# Depression Detection from Social Network Data using Machine Learning Technique

Prajakta Kumbhar
*Department of Computer Engineering*
*Modern Education Society's College*
*Of Engineering*
Pune, India

Vaishnavi Kothari
*Department of Computer Engineering*
*Modern Education Society's College*
*Of Engineering*
Pune, India

Divya Patil
*Department of Computer Engineering*
*Modern Education Society's College*
*Of Engineering*
Pune, India

Bhakti Pawar
*Department of Computer Engineering*
*Modern Education Society's College*
*Of Engineering*
Pune, India

Prof.Shilpa Khedkar
*Department of Computer Engineering*
*Modern Education Society's College*
*Of Engineering*
Pune, India

***Abstract*— Online social networks provide relevant information on users' opinions and posts on various topics. So applications, such as monitoring and detection systems can collect and analyze this data. This paper studies an information based system, which includes an emotional health monitoring system to detect users with possible psychological disorders specially depression and stress. Symptoms of this psychological disorder are usually observed passively. In this situation, the author argues that online social behavior extraction offers an opportunity to actively identify psychological disorders at an early stage. It is hard to recognize the confusion in light of the fact that the mental components considered in standard symptomatic criteria surveys can't be seen by the registers of online social exercises. Our proposed methodology is new and creative for the act of psychological disorders. In informal organizations which misuses the highlights removed from interpersonal organization information to relate to exactness potential instances of confusion discovery. We perform an analysis of the characteristics and we also apply machine learning in large-scale data sets and analyze features of the types of psychological disorders using different algorithms like ANN, RNN and Naive Bayes and output is generated and users are classified according to their features (happy, low,gloomy, etc.)**

***Keywords*— *Depression Detection, social media, Machine learning, Deep Learning***

## I. INTRODUCTION

"Mental Pain is less dramatic than physical pain, but it is more common and also harder to bear",said by C.S Lewis. Depression is a common mental illness and a leading cause of disability worldwide, which may cause suicides. e. Depression affects one in every 15 adults in a given year and the risk in women is twice than men[1]. Globally, more than 280 million people of all groups suffer from depression. Women are more affected by depression than men[2]. However, at early stages of depression, 70% of the patients would not consult doctors, which may take their condition to advanced stages.Although depression is common, the condition can be different for different people. Some of the contributing factors include genetics, exercise, and diet. Changing your diet can help lower your risk of depression[3].

According to a report released in 2017, around 17 million people in the US experience depression. In 2018, the CDC stated that almost 2 million children under the age of 3 were diagnosed with depression[3].

Social media applications provide a space for their users to share or write about their opinion and random notes which relate to their feelings and emotions in words. The text shared on social media contains valuable insight that can be used for various intelligent applications in the real world such as healthcare, entertainment, politics, view communication, and tourism. The content or points created by the user is the data that is valuable for the researchers to analyze the state of mind . Along with the growth of social media applications, people start to share their mental health battles with the world via social media instead of keeping it private. The data which is available in the user account can be analyzed to figure out their level of mental health and an opportunity to help them to recover.

Since contagious negative emotions in social networks adversely affect people, leading to depression and other mental illnesses. As a result, researchers are able to detect whether someone is depressed or not based on the comments that they make on their social media accounts.This paper focuses on the different ways used for classifying a given piece of language text in line with the opinions expressed in it.Sentiment analysis is the process to identify the tone of the text is either positive, negative, or neutral. A sentiment analysis system will assign a score for every text word based on the designed polarity. This can allow identifying the state of the user whether they are in a positive mood or negative mood.

Then researchers implement machine-learning algorithms like ANN,RNN and Naive Bayes to detect depression based on the data which was labeled with sentiment scores. The efficiency of detection is evaluated based on the accuracy of the machine learning algorithms.This process includes:

1) Collecting the data from social media;
2) Because the data is typically made in posts format, text pre-processing techniques are used to improve quality.
3) The extraction of features from the sentences;
4) the incorporation of these characteristics into machine learning algorithms to create a judgment mechanism capable of predicting whether the user is depressed or not.

## II. LITERATURE SURVEY

There has been a lot of research in the subject of depression detection during the last few years, and various research publications have been published in this topic.

The 'DAIC-WOZ' dataset was used in [1] which has class imbalances. To solve the class imbalance , the Synthetic Minority Oversampling technique(SMOTE) was used. Logistic Regression, Random Forest and SVM are used for classifying the users into depressed and non-depressed users. SVM with SMOTE analysis gave the best result with 93% of accuracy. An application named "CureD" was created where users can check their depression levels by answering a standard PHQ-8 questionnaire, followed by recording the voice by reading out the passage displayed on the screen.

User data of Sina Weibo to classify college students using a deep integrated support vector machine(DISVM) algorithm was used by authors of [4]. It classifies the input data, and then realizes the recognition of depression. It is found that DISVM makes the recognition model more stable and improves the accuracy of depression diagnosis to a certain extent. Also the proposed depression recognition scheme can detect potential depression patients in the college student population through Sina Weibo data.

[5]Authors developed a methodology where they used a combination of machine learning algorithms to get best results. They developed two ensemble models : in the first model K-NN, Support Vector Machine and Logistic Regression algorithms were used and in the second model Decision Tree algorithm, Naïve Bayes classifier, Support Vector Machine algorithms were used. The mean accuracies were calculated as running the whole code each time gave different results because each time different predictions were voted out. The ensemble model 1 gave the best result with average accuracy of 89.6%.

Authors of paper[6] proposed a two-staged method where in first stage sentimental analysis was applied on a particular individual's twitter posts to predict binary classes (i.e. depressed/not depressed) and then a deep learning module long short term memory (LSTM) and CNN was employed. In the second stage the dataset was divided into train and test set and then three vectorizers: count vectorizer, TF-IDF and n-grams were used for vectorizing tweets.

[7]Developed a system where the videos are given as an input and converted into frames which are passed through a neural network for face detection and then passed to a trained CNN model for feature extraction, classification generating an output of emotion vector. In the process of depression level detection, the two crucial components are video input and the Beck Depression InventoryII. The solution generates as a result of the correlation between emotion vector and inventory vector represented using visual graphics. The system provides the result to the user in the form of a document consisting the detected depression level.

Authors of [8] used Electroencephalogram (EEG) and eye movements (EMs) data for depression detection due to their advantages of easy recording and non-invasion. They proposed a content based ensemble method (CBEM) to promote depression detection accuracy. In the proposed model, EEG or EMs dataset were divided into subsets and then a majority vote strategy was used to determine the subjects' labels. The validation of the method is testified on two datasets which included free viewing eye tracking and resting-state EEG, and these two datasets have 36,34 subjects respectively. For these two datasets, CBEM achieves accuracies of 82.5% and 92.65% respectively. The results show that CBEM outperforms traditional classification methods.

The authors of [9] used an approach which adopts a multi-stage machine learning pipeline. First, they project mobility features of the majority class (undepressed users) using autoencoders and then the trained autoencoder classifies a test set of users as either depressed (anomalous) or not depressed (inliers) using a One Class SVM algorithm.

[10]Authors developed a multilayers deep learning model to classify users with depression. Convolutional Neural Networks (CNNs), Gated Recurrent Units (GRUs), and Multilayer Perceptrons (MLPs) were used for training. The training was done in 2 steps. The first was fitted to classify each post into either general- or mental health-related posts. The second step observed the changing patterns of users to classify users into non-depression and depression groups.

A hybrid model was proposed by authors of paper[11] which combines the factor graph model (FGM) with a convolutional neural network (CNN).The posting behaviour and social interaction of the user were the attributes of the dataset. Also algorithms like Support Vector Machine (SVM), Gradient Boosted Decision Tree (GBDT), Deep Neural Network (DNN) were applied and the results were compared with the hybrid model.

[12]Authors used the AVEC dataset which contains webcam recordings of people during a certain task. Principal Components Analysis (PCA) was used for feature selection for both audio and video recordings. Fusion and classification techniques were used. In fusion a) the feature level fusion, where the two feature sets (video and audio) were concatenated to form a unique vector, and b) the decision level fusion, where classification results from different classifiers, trained separately on video and audio data, were combined through the AND and OR operands. Classification was performed in two different modes: gender based and gender independent.

[13]Total 10,000 tweets were collected using a twitter API. The two machine learning algorithms Multinomial Naive Bayes(MNB) and Support Vector Machine(SVM)

were used for classification. To improve the quality of the training data, the tokenized text is assigned the respective parts of speech by using POS Tagger. MNB has performed the best with the F1 score of 83.29 whereas SVM has achieved a lower F1 score of 79.73.

[14]Developed a method where males and females were modeled separately and different weights were provided for different speech types and emotions according to their respective contributions in detecting depression.

Three models were used by authors of paper[15] : (a)using machine learning classifiers and WEKA ,(b)using imaging and machine learning methods and (c)risk factors. Machine learning classifiers like Bayes Net Classifier (BN), Multilayer perceptron (MLP), Logistic regression, Decision Trees, Sequential Minimal Optimisation were used. Bayes Net classifier gave the best results with accuracy of 95%.

Authors of paper[16] implemented a technique in which phrases were extracted from social networks which were passed through a sentimental analysis based on a lexical dictionary. The audio, messages and interactions of users with other users were considered. The sentiment intensities were calculated for the phrases. The sequential minimal optimization(SMO) algorithm was applied to classify the mood phrases which gave 86.1% of correct results .

An Emotion recognition system was designed by authors[17] for classrooms using a deep Convolutional Neural Networks (CNN) architecture. Student's faces were recognised during the teaching sessions and the images were converted into LBP codes to make the system more robust to illumination variations. And then multiple CNNs were used to predict the class.

[18]Authors designed a behavior recommendation system which considers the person's current behavior and health and recommends some changes in his/her behavior to improve his/her health. Contrast pattern mining was used to recommend behavior changes. For comparison purposes, the RANDKN and TOPCP approach was used. A Canadian Community Health Care Survey Data was taken as a dataset.

Authors of paper[19] used James Pennebaker and Laura King's stream-of-consciousness essay dataset. A multiple layer perceptron (MLP) with one hidden layer was trained along with CNN. Applying SVM with CNN did not improve the results but applying MLP alone improved the results.

Authors of paper[20] implemented a linear regression technique to examine social media platforms like facebook and twitter to examine how official statistical institutes interact with citizens. Authors have found out that twitter is more powerful than facebook . NodeXL is used as a powerful and convenient interactive network visualization and analysis tool that leverages the widely available MS Excel application as the platform for representing generic graph data, performing advanced network analysis and visual exploration of networks.

A Sentimeter-Br2 based on Sentimeter-Br is used by authors of paper[21] to improve the performance of music recommendation systems. Sentimeter-Br2 considers n-grams, adverbs, removes stopwords and the differing value of sentiments depending on the verbal tenses, in which a verb in the past tense is of lesser sentimental value than a verb in the present tense. The music was recommended on the basis of the user's current social media data. If there was no data updated by the user , then music of his/her style was recommended.

[22]Proposed a system that recommends songs based on the user's listening history and content based audio information with the contextual emotion information mined from user-generated articles. The authors used a factorization machine technique. It is found that effective contextual text information is more effective instead of just simple word counts of the articles the users write as the context feature.

## III. PROPOSED SYSTEM

Machine learning strategies square measure trained on datasets and a model is made for analysis. Based on the accuracy of the model, the machine learning technique is appropriate. The three methods in machine learning algorithms square measure supervised learning, unsupervised learning and reinforcement learning. In supervised learning, the model is trained victimization labeled information that contains each input and results. The sections of the process square measure the coaching section and testing phase. Unsupervised learning strategies don't use coaching information or labeled information. It finds the hidden structures or patterns from unlabeled information.

### Supervised Learning

Supervised learning needs a well-labeled dataset to coach. supervised learning is of 2 types particularly regression and classification. Classification techniques facilitate the seeking out the appropriate category labels which may predict the positive, negative and neutral sentiments. A machine learning model is developed that uses the tagged knowledge to coach, classify the posts and predict the emotions of the posts. Random Forest, Bayesian belief network, Naive Thomas Bayes and ANN classifiers are a number of the algorithms that are employed in this method.

### Unsupervised Learning

Unsupervised ways are supported by machine learning or lexicon. The necessity of the labeled datasets isn't needed in unsupervised learning. Sentiment analysis once done using unsupervised learning; it's typically supported by a Sentiment Lexicon. Text classification helps to extract phrases that contain adjectives or adverbs to estimate a phrase's linguistic orientation. linguistics orientation is then accustomed to classify the emotion.

Our main aim is to predict Depression based on their social media data which is in text format. The suggested framework contains five major phases, as shown in Figure 1:

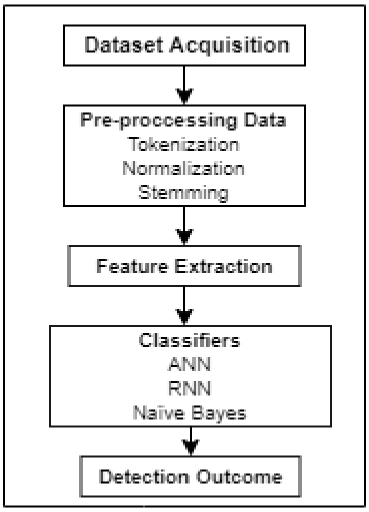

Fig.1 Proposed Model

1. Data is collected from a respected dataset.
2. On the dataset, data pre-processing techniques such as Tokenization, Normalization and stemming.
3. Extract the different features in view of psychogeriatric measurements from the user's post.
4. ANN, RNN and Naïve Bayer are classifier techniques used to process the data.
5. Model output is displayed on the screen.

The first step of implementation is gathering data from a dataset which is obtained from Reddit dataset/Kaggle. This dataset contains posts and comments of the social media in text format. Firstly, all posts for depressed and non-depressed accounts, as well as information of user accounts and activities such as number of followers, number of following, time of posts, number of mentions, and number of repost, are retrieved. We have collected a total of 3000 Posts for the generation of the training module and testing module for our model. Data collection is followed by pre-processing which is applied on text posts. we used tokenization method for splitting the sentences to each words,the tokenized sentence are then processed for normalization and lemmanization which helped to removing stop words and eliminate the affixes from the word.In feature extraction to describe and demonstrate amongst depressive and non-depressive posts, we extract the different features in view of psycholinguistic measurements from the user's post. Feature extraction process helped in the selection of keyword like from these is statement "I am happy" here the keyword is happy, so these technique is used for , remove keyword and send to trained module. These keywords are classified according to 3 levels of depression: first low level, second medium level and third high level depressed level by using algorithm ANN, RNN and naïve bayer and it will train the keywords. The last stage would predict the result whether the person is depressed or non-depressed. If a person is highly depressed, we will send a motivational message . Final output mainly contains the accuracy of the model which is compared with pre-trained data and the result is displayed.

## IV. ADVANTAGES

1. The presents Knowledge-Based Recommendation System (KBRS), which includes an emotional health monitoring system to detect users with potential psychological disturbances, specifically depression and stress.
2. Textual mood detection according to time series using text inputs.
3. Predict mood level based on score or weight with class label.
4. Successfully implemented the test model based on a training set as supervised learning approach.
5. Beneficial for early detection of depression.
6. Execute the proposed system maximum accuracy.

## V. LIMITATIONS

1. The analysis of posts is an example of this, for they are usually coupled with hashtags, emoticons and links, creating difficulties in determining the expressed sentiment. In addition, there is a need for automatic techniques that require large datasets of annotated posts or lexical databases where emotional words are associated with sentiment values. Another important aspect is that analyses are suitable for the English language, in which there is a limitation for other languages.
2. In applying automatic analysis due to the difficulty to implement it because of the ambiguity of natural language and also the characteristics of the posted content

## VI. CONCLUSION

We defined a classification problem as identifying whether a person is depressed, based on his/her social media profile activity. We develop a predictive model to predict whether user posts are depressed or not based on detecting depressed users using a machine learning approach and sentiment analysis. Different machine learning algorithms are applied and different feature datasets are explored. Many preprocessing steps are performed, including data preparation and aligning, data labeling, and feature extraction. ANN ,RNN and Naive Bayes are the selected machine learning algorithms which have been applied on post dataset to find the accuracy of the algorithm on classifying the depressed and non-depressed users. This study can be considered as a step towards building a complete social media-based platform for analyzing and predicting mental and psychological issues and recommending solutions for these users.

## VII. FUTURE SCOPE

In future, the work can be enhanced by including some additional features of online users on social media behavior .e.g time of post and interaction with other users. However the analysis is limited to text only ,further research can improve the system by adding features like text in different languages ,audio messages ,image and video based depression detection.

## ACKNOWLEDGMENT

We would like to thank the Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, for all the support provided for this research. And also we'd like to express our gratitude towards Prof. Shilpa. P. Khedkar for sharing their pearls of knowledge with us during the research process.

## REFERENCES

[1] Yalamanchili, Bhanusree, Nikhil Sai Kota, Maruthi Saketh Abbaraju, Venkata Sai Sathwik Nadella, and Sandeep Varma Alluri. "Real-time Acoustic based Depression Detection using Machine Learning Techniques." In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pp. 1-6. IEEE, 2020.

[2] W. H. Organisation, "Depression," World Health Organisation, 13 September 2021. [Online]. Available: https://www.who.int/newsroom/fact-sheets/detail/depression.

[3] N. Schimelpfening , "Why Some People Are More Prone to Depression Than Others," Verywellmind, 26 March 2021. [Online]. Available: Why Some People Are More Prone to Depression Than Others.

[4] Ding, Yan, et al. "A depression recognition method for college students using deep integrated support vector algorithm." IEEE Access 8 (2020): 75616-75629.

[5] Kumar, Piyush, Rishi Chauhan, Thompson Stephan, Achyut Shankar, and Sanjeev Thakur. "A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection." In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 568-574. IEEE, 2021.

[6] Shetty, Nisha P., et al. "Predicting depression using deep learning and ensemble algorithms on raw twitter data." International Journal of Electrical and Computer Engineering 10.4 (2020): 3751.

[7] Mulay, Akshada, Anagha Dhekne, Rasi Wani, Shivani Kadam, Pranjali Deshpande, and Pritish Deshpande. "Automatic Depression Level Detection Through Visual Input." In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pp. 19-22. IEEE, 2020.

[8] Zhu, Jing, Zihan Wang, Tao Gong, Shuai Zeng, Xiaowei Li, Bin Hu, Jianxiu Li, Shuting Sun, and Lan Zhang. "An improved classification model for depression detection using EEG and eye tracking data." *IEEE transactions on nanobioscience* 19, no. 3 (2020): 527-537.

[9] Gerych, Walter, Emmanuel Agu, and Elke Rundensteiner. "Classifying depression in imbalanced datasets using an autoencoder-based anomaly detection approach." In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pp. 124-127. IEEE, 2019.

[10] Wongkoblap, Akkapon, Miguel A. Vadillo, and Vasa Curcin. "Classifying depressed users with multiple instance learning from social network data." In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 436-436. IEEE, 2018.

[11] Lin, Huijie, et al. "Detecting stress based on social interactions in social networks." IEEE Transactions on Knowledge and Data Engineering 29.9 (2017): 1820-1833.

[12] Pampouchidou, A., O. Simantiraki, C-M. Vazakopoulou, C. Chatzaki, M. Pediaditis, A. Maridaki, K. Marias et al. "Facial geometry and speech analysis for depression detection." In Engineering in Medicine and Biology Society (EMBC), 39th Annual International Conference of the IEEE, pp. 1433-1436. IEEE, 2017.

[13] Deshpande, Mandar, and Vignesh Rao. "Depression detection using emotion artificial intelligence." In *2017 international conference on intelligent sustainable systems (iciss)*, pp. 858-862. IEEE, 2017.

[14] Jiang, Haihua, Bin Hu, Zhenyu Liu, Lihua Yan, Tianyang Wang, Fei Liu, Huanyu Kang, and Xiaoyu Li. "Investigation of different speech types and emotions for detecting depression using different classifiers." *Speech Communication* 90 (2017): 39-46.

[15] Hooda, Madhurima, Aashie Roy Saxena, and Babita Yadav. "A Study and Comparison of Prediction Algorithms for Depression Detection among Millennials: A Machine Learning Approach." In *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, pp. 779-783. IEEE, 2017.

[16] Rosa, Renata L., et al. "Monitoring system for potential users with depression using sentiment analysis." 2016 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2016.

[17] Sahla, K. S., and T. Senthil Kumar. "Classroom Teaching Assessment Based on Student Emotions."

In The International Symposium on Intelligent Systems Technologies and Applications, pp. 475-486. Springer International Publishing, 2016.

[18] Chen, Yan, et al. "Contrast pattern based collaborative behavior recommendation for life improvement." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2017.

[19] Majumder, Navonil, et al. "Deep learning-based document modeling for personality detection from text." IEEE Intelligent Systems 32.2 (2017): 74-79..

[20] Glavan, Ionela-Roxana, Andreea Mirica, and Bogdan Narcis Firtescu. "The Use of Social Media for Communication In Official Statistics at European Level." Romanian Statistical Review 64, no. 4 (2016): 37-48.

[21] Rosa, Renata L., Demsteneso Z. Rodriguez, and Graça Bressan. "Music recommendation system based on user's sentiments extracted from social networks." IEEE Transactions on Consumer Electronics 61.3 (2015): 359-367.

[22] Chen, Chih-Ming, Ming-Feng Tsai, Jen-Yu Liu, and Yi-Hsuan Yang. "Using emotional context from article for contextual music recommendation." In Proceedings of the 21st ACM international conference on Multimedia, pp. 649-652. 2013.