Early Detection of Depression from Social Media Data Using Machine Learning Algorithms

[1]Thejaswini V, [2]Yashavanth T R

[1]Student, [2]Assistant Professor,
Department of Computer Science,
Maharaja Institute of Technology
Mysore, India.

***Abstract*- Depression has become a serious problem in this current generation and the number of people affected by depression is increasing day by day. However, some of them manage to acknowledge that they are facing depression while some of them do not know it. On the other hand, the vast progress of social media is becoming their "diary" to share their state of mind. Several kinds of research had been conducted to detect depression through the user post on social media using machine learning algorithms. Through the data available on social media, the researcher can able to know whether the users are facing depression or not. Machine learning algorithm enables to classify the data into correct groups and identify the depressive and non-depressive data. The proposed research work aims to detect the depression of the user by their data, which is shared on social media. The Twitter data is then fed into two different types of classifiers, which are Naïve Bayes, SVM, Random Forest and KNN. The results will be compared based on the highest accuracy value to determine the best algorithm to detect depression. The results show both algorithms perform equally by proving same accuracy level. Finally select the best algorithm for deployment to build an application.**

***Index Terms*- Depression, Machine Learning, Sentimental Analysis**

I. INTRODUCTION

Social media and downheartedness can be linked in a variety of ways. Although social media sites can be useful, some people may experience depression that is exacerbated or caused by social media. The relationship between depression and social media can be influenced in the following ways. People's lives are frequently presented on social media in a carefully managed way that emphasizes the good parts of their lives while omitting the bad. This may cause people to socially compare their own lives to those that are portrayed on social media, thinking that their own lives are less desirable. Constantly evaluating oneself against others can lead to low self-esteem and unhappiness. Social media might make people afraid of missing out on activities, get-togethers, or experiences that other people seem to be enjoying. When a person is unwilling to engage in similar activities, their fear might cause anxiety and a sense of isolation. A person's mental health may suffer greatly if they are subjected to unfavourable remarks, cyberbullying, or internet abuse. Because of the anonymity and distance that social media provide, some people may feel empowered to act cruelly, which makes the victims feel depressed, unhappy, and alone.

When depression strikes younger children, it is more likely to take the form of absenteeism from school, separation anxiety, and parent-related anxiety. Teenagers with depression are often agitated, moody, and disruptive at school. Co-morbid anxiety, eating disorders, or drug misuse is also common in them. Older adults may experience depression more discreetly because they are less inclined to acknowledge to experiencing melancholy or grief, and because medical conditions that are more prevalent in this group of people either contribute to or are the cause of depression.

A person's daily life may be impacted by major mental health issues including anxiety and depression. Despite some overlaps in the illnesses' etiology, signs, and therapies, there are also significant distinctions between them. In contrast to depression, which typically involves a strong sense of grief and hopelessness, doctors define anxiety as excessive worry and fear. As opposed to sadness, which can cause numbness and withdrawal, anxiety is frequently characterized by a constant state of alertness.

The physical symptoms of both illnesses are also possible. For instance, a depressed person can suffer changes in their food or sleeping habits, whereas an anxious person might have chest pain or a feeling of unease. Although anxiety and depression share certain similarities, it is important to recognize the fundamental distinctions in order to guarantee the optimum management and treatment strategy. The fundamental parallels and discrepancies between anxiety and depression, including the signs, causes, and approaches to therapy, are discussed in the following paragraphs.

We can suffer a great deal from the symptoms of anxiety and despair. An sudden job loss, a rocky relationship, or financial difficulties are just a few of the unexpected pressures that can hit us at once. These demanding life events can snowball into something overpowering if they aren't addressed with, which can eventually be the root of both anxiety and depression, or both anxiety and depression. These mental illnesses can have a harmful effect on our physically as well as our attitudes and perspectives on life if left untreated. Mental health conditions including anxiety and depression increase a person's risk of suicide 20 times more than it does for the overall population. Furthermore, physical issues like high blood pressure, heart issues, and pain can be exacerbated by worry and sadness.

Machine learning assists in the discovery of interesting patterns and knowledge based on dataset exploration. On the basis of publicly accessible Facebook data, previous academics have conducted analyses of depression. Based on the grammatical and emotional tone of the words used, the researchers conducted their study.

II. RELATED PROJECTS

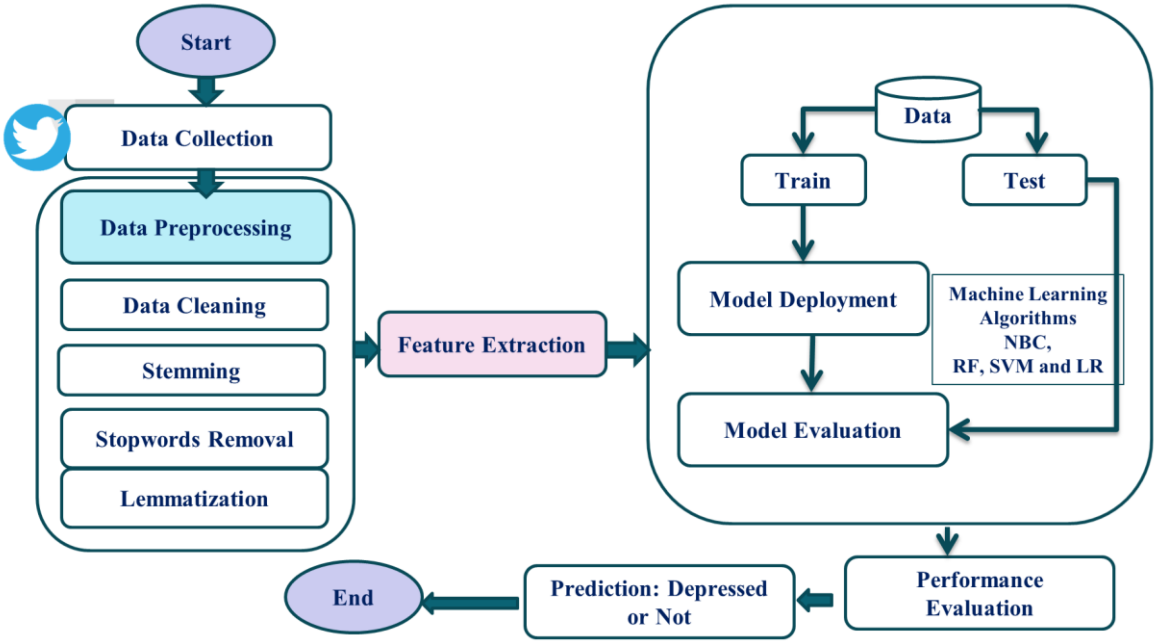
The *Fidel Cacheda et al.* investigated various machine learning (ML)-based algorithms for the primary recognition of MDDs dataset. They thoroughly analysed the dataset to determine how the subjects behaved based on the textual spreading, time gap, and temporal range of their compositions. They put out two distinct strategies based on singleton and dual machine learning. The concluding services two separate random forest (RF) classifiers, one to identify non-depressed people and another to detect depressed subjects uses random forest (RF) classifier with two threshold functions. *Sharath Chandra et al.* explained the rates of mental disease analysis have risen over the previous few eras, many cases still go undetected. Signs of mental illness can be seen on social media platforms like Twitter and Facebook as well as online forums, and automated technologies are getting better at spotting disorders like depression and other mental ailments. This application reviews previous studies that sought to use social media to predict mental illness. Selection surveys, the public posting of a identification on Twitter, or a user's involvement in an online forum have all been used to identify users who suffer from mental illnesses. *Ahmed Hosseini et al.* research problems are intricate conditions, identifying mental illness via social media can be a challenging undertaking. With social media becoming a more and more important part of people's lives, this research field has begun to change in recent years. Because of the close connection that exists between social media users and the platforms they use, personal information about users can be reflected on them, but with certain restrictions. A wealth of data on a person's life is available to researchers in such a setting.

Akshi, Aditi et al. find the main symptoms of this mixed anxiety-depressive condition include restlessness, disturbed sleep, and irregular mental processes. For the reasons stated above, we wish to select one of a few deep neural network designs that have been successfully applied to natural language processing applications as the most efficient architecture. An anxiety-related lexicon is created to identify the existence of anxiety markers. To look for abnormalities in posting behaviour, the time and frequency of tweets are examined. Opinion polarity analytics are also performed. Three classifiers (Multinomial Naive Bayes, Gradient Boosting, and Random Forest) are used to train the model, and ensemble voting classifiers are used to perform majority voting. The suggested model obtains a classification accuracy of 85.09% in preliminary results for tweets from the sampled 100 individuals.

III. PROPOSED METHODOLOGY

The main objectives are to design system to predict the level or severity of depression using tweeter dataset of any person. With the aid of machine learning algorithms, the proposed system aims to assist patients with this sickness in the early identification of depression's symptoms, which may be advantageous to both of them and their families. Determine a subject's likelihood of developing depression in the near future using information from their social media accounts. Discover the ML algorithm that offers the most accuracy. Find the best method for identifying depression in the test individuals. Recognize the amount of depression severity for a specific subject (prediction: Depression or not). Here, Created an application to identify upbeat and downbeat tweets.

Figure1: Proposed Systems



IV. SYSTEM IMPLEMENTATION

The purposed System Implementation developed to find severity & level depression on early stage using depression tweeter dataset. The implementation can classified into different modules of project and are listed as Data Collection, Data Preprocessing, Feature Selection / Extraction, Model Training and Testing, Machine Learning Model Deployment, Performance Evaluation, Comparison Study of The Model and some Exploratory Data Analysis (EDA) on dataset.

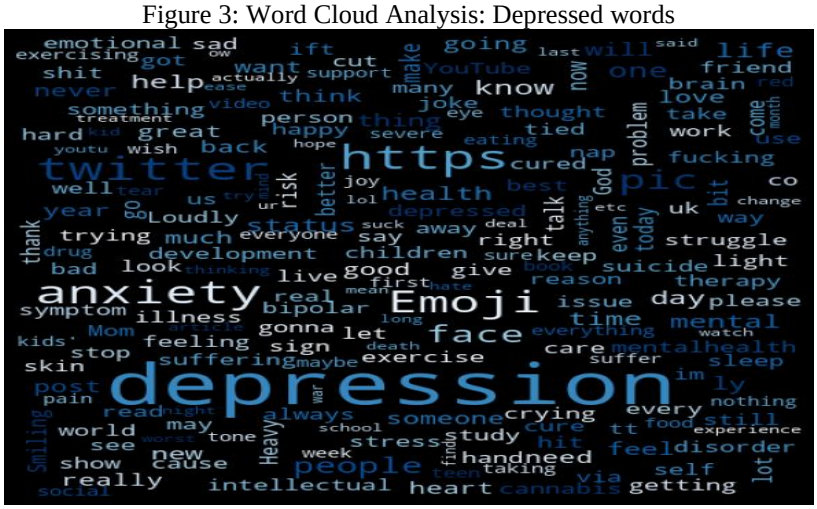
1. Text from Twitter - In this project we have used the dataset as Text collected from users through Twitter as shown in figure 2.

Figure 2: Tweeter Dataset

message	label
just had a real good moment. i missssssssss hi...	0
is reading manga http://plurk.com/p/mzp1e	0
@comeagainjen http://twitpic.com/2y2lx - http:...	0
@lapcat Need to send 'em to my accountant tomo...	0
ADD ME ON MYSPACE!!! myspace.com/LookThunder	0
so sleepy. good times tonight though	0
@SilkCharm re: #nbn as someone already said, d...	0
23 or 24ï¿½C possible today. Nice	0
nite twitterville workout in the am -ciao	0
@daNanner Night, darlin'! Sweet dreams to you	0
Good morning everybody!	0
Finally! I just created my WordPress Blog. The...	0
kisha they cnt get over u til they get out frm...	0
@nicolerichie Yes i remember that band, It was...	0
I really love reflections and shadows	0

2. Pre-processing of data – These techniques includes Tokenization, Lower casing, Stop words removal, Stemming and Lemmatization.

Figure 3: Word Cloud Analysis: Depressed words



3. Feature Extraction -Analyses the similarities between pieces of texts.

4. Classification – Building a machine learning model with NLP – Support Vector Machine, Decision tree, Naïve Bayes Classifier, Random Forest and KNN.

Figure 4: Model Accuracy

```
Svm Classifier has accuracy of 99.7431506849315 %
Decision tree Classifier has accuracy of 99.37304075235109 %
Random Forest Classifier has accuracy of 99.59090909090908 %
Naive Bayes Classifier has accuracy of 96.6286799620133 %
Knn Classifier has accuracy of 99.60282436010591 %
```

5. Predict Depression or not - Yields if the user is depressed or not. Here we created flask application to result as shown in figure 6 with the user input as a sentiment figure 5.

V. EXPERIMENTAL RESULTS

The experiment is carried on two different sentimental results i.e., first created queries (quiz), by answering the entire question and submit. The system has recognised that whether the person is depressed or not and severity level. Next, entering the direct text with user sentiment as shown in figure 5. Depression detection result shown in figure 6. Developed a flask web application, where we have number modules to be developed Login/Sign-up Module, Quiz based test, sentiment based test and remedies for the depressed one.

Figure 5: User input with sentiments

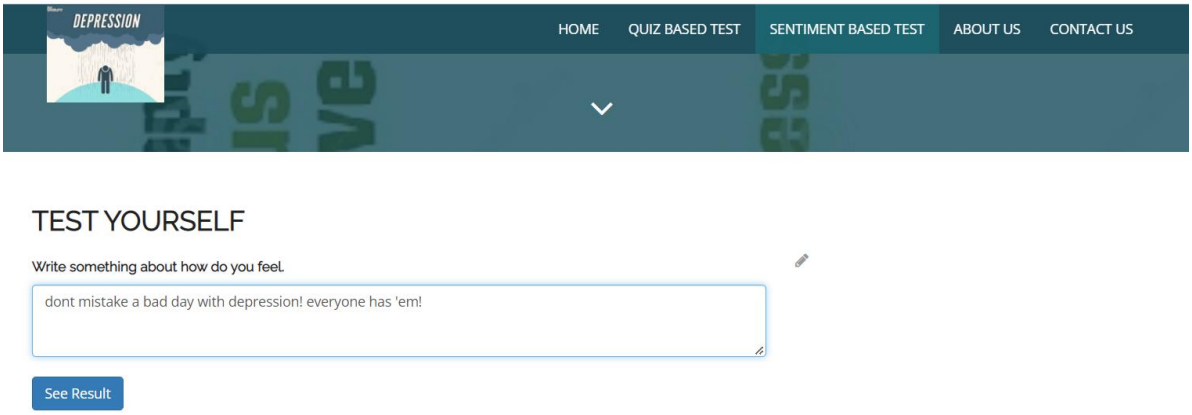
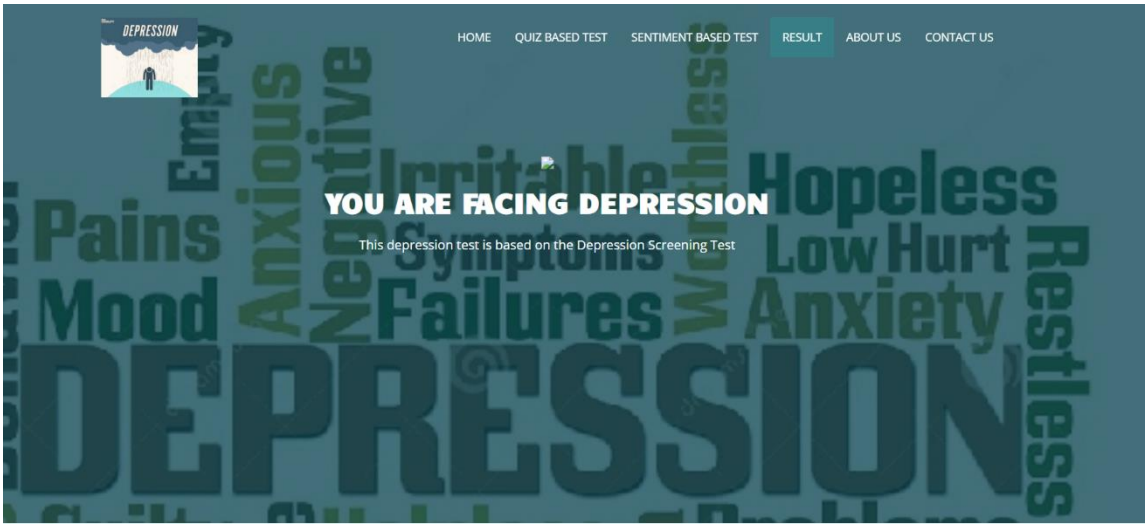


Figure 6: Prediction Result



VI. CONCLUSION

Finally, depression has gotten out of hand and is now a major issue that affects mental health. Moreover, the rapid development of social media has made it possible for users to save a wealth of information in their accounts. Twitter is a program that allows users to discuss their opinions and thoughts concisely and openly. Each tweet on Twitter is limited to 140 characters. The ability to collect and analyze the data supplied in each tweet makes it possible for researchers to do their work. Each tweet is given a sentiment score using the sentiment analysis technique, and then they are classified as positive, negative, or neutral tweets depending on their sentiment.

An algorithm that uses machine learning to classify tweets into the appropriate groups is fed the labeled tweets. To determine how well the chosen machine learning algorithms, Naive Bayes and Random Forest, classified depressive and non-depressive tweets, they were applied to two different sizes of tweet datasets. On the 3000 tweet dataset, the KNN algorithm has a classification accuracy of 99.60%, while on the 1000 tweet dataset, it has a classification accuracy of 99.34%. The 3000 tweets dataset, however, shows a Nave Bayes result of 97.31%, and the 1000 tweets dataset, a Nave Bayes result of 92.34%. By providing the same level of accuracy in the experiment, Naive Bayes and KNN are both as effective. The project, however, is just capable of producing text. To forecast the sentimental tweets, an application was made in this. Targeting a certain person and their tweets at a specific moment to assess whether they are depressed can improve the work in subsequent iterations.

REFERENCES:

[1] Fidel Cacheda, PhD; Diego Fernandez, PhD; Francisco J Novoa1, PhD; Victor Carneiro, PhD, "Early Detection of Depression: Social Network Analysis and Random Forest Techniques" ,2019.

[2] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar and Johannes C Eichstaedt, "Detecting depression and mental illness on social media: an integrative review", 2017.

[3] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi,Diana Inkpen, "Deep Learning for Depression Detection of Twitter Users",2018.

[4] Kali Cornn, Department of Statistics, Stanford University, "Identifying Depression on Social Media",2018.

[5] Akshi Kumar, Aditi Sharma, Anshika Arora, "Anxious Depression Prediction in Real-time Social Data".2019.

[6] Ravinder Ahuja, Alisha Banga, "Mental Stress Detection in University Students using Machine Learning Algorithms",2019.

[7] Md. Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M.Kamal, Hua Wang and Anwaar Ulhaq, "Depression detection from social network data using machine learning techniques",2018.

[8] Yunpeng Zhao, Yi Guo, Xing He, Jinhai Huo, Yonghui Wu, Xi Yang, and Jiang Bian, "Assessing Mental Health Signals among Sexual and Gender Minorities using Twitter Data",2018.

[9] U Srinivasulu Reddy, Aditya Vivek Thota and A Dharun, "MachineLearning Techniques for Stress Prediction in Working Employees",2018.

[10] Dritjon Gruda, Souleiman Hasan, "Feeling anxious? Perceiving anxiety intweets using machine learning",2019.

[11] Maryam Mohammed Aldarwish, Hafiz Farooq Ahmed, "PredictingDepression Levels Using Social Media Posts",2017.

[12] Kantinee Katchapakirin, Konlakorn Wongpatikaseree, Panida Yomaboot and Yongyos Kaewpitakkun, "Facebook Social Media for Depression Detection inthe Thai Community",2018.

[13] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat- Seng Chua andWenwu Zhu, "Depression Detection via Harvesting SocialMedia: A Multimodal Dictionary Learning Solution",2017.

[14] Punam B. Nalinde, Anita Shinde, "Machine Learning Framework for Detectionof Psychological Disorders at OSN", 2019.

[15] Michael M. Tadesee , Hongfel Lin, Bo Xu, and Liang Yang, "Detectionof Depression-Related Posts in Reddit Social Media Forum",2019.

[16] K Meenakshi, G Maragatham, "Computational intelligence in diagnosis and prognosis of gestational diabetes using deep learning", Research Journal of Pharmacy and Technology 12 (8), 3891-3895

[17] Meenakshi K, Maragatham, G, Agarwal, N., Ghosh, I, " A Data mining Technique for analysing and predicting the success of movie", Journal of Physics: Conference Series, Vol 1000, Issue 1, 2018.

[18] Saranya, G., Geetha, G., & Safa, M. (2017). E-antenatal assistance care using decision tree analytics and cluster analytics based supervised machine learning. 2017 International Conference on IoT and Application (ICIOT).

[19] Dhingra, S., Saranya, G., Diwakar, S., & Kumar, M. (2018). Fusion model for traffic sign detection, tracking and recognition. International Journal of Engineering & Technology, 7(3.12), 112.

[20] G Geetha, M Safa, C Fancy, D Saranya "A hybrid approach using collaborative filtering and content based filtering for recommender system", Journal of Physics: Conference Series Vol 1000, Issue 1, 2018.

[21] Geetha, G., Safa, M., Saranya, G., Subburaj, R. ,"An effective practices, strategies and technologies in the service industry to increase customer loyalty using map indicator" IEEE International Conference on IoT and its Applications, ICIOT 2017, 2017, 8073607

[22] Geetha, G., Safa, M., Fancy, C., Chittal, K., "3D face recognition using Hadoop",2017 International Conference on Energy, Communication, Data Analytics and Soft Computing,

[23] Safa, M., Pandian, A., A review on big IoT data analytics for improving QoS-based performance in system: Design, opportunities, and challenges "Lecture Notes in Networks and Systems, 2021, 130, pp. 433-4