



# Sentiment Analysis of News on Human mind Using NLP and Deep Learning Approach

<sup>1</sup>Rutwik Keskar , Bachelor of Engineering ,Dept. of Computer Science,  
Pune Institute of Computer Technology, Pune.

<sup>2</sup>Antara Joshi , Bachelor of Engineering ,Dept. of Electronics and Telecommunication  
Vivekananda Education Society's Institute of Technology, Mumbai.

**Abstract :** The chronic nature of the coronavirus disease (COVID-19) outbreak in late 2019 has brought several devastations to the world. The lack of success in treatment and cure is creating an environment that is crucial for mental wellbeing. The pandemic situation has spread across the globe. This situation has infected more than ten million people and it has disrupted various businesses in the world. The physical distancing and the use of protective masks were suggested, but still, the cases seem to rise. This led to worldwide lockdown in different phases across the world. We have extracted and classified sentiments and emotions from news channel headlines about the coronavirus disease (COVID-19). The headlines considered were those carrying keyword coronavirus news sources. The headlines are classified into positive, negative, and neutral sentiments. The text unbound polarity at the sentence level score is calculated by incorporating the valence shifters. This paper investigates the impact of news on the human mind using sentiment analysis. The research follows a deep learning mechanism to achieve the goal. Deep learning deals with a very large amount of data efficiently and performs the complex task. Deep learning is built explicitly for dealing with a significant amount of large numbers that perform complex tasks which in turn perform automatic learning that is necessary. In this paper, we perform a sentiment analysis of news data during the pandemic lockdown. We have taken a time of two and four weeks after the lockdown was imposed. Investigating the sentiments of people in the form of positive, negative, and neutral news would assist us in determining how people are dealing with the pandemic and its effects on their psychological levels. The study will be effective in the determination of people's mental well-being. Also, this will be useful in devising appropriate lockdown strategies and crisis management in the future.

**Keywords:** Coronavirus, Sentiment analysis, Lockdown, COVID-19, Cognitive study, Human psychology

## I. INTRODUCTION

The COVID-19 was discovered in Wuhan, Hubei Province, China, in late 2019, but the novel Coronavirus has been spreading at an alarming rate nowadays. The COVID-19 disease is associated with the human respiratory system. It is believed that this disease has potentially originated from either bat or the Huanan seafood market in China. There are various such types of pandemic diseases like SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome), and COVID-19 (Coronavirus Disease-19). These all diseases are associated with bronchitis, pneumonia, or severe respiratory illness. Because of the increasing number of cases, the World Health Organization declared COVID-19 as a pandemic within three months after its discovery.

The outbreak of COVID-19 has infected more than 10 million people across the world and there are reported deaths of more than a half-million worldwide. The COVID-19 has impacted on the number of deaths and several businesses worldwide. The travel industry has impacted more due to heavy travel restrictions. The impact has happened on the stock market prices and crude oil prices also. The remaining industries such as tourism, sports, and restaurants are also impacted due to this pandemic situation.

There is a lack of vaccines for COVID-19 diseases and its major factor is responsible for the uncontrollable spread of the infection. Although many researchers have suggested a variety of treatments and proposed medicines, the results are not very promising since an established cure could not be developed even after six months of the discovery of the COVID-19, social distancing, and other lockdown strategies were suggested by various government bodies all over the world. Because of the lockdown imposition across the various parts of the world, restrictions are placed on the major facilities and services. So the general public were not able to get these amenities

The main goal of the research is to evaluate the performance of the Deep learning algorithms and apply NLTK techniques on the mindset on human during COVID-19 period

*Objectives:*

- To identify the algorithms and metrics for performance evaluation of Deep Learning algorithms.
- To compare the metrics obtained from different deep learning algorithms depending on the size of datasets.
- To identify the best algorithm suited for sentiment analysis.

### *1.2 Research Questions*

The research methodology needs to answer the following questions:

- i. Which deep learning algorithm is used for analysis and performance evaluation of Fine-grain sentiment analysis on Twitter data?
- ii. Which is the best fit classification algorithm for conducting sentiment analysis on news data?
- iii. How does the performance of the deep learning algorithm vary as the quantity of labeled and unlabeled data changes?

### *1.3 Ethical Considerations*

The recording details used in this study will remain private and never be revealed to the third party. All the participants will not be compelled to share their personal information and the

## **II. LITERATURE REVIEW**

In Literature review, we have explored the past research works that have been performed with respect to the COVID-19 pandemic outbreak.

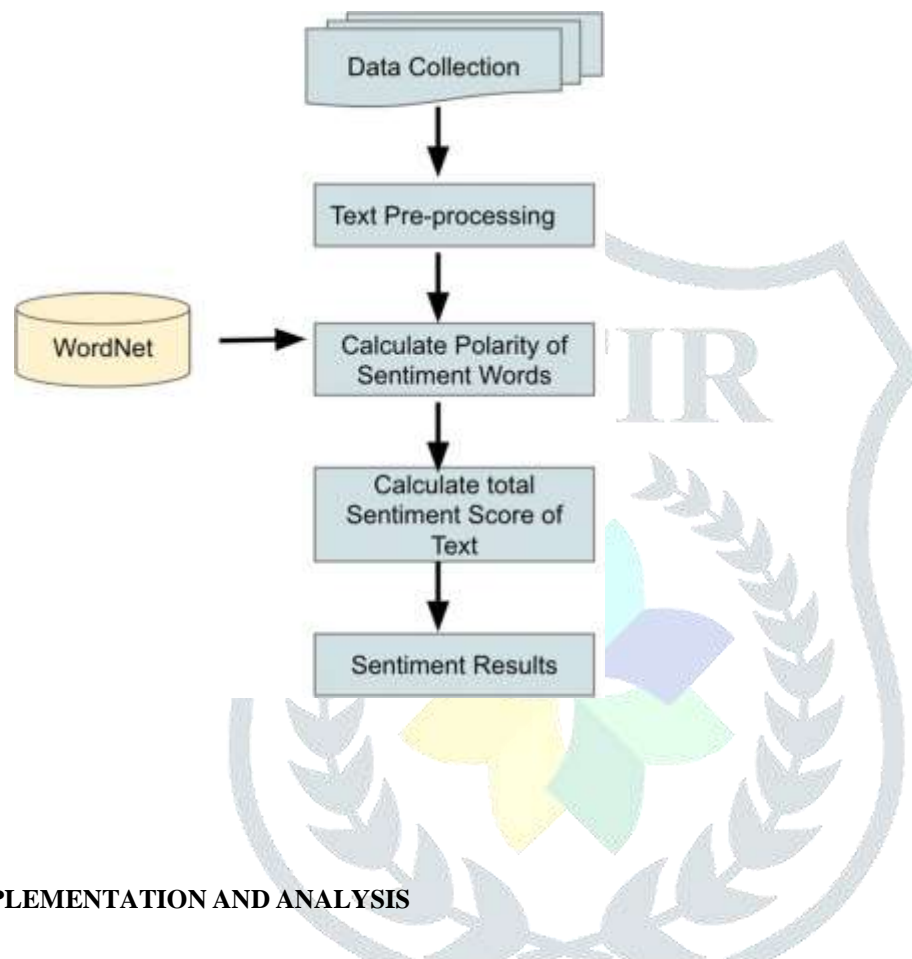
The performance of a Twitter sentiment analysis during the COVID-19 outbreak for specific countries like Australia, China, India, United States of America. This study has taken into consideration eight basic emotions such as joy, fear, anger, sadness, anticipation, disgust, surprise and trust. The tweets related to anger are highest in France. The tweets related to anticipation are highest in Germany. There are a maximum number of tweets from the USA related to this COVID-19 disease. Switzerland reports fear about this pandemic disease. India has reported the highest percentage of joy, but Switzerland reports the maximum percentage of sadness. The highest number of tweets that show surprise and trust are from Belgium. There is a limitation of this study. This study does not take into account the sentiments of the people across the world. This is only confined to a few countries only [12]. This is a report on the analysis and forecast of COVID-19 disease outbreak in countries such as China, Italy and France. This analysis is done by using susceptible-infected-recovered-deaths models. This model indicates the rate of recovery is irrespective of the country. But the death rates and infections vary from country to country. The case-fatality ratio is shown to be 4% to 8% in Italy and between 1% to 3% in China. The study reveals that at about 2500 ventilation units are a fair figure for COVID-19 pandemic strategic management in Italy. The study also shows how the case fatality ratio has been decreased. The limitation of the research is that it is narrow and it considers a very small fraction of individuals from only three countries [4] provided a risk assessment strategy for the novel Coronavirus in the form of a computational tool.

This study also reveals that the countries having low connectivity with China and having higher efficacy measures can reduce the risk in a better manner. But there is one limitation of this study, it considers the destination countries from only China and COVID-19 is a global issue. The performance analysis of the outbreak of the COVID-19 and presented a comparison of the cases confirmed, recovered, and deaths with respect to China using data visualization. The study also stated some industries that were severely affected by the pandemic, like tourism, travel, and the sports industry.

There is a limitation of this study, as it fails to discuss many global issues. The global industries such as real estate, supply chain management etc. are also affected because of this pandemic situation [1]. The analysis identified specific themes, i.e., origin of the virus, its sources, and its impact on people, countries, and the economy, ways of mitigating the risk of infection. Most of the tweets were classified as positive. The negative tweets were based on deaths due to COVID-19 and increased racism. The highest mean of tweets was for economic loss and the lowest mean of tweets was for travel bans and warnings. The analysis has been performed with respect to the COVID-19 pandemic, and not with respect to the psychology of the public following the lockdown. [16] proposed a strategy for real-time estimation of risk of death from COVID-19 infection. The study relies on the exponential growth rate of incidence to estimate the basic reproduction number and confirmed case fatality risk. The results estimated a growing number of cases by the end of January suggesting the potential possibility of the infection culminating into a pandemic. The study takes into account limited empirical data confined to analyze only the case fatality risk. The veracity of the assumed date of onset is questionable and detection window time is uncertain. [31] Suggested a binary classification and regression analysis methodology for investigating the COVID-19 outbreak for the sustainable development process. The classification model produces accurate results with respect to confirmed cases. The regression analysis has been used for comparing the fluctuations of parameters like wind, humidity, and average temperature. The results manifest that maximum daily temperature and relative humidity have the highest impact on confirmed cases. The case study is restricted to China, hence the number of cases considered is a fraction of the total number of cases. Hence, the results may be specific. [29] Conducted a sentiment analysis of students Synchronous Online Delivery of Instruction in the Philippines as a result of community quarantine during the COVID-19 pandemic. Most of the respondents forecasted that they may face issues and many respondents are worried about network connectivity. A negative sentiment analysis (66%) indicates that most of the students may not be able to adapt to the new trends of education. The limitation of this study is that the dataset considered is immensely small focusing on a specific institution. [34] Recommended a quicker method of identifying COVID-19 using an artificial intelligence framework. The framework is based on a mobile phone survey using which data may be collected during the quarantine. The data comprises information related to an individual's location details, demographic information, travel information, signs, symptoms, etc. The only limitation of the study is that it is a device and survey-based application dependent. [20] presented a review of modern technologies that may assist in tackling the COVID-19 pandemic. The methods mentioned in the study are Diagnosis using radiology images, disease tracking, Prediction outcome of patient's health condition, Computational Biology and medicine perspective, Drug discovery, protein structure predictions, and awareness through the internet. The methods prescribed are certainly interesting, however, according to the past research works, the accuracy of the suggested methods is not 100% [33] presented an analysis of 'Fear Injury Kidney' theory, according to which long term or excessive fear could lead to damage in the neuro-endocrine-immune system. This in turn leads to diseases, and several groups are observed to have the influence of stress and fear on the body which cannot be ignored. [27] analyzed the psychological impact of COVID-19 on the elderly population in China. More than fifteen hundred elderly people representing various ages were considered for the study. More than 37 % of seniors experienced depression and anxiety due to the pandemic. The study also shows that women experience more anxiety and depression as compared to men. The suggestions presented for such a situation included calling psychological hotlines, concentrating on seniors' mental health, and considering counseling and psychological crisis intervention [9] presented a study on the impact of COVID-19 on the psychology of more than fifteen thousand university students. Epidemic related stressors are observed to be related to the level of anxiety. The stressors revolve around the impact of the pandemic on the economy, daily lives, academic delays, and social support [40] explored the effect of COVID-19 lockdown on the psychology of people based on law enforcement. The enforcement of law led a large population to follow rules and restrictions during emergency situations, and quarantine people affected by the disease. The laws stated that citizens not abiding by laws would be penalized. All these impositions have been observed to cause distress among the general public.

In this section, we presented the past relevant works performed with respect to the COVID-19 outbreak. Based on the previous work, we can say that COVID-19 research is multi-faceted. Most of the research work done aims at analyzing the outbreak and finding a cure for the same. Very little work with respect to human psychology has been done in the past. One way to explore the domain is by performing sentiment analysis. Another shortcoming of the previous research works is that many research works fail to conduct the studies globally, and many times the study is limited to specific regions and a specific number of cases. In our research, we will be handling this shortcoming by targeting data from all over the globe on a social networking platform.

### III. PROPOSED SYSTEM MODEL



### IV. IMPLEMENTATION AND ANALYSIS

There is a significant number of research works related to COVID-19 outbreak analysis. There are clinical trials proposed before. But, the after-effects of the lockdown on human psychology still need to be more explored. To study the psychological effects of the COVID-19 pandemic on the minds of people, we consider the natural language processing technique (NLP) of sentiment analysis. Sentiment analysis explores the subjective opinions and feelings concerning a particular subject. It is done by collecting data from different sources like print media, electronic media, or social media. Sentiment analysis can be of various types like fine-grained, emotion detection, aspect-based, and intent analysis. In this paper, we determine the polarity of the sentences.

Also, we perform aspect-based sentiment analysis. It is a classification technique that categorizes sentences into positive, negative, and neutral categories. It is based on their polarity. The polarity or orientation is nothing but its emotion expressed in a sentence. In order to analyze the sentiment of texts, there is a need to analyze individual words or phrases' sentiment. In a sentence, there are several words that are positive and several words that are negative. The words presenting emotions and sentiments in a sentence are a challenge.

When a text is long, positive and negative then sentiments may tend to average towards neutral state. Hence it is important to separate genuine sentiments from basic phrases. The strong and clearly defined sentiments are denoted by the polar words and phrases. On the other hand, non-polar words and phrases are used in everyday life. Thus polarity gives the approximate sentiments

in a given text. In order to determine polarity, we depend on the python library named, Textblob. The Textblob is a tool built on top of the Natural Language Toolkit (NLTK). By using polarity, it is possible to determine whether a text is positive, negative, or neutral. In the TextBlob library, if the polarity of the word or sentence is greater than 0, the text is considered positive. A text is considered as negative if the polarity is less than 0. The sentence is considered as neutral, if the polarity of a text is equal.

## V. METHODOLOGY

The following steps have been performed on sentiment analysis.

**Step 1: Data Collection** - The data is collected from authentic and right sources. For our experimental purposes, we have collected data from Kaggle. We have collected news datasets of various news channels.

**Step 2: Text Preprocessing**- In this step the cleaning and preparation of text data is done. It may include several operations such as noise removal (removing unwanted data), tokenization (breaking the text into smaller components for accessing each word), stop-word removal (removing common words in the language which do not provide relevant information), etc.

**Step 3: Calculating polarity**- We can calculate polarity of the words by using the Textblob library.

**Step 4: The total sentiment score of the text is calculated.**

**Step 5: Sentiment Results**

## VI. IMPLEMENTATION DETAILS

The model has been built and trained using the CNN-LSTM algorithm [11]. By using Fine-grain sentiment analysis study will focus on the following models. Python is an interpreter, high-level, and object-oriented scripting language. It is highly readable and has fewer syntactical constructions than other languages. Python programming is used in the development of the model. In this research work, the following python libraries will be used to develop the machine learning models:

a. **TextBlob** is a popular Python library for processing textual data. It is built on top of NLTK. TextBlob uses a sentiment lexicon (consisting of predefined words) to assign scores for each word, which are then averaged out using a weighted average to give an overall sentence sentiment score. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

b. **VADER** is quite successful when dealing with social media texts, NY Times editorials, movie reviews, and product reviews. This is because VADER not only tells **about** the Positivity and Negativity score but also tells us about **how positive or negative a sentiment is**.

c. **FastText** is a CPU-based library for text representation and classification. FastText considers subwords using a collection of n-grams: for example, "train" is broken down into "tra", "rai", and "ain". In this manner, the representation of a word is more resistant to misspellings and minor spelling variations.

d. **Flair** is a contextualized representation called string embeddings. To obtain them, sentences from a large corpus are broken down into character sequences to pre-train a bidirectional language model that "learns" embeddings at the character-level. This way, the model learns to disambiguate case-sensitive characters (for example, proper nouns from similar-sounding common nouns) and other syntactic patterns in natural language, which makes it very powerful for tasks like named entity recognition and part-of-speech tagging.

e. **Gensim** It is used to extract semantic topics from documents. It is designed to process unstructured and raw textual data. Gensim has an algorithm such as Word2Vec. The Word2Vec automatically discovers the semantic structure of phrases by examining statistical co-occurrence patterns within a corpus of training documents. These algorithms are unsupervised. It first finds the statistical pattern on any plain text document. The documents can be expressed in the new, semantic representation. It queried for topical similarity against other documents.

## VII. DEEP LEARNING APPROACH



It is a class of machine learning methods and based on artificial intelligence. It uses multiple layers to extract high features progressively. The arrival of deep learning techniques has opened new doors for fresh possibilities and horizons. As indicated by Alharbi and Elise [1], deep learning has shown promising outcomes in processing tasks of natural language incorporating sentiment analysis. Deep understanding requires deep neural networks in learning multifaceted features. These features have been extracted from the assistance of minimum external contributions [7]. The deep learning approach required a large dataset to get a significant boost in the performance of a model. Le, BAC, and Nguyen utilized deep learning techniques to classify sentiments in the review of movies [5]. The researchers have utilized two different datasets that have binary and multiclass labels. The methods of deep learning used recursive neural networks and word2vec both. Such tools were used to create feature vectors in support of classifiers. The aim of the study is at extracting semantic characteristics and of the texts that matched to other previous studies. Word2vec library is used for creating vector representations in words in higher dimensions. So, the authors eased the extraction of deep semantic associations in terms.

## IX. CONCLUSION

In this paper, we performed sentiment analysis of news channel data collected from different sources which are specified during the lockdown period due to the COVID-19 pandemic. Because of this pandemic issue, many businesses are closing down, economic crises are leading and suicides are happening among the general public. As people are dependent on multimedia devices for spending their time in the lockdown period, this can lead to psychological effects such as loneliness and depression. So, there is a need to analyze the psychology of the human mind during the COVID-19 pandemic situation based on the sentiment analysis conducted.

## X. FUTURE WORK

In the future, we would expand the study and in detailed view should be performed. The performance of emotion detection such as anger, fear, joy, anxiety, sadness, panic, etc should be done using the global news channel dataset. So an integrated analysis should be performed. This conduction should be for an association of sentiments to a psychological profile, demographic characteristics, events, transactions, etc.

## XI. REFERENCES

1. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res.* 2020;22(4):e19016. doi: 10.2196/19016. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
2. Barkur G, Vibha GBK (2020) Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India. *Asian J Psychiatr* [[PMC free article](#)] [[PubMed](#)]
3. Block P, Hoffman M, Raabe IJ, Dowd JB, Rahal C, Kashyap R, Mills MC (2020) Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nat Hum Behav*:1–9 [[PubMed](#)]
4. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha FA, Röst G. Risk assessment of novel coronavirus COVID-19 outbreaks outside China. *J Clin Med.* 2020;9(2):571. doi: 10.3390/jcm9020571. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
5. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, Zheng J (2020) The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Res* 112934 [[PMC free article](#)] [[PubMed](#)]
6. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta bio-medica: Atenei Parmensis.* 2020;91(1):157–160. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
7. Dansana D, Kumar R, Das Adhikari J, Mohapatra M, Sharma R, Priyadarshini I, Le DN. Forecasting confirmed and fatalities cases of COVID-19 outbreak. *Front Public Health.* 2020;8:592. doi: 10.3389/fpubh.2020.580327. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
8. de Las Heras-Pedrosa C, Sánchez-Núñez P, Peláez JI. Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. *Int J Environ Res Public Health.* 2020;17(15):5542. doi: 10.3390/ijerph17155542. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
9. Dhar BK, Ayithey FK, Sarkar SM. Impact of COVID-19 on psychology among the university students. *Global Chall.* 2020;4(11):38–44. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
10. Dsouza DD, Quadros S, Hyderabadwala ZJ, Mamun MA (2020) Aggregated COVID-19 suicide incidences in India: fear of COVID-19 infection is the prominent causative factor. *Psychiatry Res*:113145 [[PMC free article](#)] [[PubMed](#)]
11. Dubey, Akash Dutt, Twitter Sentiment Analysis during COVID-19 Outbreak (April 9, 2020). Available at SSRN: <https://ssrn.com/abstract=3572023> or 10.2139/ssrn.3572023
12. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy, and France. *Chaos, Solitons Fractals.* 2020;134:109761. doi: 10.1016/j.chaos.2020.109761. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
13. Gautret P, Lagier JC, Parola P, Meddeb L, Sevestre J, Mailhe M, Hocquart M (2020) Clinical and microbiological effect of a combination of hydroxychloroquine and azithromycin in 80 COVID-19 patients with at least a six-day follow up: a pilot