

# A Machine Learning-based Approach for Depression Detection through Social Media

Yash Jain<sup>1</sup>, Sara Shaikh<sup>2</sup>, Soham Jadhav<sup>3</sup>, Pushkar Agnihotri<sup>4</sup> and Dr. Jayshree Bagade<sup>5</sup>

<sup>1-4</sup>Dept. of Information Technology, Vishwakarma Institute of Information Technology, Pune, India,

Email: yash.21910699@viit.ac.in, sara.21910006@viit.ac.in, ashish.21910249@viit.ac.in, soham.21910380@viit.ac.in

<sup>5</sup>Associate Professor, Dept. of Information Technology, Vishwakarma Institute of Information Technology, Pune (M.H.), India.

Email: jayashree.bagade@viit.ac.in

**Abstract**—Depression is a complex intellectual condition that causes someone to be moody and to feel persistently sad or pessimistic. Depressing syndromes are commonly a temporary answer to bereavement or anguish. Still, if the syndromes linger for more than two weeks, they grant permission to be evidence of a harsh emotional disorder. Social media has the potential to be a beneficial form for fast diagnosing and characterizing the attack of depression. We aim to check the notable attributes that form acceptable early signs of depression and accordingly detect depression earlier by deriving a dataset from Twitter of two together depressed and non-depressed users. We can predict attributes that guide public projects, behavior, and thoughts toward people, personality networks, and the use of conversation guide antidepressant causes through social media posting. The data hidden from tweets and posts discloses the consumer's concerning feelings and intuition disorder syndromes. We use machine intelligence as the foundation to extract data from these tweets utilizing methods such as NLP and the Bayes theorem to kill depression surely and efficiently. We trust that these verdicts will suffice in developing machines for use by healthcare artists or in assisting the suffering person in becoming more aware of his or her psychological state.

**Index Terms**— Twitter, Machine Learning, Natural Language Processing (NLP), TFIDF, SVM, Naïve Bayes, Gaussian Bayes, Random Forest Technique.

## I. INTRODUCTION

Worldwide, mental illness is the main cause of disability. Depression has become a serious health concern globally in the wake of the Covid 19 epidemic, with 322 million people suffering from it worldwide. The Western Pacific and Southeast Asian areas account for around half of these sad persons [2].

India has one of the world's highest rates of mental illness. According to a 2016 National Mental Health Survey, over 14% of India's population requires active mental health treatment. However, in the early phases of depression, 70% of individuals would not seek medical help, potentially advancing their disease.

Even though 87% of governments around the planet offer few basic health management duties to combat emotional disorders, 30% do not have a program, and 28% do not have a loyal budget. For insane energy (Detels, 2009). There has been a shift in the use of public news actual data for recognizing and forecasting variations in depression incidence recently. Early recognition of attraction syndromes and their situation according to the schedule [11] may prevent the onset of a large gloomy scene. De Choudhury and others [15]

assert that tweets associated with major depressive disorder, as well as their public dissemination project, can be used to classify and anticipate whether people are suffering from depression or are likely to suffer in the future. Tsugawa et al. and Coopersmith et al

investigated user activity on Twitter[13], whereas Nadeem et al., Harman et al., and Benton et al. investigated the depressive character of the user's tweets [11][12].

This study was designed to determine the likelihood of depression in a user on the sort of tweets they send and their social media activity. It might also be used to recognize the various mental disorders and abilities, as well as reform previous systems that would aid in the detection and control of depression on social media. This research makes use of Twitter API data containing over 20000 tweets. Many classifier algorithms are functioning to recognize depression levels, with the multinomial Naive Bayes classifier showing the results, with test accuracy of 0.94 and F-measure of 0.94.

The rest of the paper is further divided into sections as follows. Section 2 contains the literature review. Section 3 explains the methodology of the study, the metadata of the data used, and the machine learning technique used. Section 4 describes and discusses the results. Finally, Section 5 shows the conclusions of the study.

## II. RELATED WORK

Attempts to diagnose depression using Twitter data have increased as a result of the rising use of social media platforms and the high amount of self-disclosure on these platforms. The authors of [1] trained and classified it using classification machine algorithms in various stages of depression. In [2], the authors analyzed six different machine learning classifiers that used a few socio-demographic and mental characteristics to determine whether a person was depressed. Select K-Best Features (SelectKBest), Minimum Redundancy and Maximum Relevance (mRMR), and the Boruta feature selection algorithms are used. The authors of [3] used three different selection models to find the subset of features that best represents the characteristics of an EEG signal, and then applied six different machine learning models to all the subsets of features to find the model with the best accuracy and recall for detecting depression. The authors of [4] outline the main ML-based techniques for detecting depression and other mental diseases using behavioral data. Authors of [5] perform an analysis of a standard dataset obtained from online social media, where detection can be based on a machine learning algorithm using SVM and Naïve Bayes algorithm. Authors of [6] propose a classifier to improve the prediction accuracy of machine learning techniques on big datasets and identify sad patients based on connections between items on the QoL scale and mental illness. The authors of [7] suggest an algorithm that extracts emotion readings from the user's input text before announcing whether the material contains any depression. Authors of [8] propose a weakly supervised learning framework for detecting social anxiety and depression from long audio clips. They present a novel feature modeling technique named NN2Vec that identifies and exploits the inherent relationship between speakers' vocal states and symptoms/affective states. Using characteristics derived from a person's network and tweets, the authors of [9] trained and tested classifiers to determine if a user is sad or not. The key contribution of this work is the examination of the traits and their influence on identifying depression levels. The authors of [10] employ a semi-supervised statistical model to assess how the duration of these symptoms, as well as their manifestation on Twitter (in terms of word usage patterns and topical preferences), correspond to medical results provided via the PHQ-9. With an accuracy of 68 percent and a precision of 72 percent, our proactive and automated screening tool can detect clinical depression symptoms. The authors of [11] propose a novel strategy for building our classifier by considering social media as a text-classification problem rather than a behavioral challenge on social media sites. We achieved an 81 percent classification accuracy rate with a precision score of 0.86 utilizing a corpus of 2.5 million tweets. The authors of [12] look at post-traumatic stress disorder (PTSD), a devastating ailment that affects millions of people throughout the world, with military veterans having particularly high rates. They also offer a unique way for obtaining a PTSD classifier for social media using simple searches of publicly accessible Twitter data, which results in a considerable decrease in the cost of training data compared to earlier work. The authors of [13] examine mental health issues using publicly available Twitter data illustrating how rigorous application of simple natural language processing methods may offer insight into specific diseases as well as mental health in general, as well as indications that yet-to-be-discovered linguistic signals pertinent to mental health exist in social media. They provide a demonstration of a new approach for swiftly and cheaply acquiring data on a variety of mental diseases, with an emphasis on four, in particular, PTSD, depression, bipolar disorder, and other mental illnesses, in particular, Seasonal Affective Disorder (SAD) is a kind (SAD). The authors of [14] utilize crowdsourcing to gather a group of Twitter users

who have been diagnosed with clinical depression using a conventional psychometric instrument. They evaluate behavioral traits linked to social engagement, emotion, language and linguistic styles, ego network, and mentions of antidepressant drugs through their social media postings over a year before the development of depression. The authors of the study [15] look at a variety of language signals that might assist in detecting emotion cause events, such as the location of the cause event and the experiencer about the emotion keyword: emotional processes such as positive emotion, negative emotion, sorrow, rage, and anxiety.

### III. METHODOLOGY

Nowadays, on practically every trendy issue, people express their feelings through judgment, reviews, and opinions. Because individuals are so open about their emotions on social media sites, a machine learning method may be utilized to detect depression. We present an overview of several technological techniques for detecting depression and mental diseases using Natural Language Processing (NLP) and text classification algorithms in this work. Following the machine learning classifiers, the framework includes data pretreatment and feature extraction.

#### A. Dataset

The tweets in this dataset were obtained using the Twitter API (elevated). For the train and test datasets for our model, we collected over 20,000 tweets that were based on the following keywords: 'Diagnosed with depression,' 'I am depressed,' 'Depressed,' 'Suicidal'. To avoid overfitting, we'll use 10-fold cross-validation.

#### B. System Architecture

A study was conducted to build and analyze a variety of machine learning classification models in order to determine if a Twitter user's tweets are focused on depression or mental illness. To ensure accuracy and avoid overfitting, the model is trained using cross-validation.

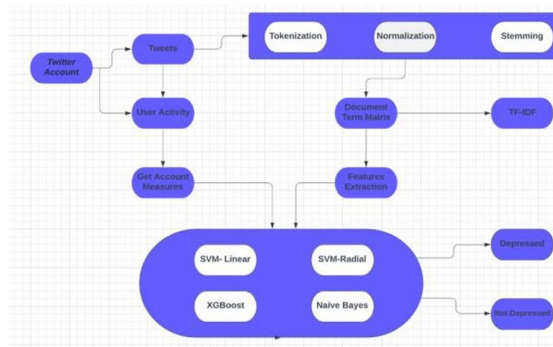


Fig 1: Architecture of the system

To begin, tweets are obtained using various keywords, and also user accounts and relevant details such as the number of followers and friends, via the Twitter API. Before being combined into a single document, tweets and user activity data are preprocessed separately.

#### C. Data Preprocessing

We have two categories of data in our dataset: text and numerical. Text data is pre-processed using natural language processing, while numerical data is pre-processed using normalization.

##### Data Cleaning

First, we tokenize and partition the tweets to remove any noisy data. URLs, hashtags, mentions, punctuation, and stopwords are all removed using regular expressions. Emojis and emoticons are not removed since they might carry vital information about the sentiment. Then we use stemming to break the words to their root kind and group related terms together.

Normalization: To prepare the data for our machine learning model, we performed normalization on the user activity data, as well as the number of friends and followers. The data is normalized using RobustScaler, which uses the first and third quartiles.

### Feature Extraction

The fundamental disadvantage of language processing is that machine learning algorithms can't handle raw text as an input. As a result, we employ feature extraction software that extracts essential textual features and converts them into numerical vectors.

Bag-Of-Words is the feature extraction technique we're using here.

*Bag-Of-Words*:- A bag-of-words tool extracts information from text for modeling purposes and is widely used in NLP machine learning algorithms. A vocabulary of unique words is built after preprocessing and formatting the text input in lowercase letters. After that, a feature matrix is built by assigning a column to each phrase or word and a row to each sentiment. Each entry in the matrix represents the existence of a word in the emotion.

To minimize overfitting, we picked terms from the lexicon that appear in more than 5 phrases and in less than 90% of the phrases.

### D. ML Classification Techniques

#### Support Vector Machine

The SVM algorithm's purpose is to find the optimum decision boundary for partitioning n-dimensional space into classes so that we can simply move fresh data points into the appropriate category. The hyperplane refers to the optimal choice boundary. SVM chooses the extreme points that will aid in the formation of the hyperplane. Support vectors represent extreme examples, which is why the technique is named a support vector machine. We utilize SVM for text classification because it can manage the large number of characteristics that text classification generates. The number of features has no effect on SVM since it implements the margin through data separation rather than the number of features. This suggests that if our data can be separated by a large margin using functions from the hypothesis space, SVM will perform well efficiently. It is unaffected by mislabeled data, which is common with textual data.

#### Naïve Bayes

The Naive Bayes algorithm is a straightforward classification method that relies on event probability. It's based on Bayes' theorem, which asserts that characteristics have no association. When compared to other text classifiers, Naive Bayes models let each feature participate in the final decision equally and independently from other features, making them more computationally.

## IV. RESULTS AND DISCUSSION

### A. Evaluation Measures

The validity of these predictions will be evaluated and cross-validated using the following metrics: average accuracy (A), average precision (P), average recall (R), average F1 scores, confusion matrix (CM), AUC score, and recipient operating classification curves (ROC).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

The proportion of positive points properly predicted by the dataset divided by the total number of positive data points in the dataset is called recall. Precision is defined as the ratio of accurately predicted positive points to the total number of positive points predicted by the model. As a result, achieving good recall and accuracy is critical. The f1 score represents the tradeoff between precision and recall.

$$\text{F1 Score} = \frac{2 \times \text{P} \times \text{R}}{\text{P} + \text{R}} \quad (2)$$

Using the metrics defined below, the Confusion Matrix can be utilized to construct a point in the ROC space:

$$\text{CM} = \begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} \quad (3)$$

Where, TP = True Positives TN = True Negatives, FP = False Positives, FN = False Negatives.  
The AUC\_ROC assessment technique uses the TPR and the FPR rate.

## B. Results

Cross-validation and other categorization methods are used to test every conceivable combination of input data (SVM-linear, Random Forest, and Naive Bayes). Depressed/not depressed is the anticipated label for every training/testing sample. According to Word Cloud, "Love" and "life" are the most often used positive terms. "Depressed," "panic attack," and "suicide" are the most common negative terms. Figure 2 indicates that all the recommended classification algorithms worked admirably on our data. SVM-F1 Linear's measure for test data is 0.994, whereas the random forest's F1 measure is 0.995, this shows abnormally high-test results. Even with varied vectorizer and feature settings, SVM-Linear surpasses all other classifiers. Both Naive bayes classifiers show reasonable results with F1-scores of 0.84 and 0.94.

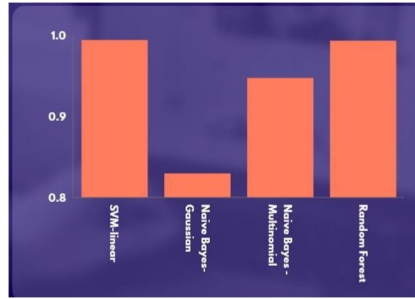


Figure 2: F1-Measures for different algorithms

Table I illustrates the relative importance of the top ten characteristics in our classifiers. The top three attributes are all positive terms, indicating that positive words are given greater weight than negative phrases. The recall score of several classifiers may be used to confirm our hypothesis. Table 2 shows different recall scores of the classifiers used.

TABLE I: TOP 10 STRONGEST FEATURES

| Serial No. | Features | Strength |
|------------|----------|----------|
| 0          | feel     | 932.40   |
| 1          | great    | 840.31   |
| 2          | love     | 822.97   |
| 3          | life     | 776.87   |
| 4          | depress  | 660.75   |
| 5          | like     | 618.98   |
| 6          | suicide  | 472.22   |
| 7          | anxiety  | 461.92   |
| 8          | attack   | 459.21   |
| 9          | panic    | 439.23   |

Table 2 demonstrates that recall is smaller than precision, proving our previous premise. It also demonstrates that the Gaussian Naive Bayes classifier is more sensitive to feature vector strength.

TABLE II: PRECISION VS. RECALL

| Algorithms              | Precision | Recall |
|-------------------------|-----------|--------|
| Gaussian Naïve Bayes    | 0.86      | 0.81   |
| Multinomial Naïve Bayes | 0.95      | 0.93   |

Min\_df is a TF-IDF vectorizer hyperparameter that sets a lower constraint on the frequency of words in the corpus. Outliers and redundant characteristics are extremely sensitive to the Naive Bayes classification method.

Table III demonstrates how varying lower constraints on word frequency affect the performance of Naive Bayes classifiers. NB performance converges at a min df of 15, removing words that appear less than 15 times in our dataset.

TABLE III: DOCUMENT FREQUENCY OF RARE WORDS

| Min_df    | 1    | 5    | 15   |
|-----------|------|------|------|
| Accuracy  | 0.92 | 0.93 | 0.95 |
| Precision | 0.95 | 0.94 | 0.96 |
| Recall    | 0.89 | 0.92 | 0.94 |
| F1_Score  | 0.92 | 0.93 | 0.95 |

TABLE IV: COMPARISON FROM PAST PAPERS

| Name of Paper                                                                               | Acc  | F1 score |
|---------------------------------------------------------------------------------------------|------|----------|
| Depression Detection of Tweets and A Comparative Test.                                      | 0.90 | 0.74     |
| A profile-based sentiment-aware approach for depression detection in social media           | 0.69 | 0.89     |
| Depression Detection via Harvesting social media: A Multimodal Dictionary Learning Solution | 0.85 | 0.85     |
| Predicting Depression via social media                                                      | 0.51 | 0.72     |
| Our Paper                                                                                   | 0.94 | 0.94     |

## V. CONCLUSIONS

The system gathers information from social media and analyzes the user's mood. We can access a number of candidates by keeping an eye on their social media accounts because practically everyone nowadays utilizes social media platforms to describe their life experiences. The tweets are computed using a natural language processing algorithm. Support Vector Machines (SVM), Naive Bayes (Gaussian Naive Bayes, Multinomial Naive Bayes), and Random Forest are used in this technique. The system can identify mental states such as happiness, anxiety, melancholy, pleasure, anxiety, and so on. The system collects tweets written by a Twitter user and classifies the kind of depression using the Twitter API. As a result, it also conveys the intensity of depression. Psychiatrists, NGOs, and other charitable organizations can utilize this technique to keep track of their patients' or applicants' emotions.

## REFERENCES

- [1] Machine Learning Based On Approach For Detection Of Depression Using Social Media Using Sentiment Analysis Geeta Tiwari<sup>1</sup> , Gaurav Das <sup>2</sup> Assistant Professor, Department of CSE Compucom Institute of Technology and Management , Jaipur Volume 6 Issue 3 March 2021.
- [2] Current Research in Behavioral Sciences. An in-depth analysis of machine learning approaches to predict depression 2021 Md.Sabab Zulfike rNasrin Kabir Al Amin Biswas Tahmina Nazneen Mohammad Shorif Uddin.
- [3] International Journal of Pharma Medicine and Biological Sciences Vol. 10, No. 1, January 2021. An EEG-Based Depression Detection Method Using Machine Learning Model Ran Bai<sup>1</sup> , Yu Guo<sup>2</sup> , Xianwu Tan<sup>1</sup> Lei Feng<sup>3,4,5</sup>, and Haiyong Xie<sup>5,1</sup> <sup>1</sup>National Engineering Laboratory for Public Safety Risk Perception and Control by Big Data (NEL-PSRPC), Beijing, China. Journal of personalized medicine.
- [4] Review Machine Learning-Based Behavioral Diagnostic Tools for uuDepression: Advances, Challenges, and Future Directions Thalia Richter <sup>1,\*</sup>,Barak Fishbain <sup>2</sup> , Gal Richter-Levin <sup>1</sup> and Hadas Okon-Singer.
- [5] Journal of trends in Computer Science and Smart technology (TCSST) (2021) Vol.03/ No. 01 Pages: 24-39 <https://www.irojournals.com/tcsst/> DOI:
- [6] Detecting depression using an ensemble classifier based on Quality of Life scales Xiaohui Tao<sup>1\*</sup> ,Oliver Chi<sup>2</sup> , Patrick J. Delaney<sup>1</sup> , Lin Li<sup>3</sup> and Jiajin Huang<sup>4</sup>.
- [7] Depression Detection using Machine Learning K. Sudha , S. Sreemathi , B. Nathiya , D. Rahini Priya Associate Professor, CSE Department, Muthayammal Engineering College, Tamil Nadu 2020.

- [8] A Weakly Supervised Learning Framework for Detecting Social Anxiety and Depression. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 2, No. 2, Article 81. Publication date: June 2018.
- [9] Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features Hatoun AlSagri 1,2 Mourad Ykhlef
- [10] A.H. Yazdavar, H.S. Al-Olimat, T. Banerjee, K. Thirunarayan, and A.P. Sheth, “Analyzing clinical depressive symptoms in twitter”,2016.
- [11] M. Nadeem, “Identifying depression on twitter,” arXiv preprint arXiv:1607.07384, 2016.
- [12] G. Harman and M.H. Dredze, “Measuring post-traumatic stress disorder in twitter,” ICWSM, 2014. G. Coppersmith, M. Dredze, and C. Harman.
- [13] “Quantifying mental health signals in twitter,” Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality, pp.51–60, 2014. G. Coppersmith, M. Dredze, and C. Harman.
- [14] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. Predicting depression via social media. ICWSM. 13, 1-10 (2013).
- [15] Depression detection from social network data using machine learning techniques Md. Rafiqul Islam,<sup>1</sup> Muhammad Ashad Kabir,<sup>2</sup> Ashir Ahmed,<sup>3</sup> Abu Raihan M. Kamal,<sup>1</sup> Hua Wang,<sup>4</sup> and Anwaar Ulhaq<sup>5</sup>.