iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33962

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

Social Media Analysis for Mental Health Evaluation

Benita Rego[1], Nolita Rego[2], Mohit Kunder[3]
[1, 2, 3]*Department of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, University of Mumbai*

Abstract: ***The system provides a background on depression, use of social media platforms for prediction and analysis using machine learning and deep learning algorithms. The system monitors the social media activities of each person and predicts their mental health factors such as depression, anxiety, stress, etc. This system also uses real time online social media data by exploring the parallelism between the users' mental health and the content they post on their social media handle. This system could be used by mentors such as teachers, doctors, etc to acquire a weekly analysis of the person's stress levels, thereby helping in providing consultation accordingly.***

Keywords: Social computing, healthcare, artificial intelligence, deep learning, data analysis, natural language processing, convolutional neural network

I. INTRODUCTION

Social Media is universally popular among the current generation, making it effortless for people to share their lives on a platform where everyone can view it, also observing their emotional and mental mindset through posts and tweets on their social media handle. Although it is easier to reach out to people, it is difficult to accurately analyse a person's state of mind. But, it may be possible to detect their mental health using their social network data. For this, real-world social media data should be analysed. Along with detection, the person will also require appropriate medical assistance to help him get through their declining mental state. Social media platforms have been an indispensable part of the people where most of them express their opinions, point of views and expressions in the form of tweets, posts, comments, etc. There have been perceptions of users expressing their feelings, especially the negative ones such as depression, anxiety, mental health issues through cryptic tweets or posts. The close relation between the contents posted on such platforms and their mental state can be used to identify psychological health issues like anxiety, depression, stress, etc which is common among people nowadays [3]. There are various ways in detecting such issues with the help of machine learning and deep learning along with the social media API being the backbone of the identification. The identification of mental state of mind i.e detecting the mental health of the user using social media with the help of web mining and emotional analysis is one such way towards evaluation and mental health awareness [2]. This evaluation can be proven helpful to people under tremendous negative mental state, further aiding to overcome them. Convolutional neural network have proven to be the best algorithm having one of the best practical efficiency. Mental illnesses, such as depression, are highly prevalent in the current generation which has shown major impacts on a person's physical health. Lately, artificial intelligence (AI) methods have been widely established to serve mental health contributors which includes doctors, counsellors and psychiatrists, for making decisions based on persons' historical data (e.g., medical records, behavioural data, social media usage, etc.). One of the most recent generations of AI technologies, Deep learning has exhibited higher-level production in many real-world applications that range from various visual perceptions to medical care management.

II. LITERATURE REVIEW

Nowadays social media is constantly used by people all over the world to not only interact with other people but to also share certain aspects of your life. The people on the various social media platforms engage with each other in more than just conversations. Although social media is an extremely useful tool to connect friends who live across the world as well as in helping people meet someone with common interests, opinions, etc they wouldn't meet under normal circumstances, it's overuse can be really harmful.

The COVID-19 pandemic has struck over 200 countries and has caused global panic and hence a decline in people's mental health and mental stress [5]. Also, in this age of social media, several people are getting overloaded with rumors and misinformation regarding the virus which is unverified and false [6]. Such news also created unnecessary panic, fear and anxiety amongst the general populace of India.

Twitter is a free social media platform that's widely used across the world. The tweets that the users posts are public unless the user specifically marks them private. The users can also label their tweets with the help of a hash tag in order to categorize it. Therefore, due to the fact that the tweets are so public in nature, Twitter can be really valuable to have information about what people think about various topics or subjects as well as analyze a user's state of mind and their mental state and is a valuable source of information for working towards treatment and prevention of mental health related issues [7].

According to a study [8] there are eight types of gratifications for a person to use social media. These motives include self-documentation, information sharing, social interaction, entertainment, passing time, self-expression, medium appeal and convenience. Based on these the users that are motivated by self-documentation, self-expression and social interaction are the ones who are most likely to use social media in order to talk about themselves and their mental health.

Users and social media are related in various ways like social media anxiety disorder, social anxiety and anxious depression social media verbalization [1]. In order to detect them, these disorders are evaluated using a deep learning model which will help in analyzing the mental health condition of the user with respect to their social media contents. One of the papers referred to is by Wajdhi Zanghouani which revolves around detection of depression among the Youth of Middle East and North Africa (MENA) region using social media data that aims to create a large scale dataset of users with self-reported messages indicating that they have depression [9]. This research made use of NLP techniques and ML techniques such as Support Vector Machines (SVM). To gather data from people who are tweeting about the required topics Twitter API was used, which also gives relevant information such as location and profile description. Sentiment analysis is done on the tweet based on [10] Ekman's six emotions – joy, sadness, anger, fear, disgust and surprise.

In another paper by Neethu, sentiment analysis was done on a sentence level on the tweets in three steps that included preprocessing, creating a feature vector using relevant features and then classifying the tweets into positive and negative using various classifiers [11]. Positive emotions are given the weight of "1" and negative ones are given the weight of "-1". Twitter specific features such as hash tags, emoticons, etc are removed and then further feature extraction is done to convert the tweet into simple text. Tweets are then represented as a collection of words using a unigram approach. Different classifiers such as Multinomial Naive Bayes, SVM, Random Forest, K-Nearest Neighbours, Maximum Entropy and Ensemble classifiers were tested and found to have similar accuracy.

III. PROBLEM DEFINITION

The model is a user-friendly application that analyzes and detects the mental state of a person through their social media contents and further provides the resources to help the said person find an appropriate assistance. It also gives periodic statistics and analysis in the form of a graphical representation. On detection of negative polarity, it alerts the user if the stress level exceeds a certain limit (threshold) and furthermore, recommends medical assistance. The recommendation system provides the locations of nearby counseling firms when required.

IV. PROPOSED SYSTEM

The entire system is designed in such a way that it can be accessed easily through the mobile application. It is primarily designed for Twitter users where the tweets are used for detection and analysis. This system involves technologies, each playing a crucial role in the development and working of the system. After careful observations and testing of different algorithms, the algorithm used for the system is convolutional neural networks also known as CoVNet because of its high end accuracy as compared to other algorithms listed in the result section of this article. [12] The results obtained from deep learning models had surpassed the other algorithms with close accuracy like naive bayes algorithm**.** CoVNet is ideally used for image analysis, but recently it has been used for many other data categories like texts, numerical, and multimedia. CNN is a multi-layer artificial neural network that consists of multiple layers such as convolutional, pooling and fully connected layers which are the hidden layers of CNN. In a convolution layer, the data is streamed across a filter and features are extracted through Hadamard product. The function of the convolutional layer containing filters for feature maps, is feature extraction [12]**.** The system consists of a convolutional neural network model which is trained with the dataset fed to the network [4]. The dataset is purely the tweets of the user which is fed in real time for accurate analysis. The model is made solely for the user as it is a single party application where the user logs in using their Twitter handle, authenticated by Twitter OAuth 2.0 (Open Authorization). The proposed application can extract public tweets with all the legal authorization by the Twitter Developer system. The model is trained with the live data fetched over 40, 000 twitter data using the Twitter API (Application Programming Interface). The fetched data is cleaned, preprocessed and trained under convolutional neural network where the classification training type is categorical cross entropy**.**

In the study in [13], the author has used word2vec which is a learned model trained on 100 billion words from Google News which has been used in our model as well. Once the model is trained, it is used for analyzing the user's data which is streamed in using live Twitter API where the classification into positive, negative and neutral tweets are done along with evaluation using a graphical format. The user can view the analysis on a weekly and monthly basis. If the negative tweets cross a particular threshold set by the system, the said user will be alerted with a notification in the form of a message on Whatsapp which is integrated using Twilio API. Along with that, the user can make use of the medical assistance recommendation given by the application after the alert which is implemented using Google Maps API (Nearby Places). The mobile application is user friendly, built using Flutter and primarily for the user. The users' data analysis is secured and private. The performance of the application is dependent on the data fetched from the users' Twitter account.

V. DESIGN AND METHODOLOGY

The application is exclusively for the user, specifically Twitter user. The sole requirement of the application is that the user should have a Twitter account with adequate tweets for accurate evaluation. As shown below in Figure 1, the user logs into the application using his/her Twitter handle where further authentication and legal permission for data extraction is provided by the API layer that is, Twitter OAuth 2.0. The data fetched is sent to the application layer where the trained model which is done using convolutional neural network integrated with Flask uses the Python libraries to perform the functions of cleaning and pre-processing using Natural Language Processing (NLP), vectorization and further classification and analysis of the data that is, tweets. The user after logging into his/her account is routed to the dashboard page of the application where the analysis is displayed in graphical format as positive, negative and neutral. The Database layer where the live streamed Twitter data is stored in JSON format in Real-time Firestore is in consistent functioning with the application and API layer.

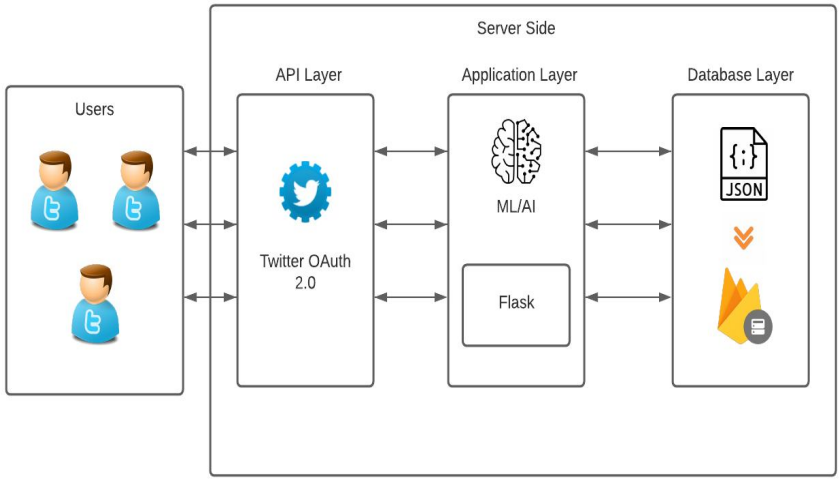


Fig. 1 Architectural Diagram of the complete system

The architectural model is explained as shown below in Figure 2:

1) *Data Exploration:* The data which is fetched from the Twitter API is public and secured and further stored in comma-separated values (csv) format, read using python library, pandas. Since it is fetched in an unstructured format, the data is formatted using the pandas data frame and delimiters.
2) *Data Cleaning and Preprocessing:* The data is cleaned using regular expression (RegEx). Then it is tokenized, converted to lowercase and remove stop words using natural language processing toolkit (NLTK) [1].
3) *Splitting Data into Test and Train* The cleaned data are split into test and train where 90% and 10% are given for training and testing respectively. Then the training and testing vocabulary is built to get maximum training and testing length with the total number of words in the data. Further, we tokenize and pad sequences by assigning an integer to each training and testing words.
4) *CNN:* The text is passed to the convolutional neural network where 5 different filter sizes are put in each text with the type of classification being categorical cross entropy as shown in Figure 3. Then the model is trained with a number of epochs given for the model to loop and learn.

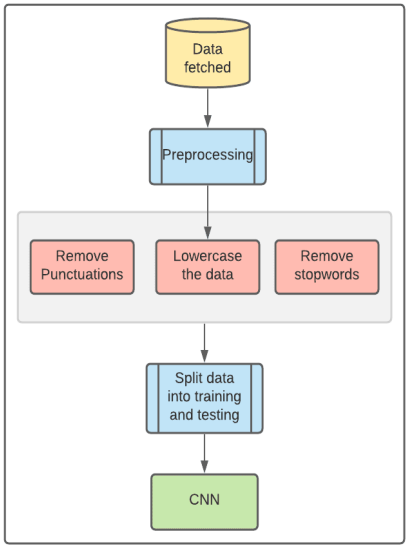


Fig. 2 Architecture Diagram of Tweets Extraction and Classification

```
Model: "model"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #       Connected to
input_1 (InputLayer)         [(None, 50)]              0
_________________________________________________________________________________________________
embedding (Embedding)        (None, 50, 300)           18687600      input_1[0][0]
_________________________________________________________________________________________________
conv1d (Conv1D)              (None, 49, 200)           120200        embedding[0][0]
_________________________________________________________________________________________________
conv1d_1 (Conv1D)            (None, 48, 200)           180200        embedding[0][0]
_________________________________________________________________________________________________
conv1d_2 (Conv1D)            (None, 47, 200)           240200        embedding[0][0]
_________________________________________________________________________________________________
conv1d_3 (Conv1D)            (None, 46, 200)           300200        embedding[0][0]
_________________________________________________________________________________________________
conv1d_4 (Conv1D)            (None, 45, 200)           360200        embedding[0][0]
_________________________________________________________________________________________________
global_max_pooling1d (GlobalMax (None, 200)            0             conv1d[0][0]
_________________________________________________________________________________________________
global_max_pooling1d_1 (GlobalM (None, 200)               0          conv1d_1[0][0]
_________________________________________________________________________________________________
global_max_pooling1d_2 (GlobalM (None, 200)            0             conv1d_2[0][0]
_________________________________________________________________________________________________
global_max_pooling1d_3 (GlobalM (None, 200)            0             conv1d_3[0][0]
_________________________________________________________________________________________________
global_max_pooling1d_4 (GlobalM (None, 200)            0             conv1d_4[0][0]
_________________________________________________________________________________________________
concatenate (Concatenate)       (None, 1000)           0             global_max_pooling1d[0][0]
                                                                     global_max_pooling1d_1[0][0]
                                                                     global_max_pooling1d_2[0][0]
                                                                     global_max_pooling1d_3[0][0]
                                                                     global_max_pooling1d_4[0][0]
_________________________________________________________________________________________________
dropout (Dropout)            (None, 1000)              0             concatenate[0][0]
_________________________________________________________________________________________________
dense (Dense)                (None, 128)               128128        dropout[0][0]
_________________________________________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0             dense[0][0]
_________________________________________________________________________________________________
dense_1 (Dense)              (None, 3)                 387           dropout_1[0][0]
_________________________________________________________________________________________________
Total params: 20,017,115
Trainable params: 1,329,515
Non-trainable params: 18,687,600
_________________________________________________________________________________________________
```

Fig. 3 CNN Model

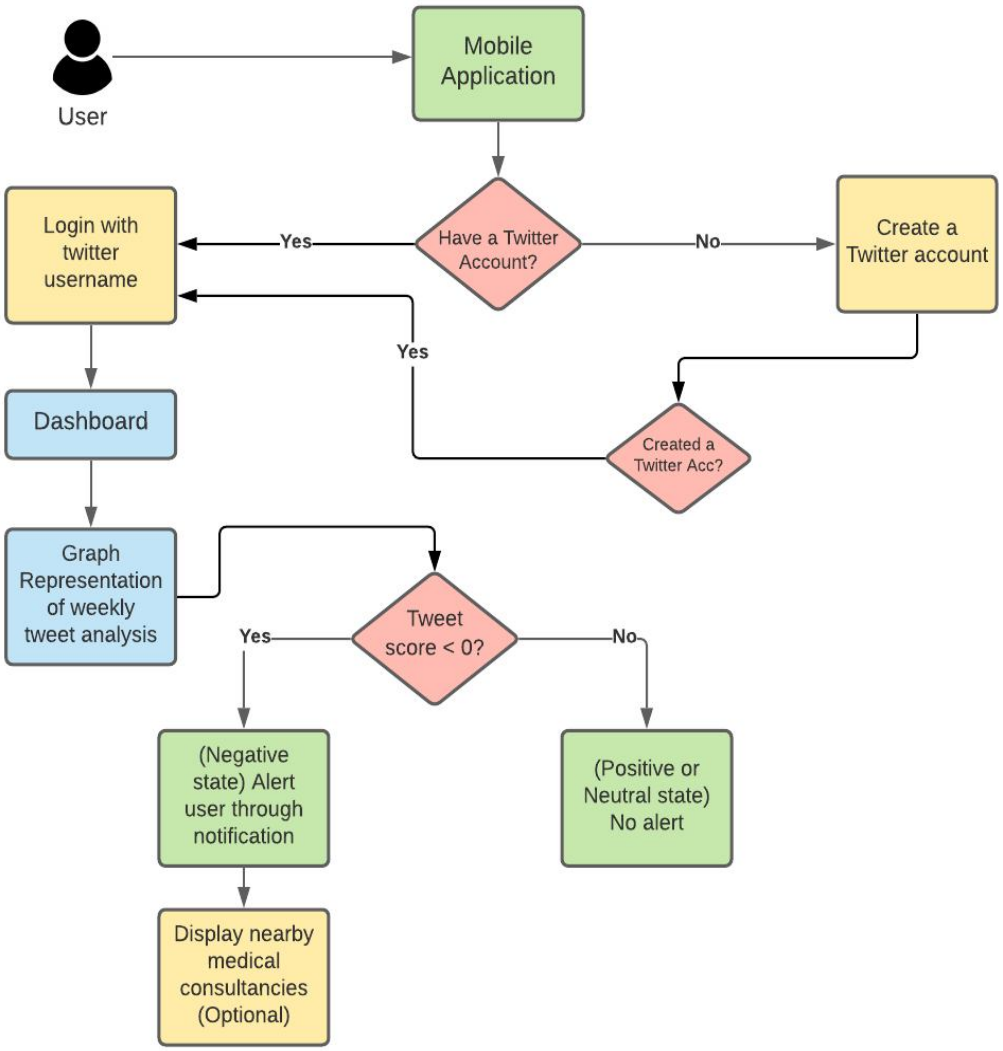


Fig. 4 Flow Chart of the Application

As shown in figure 4, in the frontend, a Flutter-based mobile application will be used by the user. In order to have a proper scaled and distributed architecture, we have split the backend into two, a Flask based server connected to our Real-time Firestore database in order to keep track of data and to handle tweets classification requests powered by our convolutional neural network model. The system works with the Twitter account of the data from where the data is fetched. Using the users' data fetched, the user gets access to the analysis of the tweets from where they can keep track of their data going towards extreme sides like negative or positive and take medical assistance appropriately. Initially the user will have to login into the system and provide it permission to connect to their Twitter account. From there the system will fetch the user's Twitter data like tweets, date and time. Using the trained model the tweets will be assigned a score and classify it to positive, negative or neutral. The score will be displayed in the form of a graphical representation as per the weekly analysis of the tweets of the user. If the negative graph crosses the threshold value set by the application, then the user is alerted via a notification on WhatsApp, further displaying nearby medical consultation using Google Maps which will be optional for the user. The model used in this project for feature extraction is Twitter OAuth 2.0 API and for classification is convolutional neural network (CNN or CoVNet) which is also called space invariant artificial neural networks (SIANN). This model has been used to perform categorical cross entropy classification which classifies into three different factors and further achieves 84.2% test accuracy, but can be improved by training more data. After increasing the data size and further training gave us a test accuracy of 88.16%. This is the best algorithm for our project as compared to other algorithms that gives accuracy less than what CNN gives.

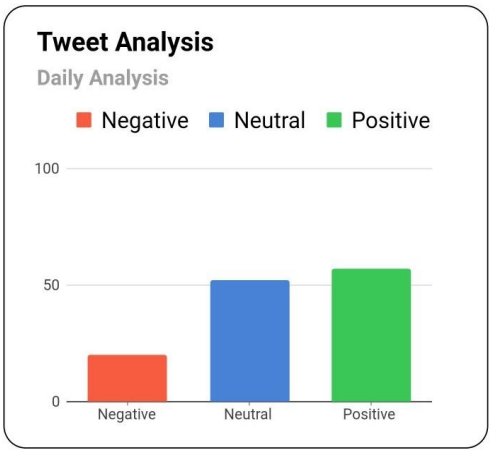


Fig. 5(a) Daily Analysis (Bar Graph)

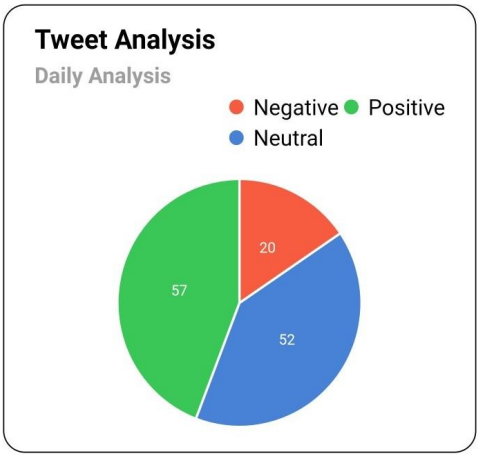


Fig. 5(b) Daily Analysis (Pie Chart)

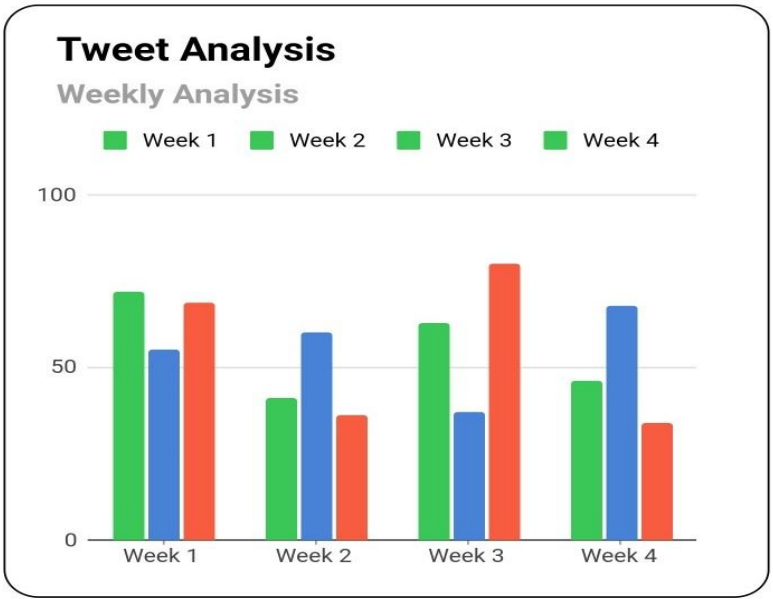


Fig. 5(c) Weekly Analysis

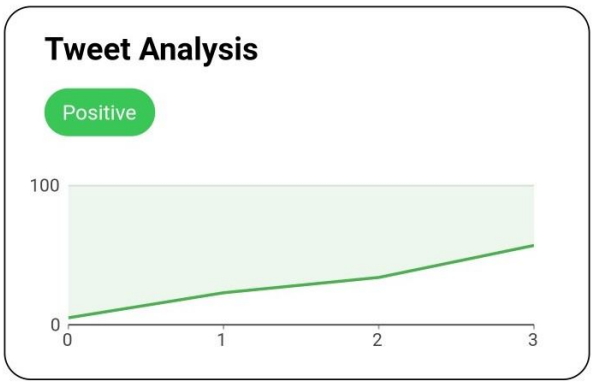


Fig. 5(d) Overall Analysis (Positive Tweets)

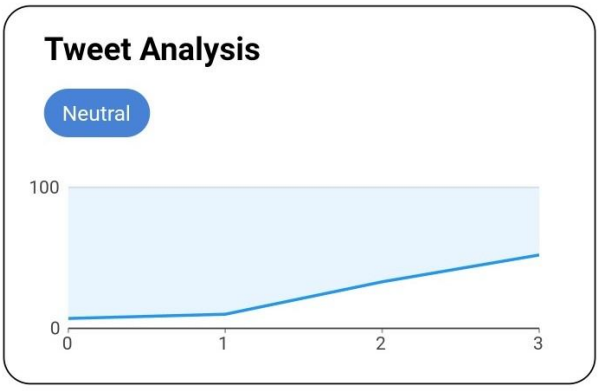


Fig. 5(e) Overall Analysis (Neutral Tweets)

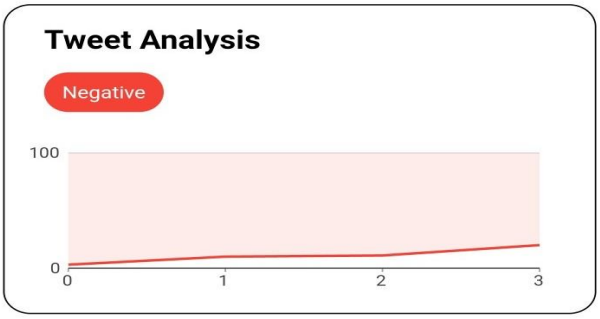


Fig. 5(f) Overall Analysis (Negative Tweets)

VI. METHODS TO IMPROVE ACCURACY

The decision to use this model was taken by comparing it against various algorithms like Multinomial Naive Bayes (MNB), Support Vector Machine (SVM), Random Forest (RF) and K-nearest Neighbours (KNN) models. A lot of planning and analysis has gone into locking down one model for execution and this has increased the clarity about the design and implementation procedure as a whole.

Test Algorithms With Accuracies

Sr no.	Algorithms	Standard accuracy	Test accuracy
1	Convolutional Neural Network (CNN)	87.957%[2]	88.16%
2	Random Forest (RF)	81.04%[1]	81.76%
3	Support Vector Machines (SVM)	79.40%[14]	70.22%
4	Multinomial Naive Bayes (MNB)	77.89%[1]	77.42%
5	K-Nearest Neighbours (KNN)	79.3%[15]	55.55%

Convolutional neural network (CNN) model can maintain its embeddings along with training and testing the CNN data to get the right epochs according to its batch size and providing a greater accuracy compared to other algorithms. The resultant accuracy developed after the experiment is 88.16% which is better than the standard accuracy.

VII. RESULT

We have successfully completed the stages of detection and analysis of social media data wherein the model for our system has provided better accuracy than the standard marked for mental health evaluation through convolutional neural networks. Considering the current situation, the causes for mental balance and stability have become worse and therefore, the analysis of Twitter data for the detection and analysis along with further assistance has been of utmost importance in the growing world where research continues to foster. This system implements a method of tweet extraction that is further used to analyze the state of the person, giving good accuracy results. The system uses the convolutional neural networks model for implementation.

VIII. CONCLUSION

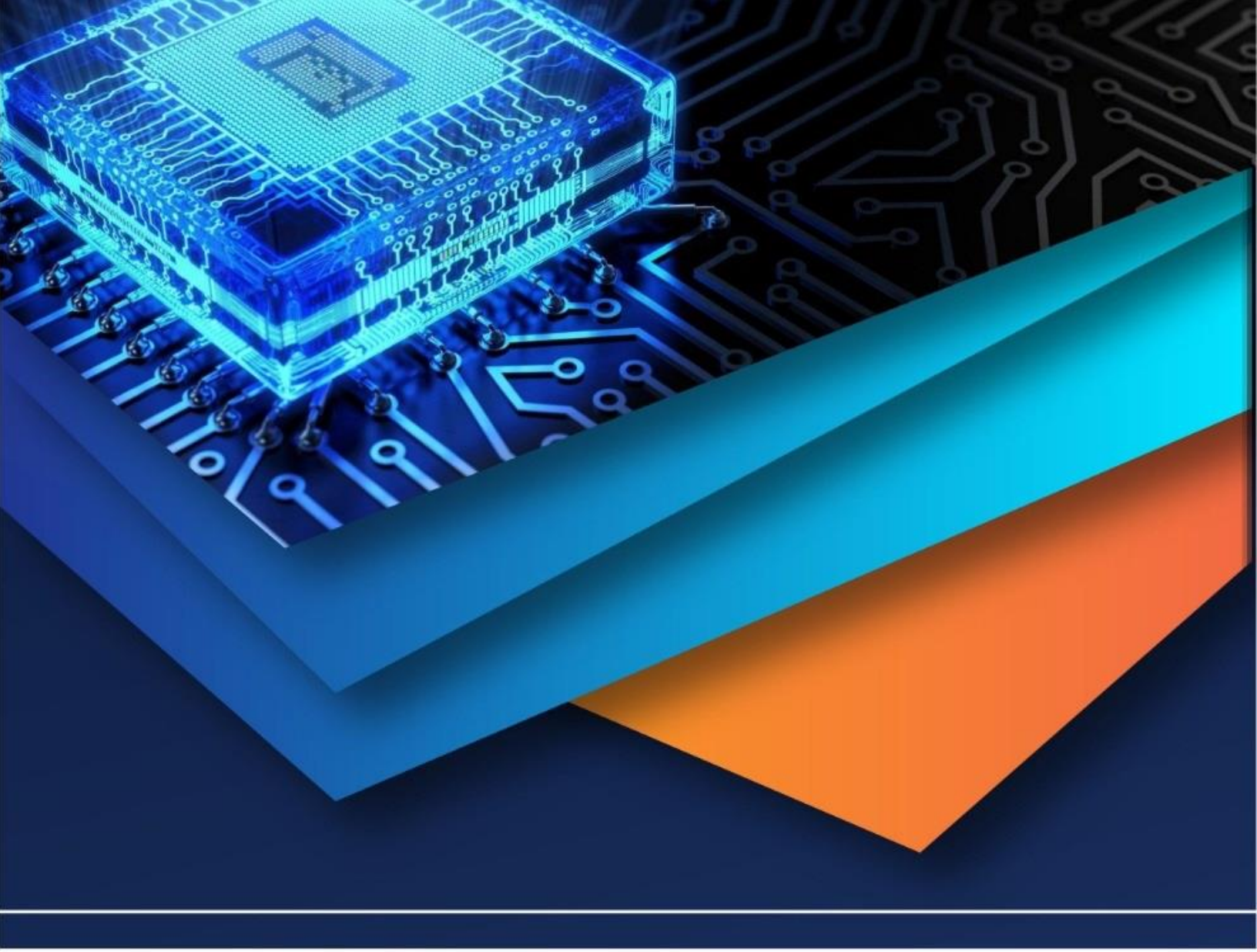
This application successfully gives analysis of social media contents, represented in the form of graphical format and further provides recommendations to nearby medical assistance if the negative value in the graph crosses the threshold set by the system. It is very considerate to people going through mental health issues like stress and depression and it helps them get through it while giving them analysis of what they are posting on social media, keeping in mind how their mindset has been working on it and how they can improve. Furthermore, our application gives an appropriate solution which is essential after detection which would help them ameliorate their mental health.

IX. ACKNOWLEDGMENT

We would like to thank our professor and our project mentor Prof. Sunil Chaudhari for his constant support and guidance throughout the project. We are very grateful to the Head of the Department of Computer Engineering, Dr. B.S. Daga and college principal, Dr. Srija Unnikrishnan for their generosity and co-operation during our project.

REFERENCES

[1] Akshi Kumar, Aditi Sharma, Anshika Arora, Anxious Depression Prediction in Real-time Social Data, International Conference on Advanced Engineering, Science, Management and Technology . 2019 (ICAESMT19), March 2019
[2] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, Diana Inkpen, Deep Learning for Depression Detection of Twitter Users, University of Ottawa, ONKIN 6N5 Canada, June 2018
[3] Huijie Lin, Jia Jia, Jiezhon Qiu, Yongfeng Zhang, Lexing Xie, Jie Tang, Ling Feng, and Tat-Seng Chua, Detecting Stress Based on Social Interaction in Social Networks, Journal of Latex Class Files, Vol. 13, No. 9, September 2014
[4] Nhan Cach Dang, María N. Moreno-García, Fernando De la Prieta, Sentiment Analysis Based on Deep Learning: A Comparative Study, MDPI, Biotechnology, Intelligent Systems and Educational Technology (BISITE) Research Group, University of Salamanca, 37007 Salamanca, Spain, 14 March 2020
[5] Golechha M, COVID-19, India, lockdown and psychosocial challenges: What next?, International Journal of Social Psychiatry. 2020;66(8):830-832. doi:10.1177/0020764020935922
[6] Anant Kumar, K. Rajasekharan Nayar, COVID 19 and its mental health consequences, Journal of Mental Health, DOI: 10.1080/09638237.2020.1757052
[7] Chandler McClellan, Mir M Ali, Ryan Mutter, Larry Kroutil, Justin Landwehr, Using social media to monitor mental health discussions − evidence from Twitter, Journal of the American Medical Informatics Association, Volume 24, Issue 3, May 2017, Pages 496–502
[8] Liu, Ivy L.B.; Cheung, Christy M.K.; and Lee, Matthew K.O., Understanding Twitter Usage: What Drive People Continue to Tweet, PACIS 2010 Proceedings. Paper 92. http://aisel.aisnet.org/pacis2010/92
[9] Wajdi Zaghouani, A Large-Scale Social Media Corpus for the Detection of Youth Depression (Project Note), Procedia Computer Science, Volume 142, 2018, Pages 347-351, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.10.483.
[10] Paul Ekman, An argument for basic emotions, Cognition and Emotion, 6:3-4, 169-200, DOI: 10.1080/02699939208411068, 1992
[11] 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), Tiruchengode, India, 2013, pp. 1-5, doi: 10.1109/ICCCNT.2013.6726818.
[12] Shuichi Hashida, Keiichi Tamura, Tatsuhiro Sakai, Classifying Tweets using Convolutional Neural Networks with Multi-Channel Distributed Representation, IAENG International Journal of Computer Science, 46:1, IJCS_46_1_08, 1 February 2019
[13] Y. Kim, Convolutional neural networks for sentence classification, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, A. Moschitti, B. Pang, and W. Daelemans, Eds. ACL, 2014, pp. 1746– 1751.
[14] Munir Ahmad, Shabib Aftab, Syed Shah Muhammad, Sarfraz Ahmad, Machine Learning Techniques for Sentiment Analysis: A Review, International Journal of Multidisciplinary Sciences and Engineering, Vol. 8, No. 3, April 2017
[15] Abhilasha Tyagi, Naresh Sharma, Sentiments Analysis of Twitter Data using K- Nearest Neighbour Classifier, International Journal of Engineering Science and Computing, April 2018



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY