng* 30(4):701–712
3. Belkacem S (2020) COVID-19 data analysis and forecasting: algeria and the world. [arXiv:2007.09755](https://arxiv.org/abs/2007.09755), RoessleinJoshua. “tweepy Documentation. <http://tweepy.readthedocs.io/en/v3.5> (2009)
4. Cambria E, Poria S, Gelbukh A, Thelwall M (2017) Sentiment analysis is a big suitcase. *IEEE Intell Syst* 32(6):74–80
5. Ćosić K, Popović S, Šarlija M, Kesedžić I (2020) Impact of human disasters and Covid-19 pandemic on mental health: Potential of digital psychiatry. *PsychiatriaDanubina* 32(1):25–31
6. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
7. Galbraith N, Boyda D, McFeeters D, Hassan T (2020) The mental health of doctors during the Covid-19 pandemic. *BJ Psych Bull*:1–4
8. Huang L, Rong Liu H (2020) Emotional responses and coping strategies of nurses and nursing college students during COVID-19 outbreak
9. Hughes DJ, Rowe M, Batey M, Lee A (2012) A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Comput Hum Behav* 28(2):561–569
10. Indra ST, Wikarsa L, Turang R (2016) Using logistic regression method to classify tweets into the selected topics. In: 2016 International conference on advanced computer science and information systems (ICACSIS). IEEE, pp 385–390
11. Kywe SM, Hoang TA, Lim EP, Zhu F (2012) On recommending hashtags in twitter networks. In: International conference on social informatics. Springer, Berlin, Heidelberg, pp 337–350
12. Mardani A, Saraji MK, Mishra AR, Rani P (2020) A novel extended approach under hesitant fuzzy sets to design a framework for assessing the key challenges of digital health interventions adoption during the COVID-19 outbreak. *Appl Soft Comput*:106613
13. Mittal S (2020) An exploratory data analysis of COVID-19 in India. *Int J Eng Res Technol (IJERT)*. ISSN: 2278-0181
14. Mohammad SM, Kiritchenko S (2015) Using hashtags to capture fine emotion categories from tweets. *Comput Intell* 31(2):301–326
15. Nabity-Grover T, Cheung CM, Thatcher JB (2020) Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media. *Int J Inf Manag*:102188
16. Naseem U, Razzak I, Musial K, Imran M (2020) Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Futur Gener Comput Syst* 113:58–69
17. Pang B, Lee L (2004) A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. [arXiv preprint cs/0409058](https://arxiv.org/abs/preprint/cs/0409058)
18. Parise S, Whelan E, Todd S (2015) How Twitter users can generate better ideas. *MIT Sloan Manag Rev* 56(4):21
19. Patel M (2019) TinySearch--semantics based search engine using bert embeddings. [arXiv:1908.02451](https://arxiv.org/abs/1908.02451)
20. Sahoo S, Bharadwaj S, Parveen S, Singh AP, Tandup C, Mehra A, Grover S (2020) Self-harm and COVID-19 pandemic: an emerging concern—a report of 2 cases from India. *Asian J Psychiatry*
21. Sujath R, Chatterjee JM, Hassanien AE (2020) A machine learning forecasting model for COVID-19 pandemic in India. *Stochast Environ Res Risk Assess*:1

22. Tuli S, Tuli S, Tuli R, Gill SS (2020) Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet Things*:100222
23. Wang J, Tang K, Feng K, Lv W (2020) High temperature and high humidity reduce the transmission of COVID-19. SSRN 3551767
24. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Brew J (2019) HuggingFace's transformers: state-of-the-art natural language processing. Arxiv-1910
25. Zhang Z (2016) Model building strategy for logistic regression: purposeful selection. *Annals Trans Med* 4(6)