Implementation of Therapy Bot for Potential Users With Depression During Covid-19 Using Sentiment Analysis

Mr. Muneshwara M S[1], Mrs. Swetha M S[2], Ms. Poorvi Rohidekar[3] ,Mr. Pranove A B[4]

[1]Assistant Professor, Dept. of CSE , BMS Institute Of Technology and Management, Yelahanka, Bengaluru-560064, Karnataka, India.
[2]Assistant Professor, , Dept. of ISE ,BMS Institute Of Technology and Management, Yelahanka, Bengaluru-560064, Karnataka, India.
[3,4]Student, BMS Institute of Technology and Management, , Yelahanka, Bengaluru-560064, Karnataka, India.

E.Mail :[1]muneshwarams@bmsit.in, [2]swethams_ise2014@bmsit.in, [3]prohidekar8@gmail.com, [4]pranove.ab@gmail.com

Abstract:
In these past few years, there has been a huge growth in the use of microblogging and social media updating. Prodded by that development, organizations are looking for approaches to mine data about the thing individuals are feeling. Our undertaking is to fabricate a compelling framework to mine the estimations/psychological wellness of individuals during Coronavirus pandemic. Our project is to build an effective system to mine the sentiments/ mental health of people during COVID-19 pandemic. It is to effectively analyze, extract, and train a vast amount of data. The results will classify the customer's perspective into negative and positive sentiments. To help us classify the texts in an easier way, hashtags will be an added advantage. By collecting valuable information, we apply the machine learning algorithm BERT to train our model to recognize the sentiments to provide the best accuracy. confidentiality. Main objective is to help a user understand his/ her mental state as much as possible and to self-regulate. To provide end-to-end encryption of user bot interaction since it is sensitive information.

Keywords: Chatbot, COVID-19, Mental Health, Sentiment Analysis.

1. INTRODUCTION

Sentiment analysis refers to the broad area of natural language processing which deals with the computational study of opinions, sentiments and emotions expressed in text. Sentiment analysis is known as opinion mining. It is extremely useful in business fields, social media etc. Using sentiment analysis, the polarity of opinions can be found, such as positive, negative, or neutral by analysing the text of the opinion. The online medium has become a significant way for people to express their opinions and feelings.

While there has been a fair amount of research on sentiments expressed in reviews, new articles, our main aim in this project is to analyse the sentiments of users during the COVID-19 pandemic. The outbreak of coronavirus disease has upended the lives of people worldwide. By mining the data from these social networking sites such as Twitter, Reddit, Facebook using their public API (Application programming interface) we will classify the texts and posts by users based to check if the user is feeling a positive or negative emotion by using various machine learning algorithms. The tweets and posts reflect the people's psychological well-being and spread of negative sentiments may lead to social disruption. This analysis provides insightful information in making

appropriate health responses of the public. A chat-bot powered by NLP can support analysis of sentiment to identify a person's emotion. To develop chat bots that can analyse the sentiments of a person in real time and that can mine answers to questions from given predefined responses. The main purpose is to relieve the user from generalized anxiety, depression, and other mental health conditions. A tree structure is created for the responses for the related questions according to the result obtained after applying sentiment analysis. The NLP algorithm sentiment analysis handles flow of conversation by recognizing keywords and terms.

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. It can be used to focus on the customer feedback verbatims where the sentiment is strongly negative that a person is suffering from depression. Likewise, we can look at positive tweets where a person is not. Only after this analysis has been conducted successfully, we can focus on developing a bot for potential users. To effectively ensure faster onboarding, cost reduced, self-therapized, automated bot. To train them once so that they will communicate with our target audience.

Literature Review:

In [1] as aforementioned, the rise of online social networks provides unprecedented opportunities for solving problems in a wide variety of fields with information techniques. For example, traditional psychology research is based on questionnaires. An association model is established between features abstracted from Micro-blog systems and depression inclination. The model also determines the principal features which affect depression detection significantly.

In [2] We see that CHATBOTS are automated systems which are replicating the user's behavior on one side of the chatting communication. They are mimic systems which imitate the conversations between two individuals. They provide a stimulating platform for effective and smart communications with the user on the other end. They copy marketers, sales persons, counsellors and other mediators and work to provide services that the above mentioned people provide. The chatbots used by the telecom and marketing sectors for customer service are scripted types of chatbots. They help the customers on some predefined customer care questions. Research is being carried out in making the conventional monotonous chatbots to be communicative, responsive and carry out the communication in a natural (conversational) language. Selection of an appropriate method is based on the domain of the chatbot, the functionalities it intends to provide, the language of communication, the end user, etc.

In [3] we see that an unprecedented situation such as the COVID-19 outbreak makes it important to rapidly define the zeitgeist as new issues arise in a difficult time. " The 5 most common themes were health care environment, business economy, emotional support, social change, and psychological stress, which represent the biggest concerns for the public. The sentiments can clarify the public response to COVID-19 and help guide government officials, private entities, and the public with information as they navigate the pandemic.

In [4] a chatbot system is built that is able to converse and interact with human users using spoken, written, and visual languages. Chatbots have potential to increase access to mental health interventions. In particular, chatbots may encourage interaction by those who have traditionally been reluctant to seek mental health advice due to stigmatization. This paper shows that the majority of chatbots (92.5%) depend on only decision trees to generate their responses; only 7.5% use machine learning approaches. This may

indicate that chatbots in mental health lag behind chatbots in other fields (e.g. customer services) where artificial intelligence chatbots are more common. The extensive use of rule-based approaches in the studies have to be attributed to the fact that they are more appropriate for chatbots that perform simple, straightforward and well-structured task. In contrast to artificial intelligence chatbots, users of rule-based chatbots cannot usually control the dialogue because their inputs are restricted to predefined words and phrases.

In [5] we see that sentiment analysis is replacing traditional and web based surveys conducted by companies for finding public opinion about entities like products and services. Feature based sentiment analysis included feature extraction, sentiment prediction, sentiment classification and optional summarization modules. Feature extraction identifies those product aspects which were being commented by customers, sentiment prediction identifies the text containing sentiment or opinion by deciding

sentiment polarity as positive, negative or neutral and finally the summarization module aggregates the results obtained from previous two steps.

2. PROPOSED METHODOLOGY

The task of sentiment analysis, especially in the domain of micro-blogging, is still in the developing stage and far from complete. So, to improve the current analysis and to bridge the gap between the existing system, we take our analysis into a deeper level by analyzing those suffering from depression, and other mental health issues during the COVID-19 pandemic.

Issues may occur from loneliness, loss of job, stress around the disease itself etc. Rather than just analyzing the positivity and negativity of generic tweets, posts by users we will make use of the hashtag system (#depression, #mentalhealth) to arrive at the best possible result. To then apply our analysis into building a bot to help users to interact and elevate their spirits. The main purpose is to relieve the user from generalized anxiety, depression, and other mental health conditions.
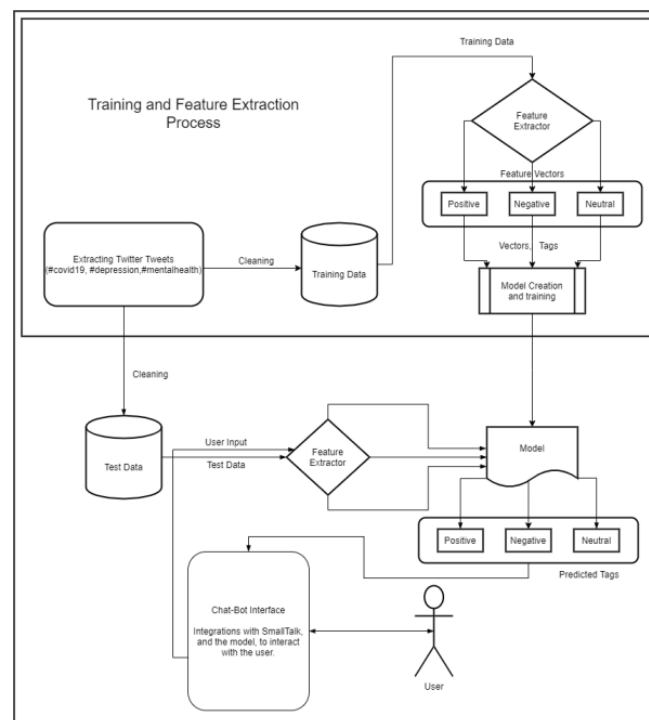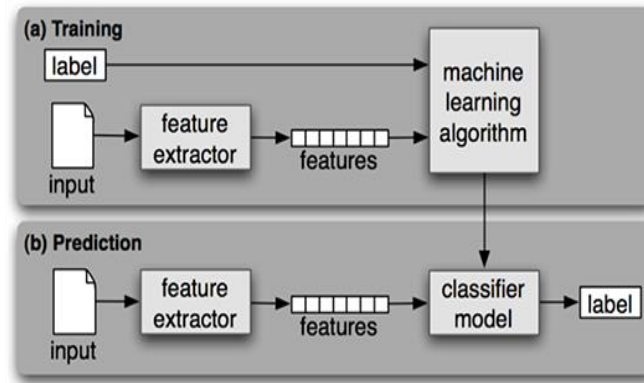

Figure 1. Behavioral Design

Training and Prediction Process:

In the training process (a), our model learns to associate a particular input (i.e., a text) to the corresponding output (tag) based on the test samples used for training. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g., positive, negative, or neutral) are fed into the machine learning algorithm to generate a model.

In the prediction process (b), the feature extractor is used to transform unseen text inputs into feature vectors. These feature vectors are then fed into the model, which generates predicted tags (again, positive, negative, or neutral).

Feature Extraction from Text:

The first step in a machine learning text classifier is to transform the text extraction or text vectorization.

More recently, new feature extraction techniques have been applied based on word embeddings (also known as word vectors). This kind of representation makes it possible for words with similar meaning to have a similar representation, which can improve the performance of classifiers

Algorithm:

A new technique for NLP pre-training called Bidirectional Encoder Representations from Transformers, or BERT. With this release, anyone in the world can train their own state-of-the-art question answering system (or a variety of other models) in about 30 minutes on a single Cloud TPU, or in a few hours using a single GPU.

BERT builds upon recent work in pre-training contextual representation including semi-supervised sequence learning, generative pre-training. However, unlike these previous models, BERT is the first deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus.

System Design:

System architecture is a conceptual model that defines the structure, behavior and more views of a system. A system architecture can comprise system components, the expanded system developed, that will work together with the overall system.

Data Acquisition:

The data creation is done using Twint, which is an advanced Twitter scraping tool written in Python that allows for scraping Tweets from Twitter profiles without using Twitter's API. Twint utilizes Twitter's search operators to let you scrape Tweets from specific users, scrape Tweets relating to certain topics, hashtags & trends, or sort out sensitive information from Tweets like email and phone numbers. The data is scraped with the hashtag "covid19 depression sad" and in the csv format with a large set of tweets.

Pre-Processing Data:

Before utilizing the data to train the model we have to clean the data and remove unwanted texts such as removing special characters , links, images, videos, emoticons, numbers , capital letters. During this we also make the dataset ready for training and testing by categorizing and labelling them into positive tweets(1) and negative tweets (0), called vector features.

Figure 2. Word Cloud

Building and Fine-tuning The Model
We use vector tags and train the model for multiple algorithms to determine the best possible classification algorithms for

sentiment analysis. We use this analysis to provide accuracy, precision and comparison among them.

Figure 3. Algorithm Flow

Developing API's For the Model
Once the training and testing of the data has been done the best possible classifier is chosen and we build an API to present the results of any given tweet classification. Any text given as a parameter to this API will be analyzed by the trained model and sentiment will be classified as negative or positive. This API is built in Python on Flask.

Developing The User Interface We will design a UI for the target user to interact with the model through the API in the backend. Here the user can type in text to be analyzed by the model, and based on the sentiment the chat-bot will reply to the user appropriately.
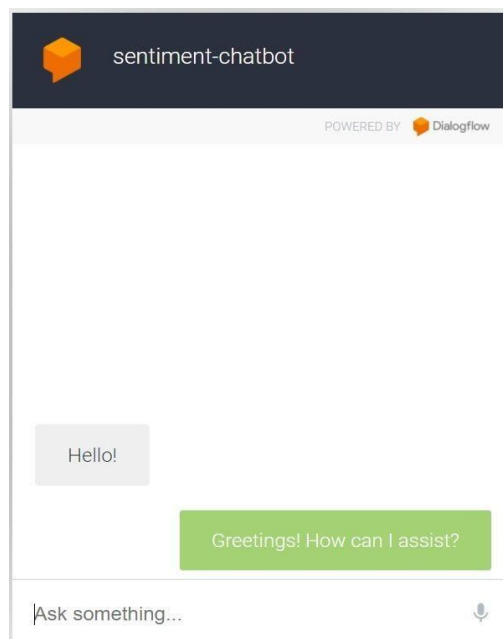
Figure 4. Chat-Bot UI
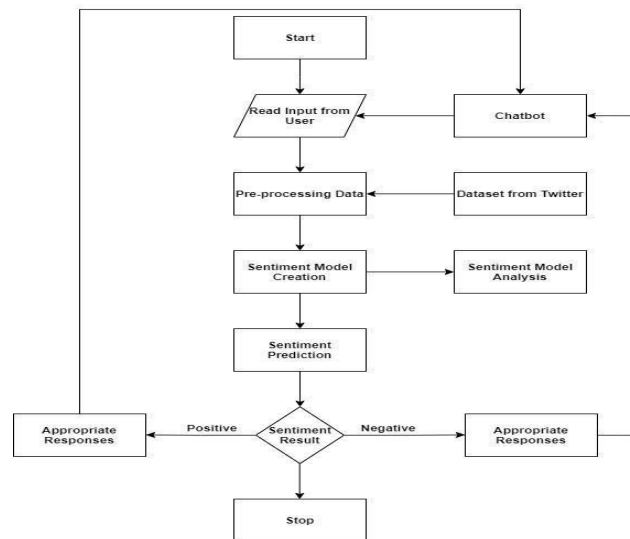
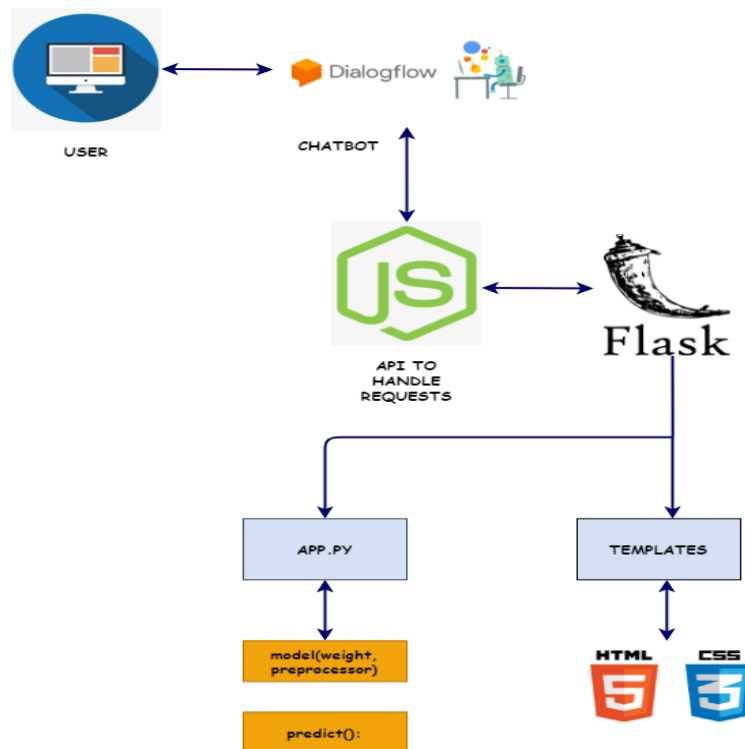**Figure 5. Flow Diagram**
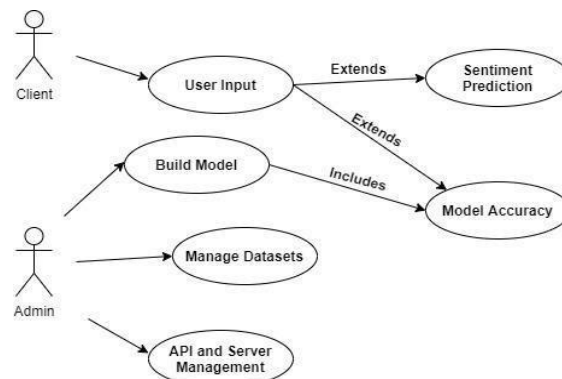
Figure 6. Use Case Diagram

Figure 7. Component Diagram

3. RESULTS
Dataset Showing Tweets, Subjectivity, Polarity, and Analysis

	tweets	Subjectivity	Polarity	Analysis
56054	getting a puppy or dog during the quarantine? ...	0.611111	0.477778	Positive
35610	10 launched fearnotthereisgod twitter campaign...	0.266667	-0.025000	Negative
47391	we met with our clients to discuss their fears...	0.625000	0.083333	Positive
39311	with funding from nsws is now able to offer up...	0.485000	0.260000	Positive
56154	if normals panic its cuz of the pandemic; if m...	1.000000	-0.500000	Negative
...	...	...	...	...
39520	anxiety over coronavirus can manifest in myria...	0.200000	-0.100000	Negative
32260	a thread about my anxiety on reopening back in...	0.450000	0.200000	Positive
16404	of course it's high the state was shut down. t...	0.476296	0.001481	Positive
955	take the positive step to make a change now. t...	0.415152	0.209091	Positive
71648	respected madame looking at crazy volatility a...	0.390972	-0.220139	Negative

Figure 8. Data Frame

Model Evaluation:

```
learner.evaluate()

              precision    recall   f1-score    support

          0        0.95      0.96       0.96       4978
          1        0.96      0.95       0.96       5022

   accuracy                             0.96      10000
  macro avg        0.96      0.96       0.96      10000
weighted avg       0.96      0.96       0.96      10000

array([[4771,  207],
       [ 227, 4795]])
```

Figure 9. Model Evaluation
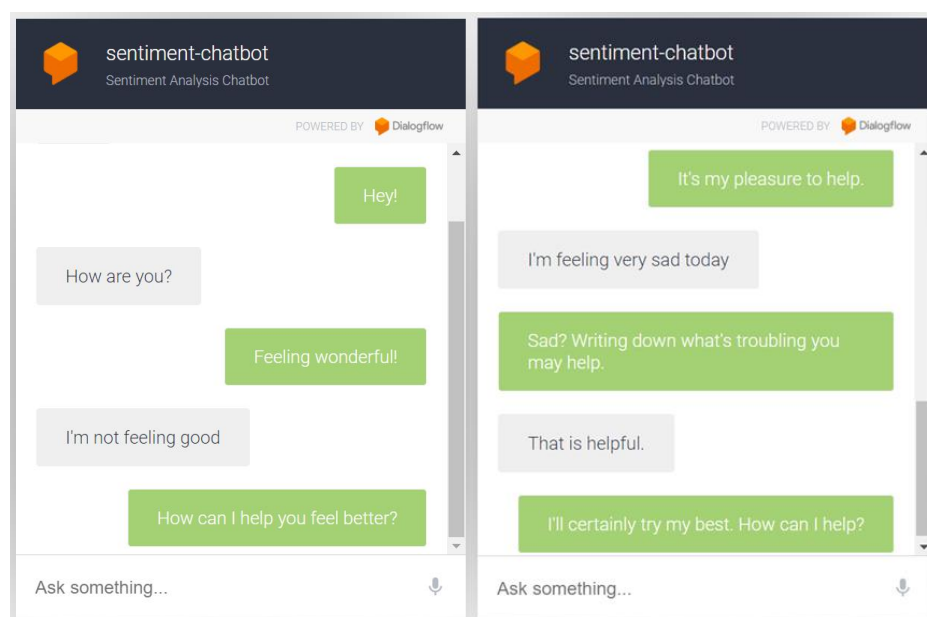
Chatbot UI :

Figure 10. ChatBot UI

4. CONCLUSION

After the extensive analysis of the objective, This project works to expand the scope of social media-based mental health measures and use existing research that has proven the correlation between depression and specific linguistic features in order to build an algorithm i.e BERT classifier that can predict text-based signs of depression. By analysing linguistic markers in social media posts, that can give an individual insight into his or her mental health far earlier than traditional approaches. Building an algorithm that can analyse Tweets exhibiting self-assessed depressive features can make it possible for individuals, parents, caregivers, and medical professionals to analyse social media posts for linguistic clues that signal deteriorating mental health far before traditional approaches currently do and once the sentiment has been analysed it can be passed through the flask API to make appropriate predictions and revert back with a response. The future scope of this bot is that it can be integrated with other messaging apps like Slack, Facebook Messenger, etc so as to recognise and diagnose mental health problems of users on a day to day basis, based on the context of their chat.

5. REFERENCES

[1] Wang, Xinyu & Zhang, Chunhong & Ji, Yang & Sun, Li & Wu, Leijia & Bao, Zhana. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. Pacific Asia Conference on knowledge discovery and data mining, **(2013)**.

[2] Nivedita Bhirud, Subhash Tataale, Sayali Randive, Shubham Nahar A Literature Review On Chat-bots In Healthcare Domain. International Journal of Scientific & Technology Research Volume 8, Issue 07, **(2019)** .

[3] Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, Hon SD, Park J, Dang P, Lipsky MS. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. J Med Internet Res. **(2020)**.

[4] Muhammad Zubair Asghar1 , Aurangzeb Khan2 , Shakeel Ahmad1 , Fazal Masud Kundi1. Review of Feature Extraction in Sentiment Analysis. Journal of basic and applied research, **(2014)**.

[5] Kaliyar, Rohit Kumar. "A Multi-layer Bidirectional Transformer Encoder for Pre-trained Word Embedding: A Survey of BERT." 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), **(2020)**.

[6] Agarwal, Apoorv, Fadi Biadsy and K. McKeown. "Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-Grams." Proceedings of the 12th Conference of the European Chapter of the ACL, pages 24–32, Athens, Greece, 30 Mar – 3 April **(2009)**.

[7] Pang, Bo & Lee, Lillian. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Computing Research Repository - **(2004)**.

[8] S. Liu, H. Tao and S. Feng, "Text Classification Research Based on Bert Model and Bayesian Network," 2019 Chinese Automation Congress (CAC), **(2019)**.

[9] Navin Sabharwal, Amit Agrawal. "Hands-on Question Answering Systems with BERT", Springer Science and Business Media LLC, **(2021)**.

[10] Yuksel Pelin Kilic, Duygu Dinc, Pinar Karagoz. "Named Entity Recognition on Morphologically Rich Language: Exploring the Performance of BERT with varying

Training Levels", 2020 IEEE International Conference on Big Data (Big Data), **(2020)**.

[11] H. Nandigramwar, A. Mittal, A. Bhatnagar and M. Rashid, "A distributed and unified API service for machine learning models," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), **(2021)**.

[12] P. Bhattacharyya, "Sentiment Analysis," 2013 1st International Conference on Emerging Trends and Applications in Computer Science, **(2013)**.

[13] Brett Duncan, Yanqing Zhang. "Neural networks for sentiment analysis on Twitter", 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), **(2015)**.

[14] Habibe Karayiğit, Çiğdem İnan Acı, Ali Akdağlı. "Detecting abusive Instagram comments in Turkish using convolutional Neural network and machine learning methods", Expert Systems with Applications, **(2020)**.

[15] Dubey, Akash. Twitter Sentiment Analysis during COVID19 Outbreak. SSRN Electronic Journal, **(2020)**.

[16] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), **(2013)**.

[17] H. P. Patil and M. Atique, "Sentiment Analysis for Social Media: A Survey," 2015 2nd International Conference on Information Science and Security (ICISS), **(2015)** .

[18] ThippaReddy Gadekallu, Akshat Soni, Deeptanu Sarkar, Lakshmanna Kuruva. "chapter 6 Application of Sentiment Analysis in Movie reviews", IGI Global, **(2019)** .

[19] Tianyi Wang, Ke Lu, Kam Pui Chow, Qing Zhu."COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model", IEEE Access, **(2020)** .

[20] Manju Venugopalan, Deepa Gupta. "Exploring sentiment analysis on twitter data", 2015 Eighth International Conference on Contemporary Computing (IC3), **(2015)** .