*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 8, Issue 2, pp 128-132, February-2024**

https://doi.org/10.47001/IRJIET/2024.802018

# Depression Prediction Using LSTM and SVM

**[1]Srinju M, [2]Prof. P. Gopika**

[1]PG Student, Dept. of Computer Science and Engineering, EASA College of Engineering and Technology, Tamilnadu, India
[2]Professor, Dept. of Computer Science and Engineering, EASA College of Engineering and Technology, Tamilnadu, India

***Abstract* - As there has been an increase in the number of mental illness cases, there is a need to curb this problem. Due to the complexity of traditional techniques based on clinical diagnosis, there should be an automated system for the detection and prevention of illness and hence there comes the need to develop a depression prediction system. The data is collected from twitter, followed by preprocessing and cleaning which includes removal of stop words, URL and HTML tags, expanding abbreviations etc. Following this, the process of feature extraction will be used to extract word count, pronouns, negations and other features from the comments made by the users following selection. Long Short Term Memory and Support Vector Machine classifiers are applied to obtain the results. Further, the tokenized words are embedded into a vector and passed to the encoder in and then to perform depression prediction, a classification layer is added on the top of the transformer output. Similarly the tokenized input is embedded into a vector of SVM, which is mapped to a class label, followed by classification into depressed and non-depressed classes. Both methods are compared for depression analysis. This helps in providing an early detection of depression in people. Social network and microblogging sites such as Twitter are widespread amongst all generations nowadays where people connect and share their feelings, emotions, pursuits etc. Depression, one of the most common mental disorder, is an acute state of sadness where person loses interest in all activities.**

***Keywords:*** Depression Detection, Natural Language Processing, Machine Learning, Social Media.

## I. INTRODUCTION

Mental illness is a very serious issue which has its impact all over the world. About thirteen percent of the world's total population is grieving from depression and anxiety. Nowadays, due to the stigma of mental illness, it becomes difficult for the affected ones to confide in their peers and hence such people tend to use social media as a way to express their feelings through posts, comments, etc and therefore, social media has become a way to analyze a mentally ill person. Depression prediction systems when used by users can help them in early detection of this illness which may help them in timely diagnosis and treatment which is very essential. The primary aim of this system is to predict if a user is depressed or not.

Despite the massive quantity of analysis on understanding individual moods together with depression, anxiety, and stress supported activity logs collected by pervasive computing devices like smartphones, foretelling depressed moods continues to be an open question. In this paper, we have proposed a depression analysis and suicidal ideation detection system, for predicting the suicidal acts based on the level of depression. Then, classification machine algorithms are used to train and classify it in five stages of depression depending on severity - Minimal or none, mild, moderate, moderately severe and severe.

The proliferations of internet and communication technologies, especially the online social networks have rejuvenated how people interact and communicate with each other electronically. The applications such as Facebook, Twitter, Instagram and alike not only host the written and multimedia contents but also offer their users to express their feelings, emotions and sentiments about a topic, subject or an issue online. On one hand, this is great for users of social networking site to openly and freely contribute and respond to any topic online; on the other hand, it creates opportunities for people working in the health sector to get insight of what might be happening at mental state of someone who reacted to a topic in a specific manner. In order to provide such insight, machine learning techniques could potentially offer some unique features that can assist. Depression as a disorder has been a great concern in our society and has been perpetually a hot topic for researchers in the world.

A traditional signature and experts rules-based techniques are no longer sufficient to detect ever-evolving threats. Machine Learning (ML) algorithms play a vital role in network cyber security.

Social network and sites such as Twitter are widespread amongst all generations nowadays where people connect and share their feelings, emotions, pursuits etc. Depression, one of the most common mental disorders, is an acute state of sadness where person loses interest in all activities. If not treated immediately this can result in dire consequences such as death. In this era of virtual world, people are more

 

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 2, pp 128-132, February-2024

https://doi.org/10.47001/IRJIET/2024.802018

comfortable in expressing their emotions in such sites as they have become a part and parcel of everyday lives. The research put forth thus, employs machine learning classifiers on the twitter data set to detect if a person's tweet indicates any sign of depression or not. Natural Language Processing as well as Opinion Mining are done as sub-domains of web content mining, with product input from customers 'polarities.

Depression and anxiety are psychiatric disorders that are observed in many areas of everyday life. For example, these disorders manifest themselves somewhat frequently in texts written by non diagnosed users in social media. However, detecting users with these conditions is not a straightforward task as they may not explicitly talk about their mental state, and if they do, contextual cues such as immediacy must be taken into account. When available, linguistic flags pointing to probable anxiety or depression could be used by medical experts to write better guidelines and treatments. In this paper, we develop a dataset designed to foster research in depression and anxiety detection in Twitter, framing the detection task as a binary tweet classification problem. We then apply state-of-the-art classification models to this dataset, providing a competitive set of baselines alongside qualitative error analysis. Our results show that language models perform reasonably well, and better than more traditional baselines. Nonetheless, there is clear room for improvement, particularly with unbalanced training sets and in cases where seemingly obvious linguistic cues (keywords) are used counter-intuitively.

In this project, from vectors trained on Twitter data, several word embeddings were generated for each tweet. The vectors and the averaged embedding generated have a dimensionality of 200. Prior to giving these embeddings as an input to the LSTM And SVM model, text preprocessing was carried out. In a particular series of the SVM experiment performed, only tokenization and lowercasing was carried out for all tweets, but in the subsequent one, specific preprocessing of tweets was performed. SVM model used a linear kernel.

- There are many experiences that can cause us depression, such losing a loved one, losing a job, getting a divorce and other tough situations can lead a person to feel blue, lonely and overwhelmed.
- It's normal to get these feelings when we feel stressed.
- Everybody have experimented sadness at times. However, depression is very different from this.
- Depression is a psychiatric disorder that needs to be addressed with medication.

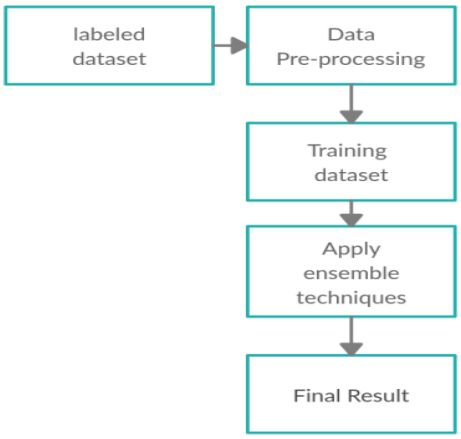


**Figure 1: The system model of our scheme**

## II. PROBLEM FORMULATION

Depression is the leading cause of disability worldwide. Almost 75% of people with mental disorders remain untreated in developing countries with almost 1 million people taking their lives each year. In addition, according to the World Health Organization (WHO), 1 in 13 globally suffers from anxiety. The WHO reports that anxiety disorders are the most common mental disorders worldwide with specific phobia, major depressive disorder and social phobia being the most common anxiety disorders.

### Depression on Social Media

Social media platforms are becoming an integral part of people's life. They reflect user's personal life. People like to share happiness, joy and sadness on social media. These platforms are used for researchers to identify the causes of depression and detect it.

I was reading an old article in the Time News Website about how Twitter knows when you're depressed and the possibility of creating an artificial intelligence model that can scan your Twitter feed and tell you if you're at risk for depression or receive notices from third parties, for instance, that warned you that you may want to seek help, just based on an automated scan of your tweets. I guess, that day has finally come. There are many ways for detecting sentiment in tweets.

Detecting earlier depression can be a huge step to address the mental illness and offer support to the people suffering from this terrible mental illness. Creating a model to detect depression in tweets

In Machine Learning, there are many ways for sentiment analysis such: decision-based systems, Bayesian classifiers, support vector machine, neural networks and sample-based methods.

 

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 2, pp 128-132, February-2024

https://doi.org/10.47001/IRJIET/2024.802018

After reading some papers about using different Machine Learning and artificial intelligence techniques to detect depression on Social Media, I decided to apply sentiment analysis through a powerful theorem from probability theory called Baye's Theorem. The model will be write in python and it will tell whether a given tweet is depressive or not.

TF-IDF stands for Term Frequency Inverse Document Frequency of records. It can be defined as the calculation of how relevant a word in a series or corpus is to a text. The meaning increases proportionally to the number of times in the text a word appears but is compensated by the word frequency in the corpus (data-set).

## III. BASIC IDEA OF OUR SCHEME

Term Frequency: In document d, the frequency represents the number of instances of a given word t. Therefore, we can see that it becomes more relevant when a word appears in the text, which is rational. Since the ordering of terms is not significant, we can use a vector to describe the text in the bag of term models. For each specific term in the paper, there is an entry with the value being the term frequency.

The weight of a term that occurs in a document is simply proportional to the term frequency. tf(t,d) = count of t in d / number of words in d Document Frequency: This tests the meaning of the text, which is very similar to TF, in the whole corpus collection. The only difference is that in document d, TF is the frequency counter for a term t, while df is the number of occurrences in the document set N of the term t. In other words, the number of papers in which the word is present is DF.

Inverse Document Frequency: Mainly, it tests how relevant the word is. The key aim of the search is to locate the appropriate records that fit the demand. Since tf considers all terms equally significant, it is therefore not only possible to use the term frequencies to measure the weight of the term in the paper. First, find the document frequency of a term t by counting the number of documents containing the term:

Term frequency is the number of instances of a term in a single document only; although the frequency of the document is the number of separate documents in which the term appears, it depends on the entire corpus. Now let's look at the definition of the frequency of the inverse paper. The IDF of the word is the number of documents in the corpus separated by the frequency of the text.

The more common word is supposed to be considered less significant, but the element (most definite integers) seems too harsh. We then take the logarithm (with base 2) of the inverse frequency of the paper. So the if of the term t becomes:

Computation: Tf-idf is one of the best metrics to determine how significant a term is to a text in a series or a corpus. tf-idf is a weighting system that assigns a weight to each word in a document based on its term frequency (tf) and the reciprocal document frequency (tf) (idf). The words with higher scores of weight are deemed to be more significant.

## IV. RELATED WORK

**1) Sharifa Alghowinem, Roland Goecke, Michael Wagner, Gordon Parker, Michael Breakspear, "Eye Movement Analysis For Depression Detection"**

Depression is a common and disabling mental health disorder, which impacts not only on the sufferer but also on their families, friends and the economy overall. Despite its high prevalence, current diagnosis relies almost exclusively on patient self-report and clinical opinion, leading to a number of subjective biases. Our aim is to develop an objective affective sensing system that supports clinicians in their diagnosis and monitoring of clinical depression.

**2) Stankevich Maxim, Nikolay Ignatiev, Ivan Smirnov, "Predicting Depression with Social Media Images".**

The study is focused on the task of depression detection by analyzing images related to social media users. We formed a dataset that consists of 485,121 images from profiles of 398 volunteers that provided access to their data in popular Russian-speaking social media Vkontakte. The results of the depression questionnaire were used to distinguish depression and control groups and set the binary classification task

**3) Rutuja K Bhoge, Snehal A Nagare, Swapanali P Mahajan, Prajakta S Kor, "Depression Detection by Analyzing Social Media Post of User".**

Nowadays the problem of early depression detection is one of the most important in the field of psychology .Mental health issues are widely accepted as one of the most prominent health challenges in the world, with over 300 million people currently suffering from depression alone.

**4) Hamad Zogan, Imran Razzak, Xianzhi Wang, Shoaib Jameel & Guandong Xu, "Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media".**

The ability to explain why the model produced results in such a way is an important problem, especially in the medical domain. Model explainability is important for building trust by providing insight into the model prediction. However, most existing machine learning methods provide no explainability, which is worrying.

 

ISSN (online): 2581-3048
Volume 8, Issue 2, pp 128-132, February-2024
https://doi.org/10.47001/IRJIET/2024.802018

**5) Chun Yueh Chiu •Hsien Yuan Lane, •Jia Ling Koh •Arbee L. P. Chen," Multimodal depression detection on instagram considering time interval of posts".**

Depression is a common and serious mental disorder that causes a person to have sad or hopeless feelings in his/her daily life. With the rapid development of social media, people tend to express their thoughts or emotions on the social platform. Different social platforms have various formats of data presentation.

## V. CONCLUSION

Often traditional survey-based questions fail to uncover the extent of users 'mental health depreciation. Nowadays, social media is a common platform, which people use to reach out. Therefore, the above proposed methodology cashes in this popularity of SNS and evaluates the depression levels of the users by employing natural language and machine learning techniques. In future, by collecting more data, more frequently with the aim of improving the accuracy of the work to give a better diagnosis. Such tools which can predict variations in person's mood can be an important method for both clinical observations and self-diagnosing. The method can be time consuming and hence steps must be taken in this regard to improve upon. The real-time application focuses on the analysis of sentimental and behavioural, that follows tweets are classified and summarizing on the twitter. Here, sentiment classifiers are dependent on user behaviour and Naive Bayes classification algorithm, which is used to classification of different emotions of the student's. The analysing and classifying student data is implemented. The developed is classifying tweets into a different, that is depends on the student emotions. The learning is not only involves the negative impacts for education fields about also involves the positive. Like, cultural activities and other extra work and experience not only stressful life in this environment, where it also enjoy the fun activities. This project defines a binary classification problem as identifying whether a person is depressed, based on his tweets and Twitter profile activity. Different machine learning algorithms are exploited and different feature datasets are explored. Many preprocessing steps are performed, including data preparation and aligning, data labeling, and feature extraction and selection.

## VI. FUTURE WORK

The experiences of learning and predicts the pressure with the social media in mining techniques are follows some future and enhancements are given to below, The minimum spanning tree for classification of the tweets to the problem of the students. The prediction of the pressure for each student is measured from peoples problem. The Latent Dirichlet Allocation checks the most relevant words and most important tweets will be done. The measure of fuzzy similarity that will be done in each topic and its count based on the analysis of performance is measured from the clustering and the classifications. The rate of false positive is the proportion of the cases of negatives that are differentiates as positive, the rate of true negative is the proportion of cases of negatives that are correctly. The rate of False Negative is the proportion of positives that incorrect as negative.

## REFERENCES

[1] N. A. Asad, M. A. Mahmud Pranto, S. Afreen and M. M. Islam, "Depression Detection by Analyzing Social Media Posts of User," 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2019, pp. 13- 17, doi: 10.1109/SPICSCON48833.2019.9065101.

[2] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot and Y. Kaewpitakkun, "Facebook Social Media for Depression Detection in the Thai Community," 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, 2018, pp. 1-6, doi: 10.1109/JCSSE.2018.8457362.

[3] P. Arora and P. Arora, "Mining Twitter Data for Depression Detection," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 186-189, doi: 10.1109/ICSC45622.2019.8938353.

[4] A.U. Hassan, J. Hussain, M. Hussain, M. Sadiq and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 138-140, doi: 10.1109/ICTC.2017.8190959.

[5] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 858-862, doi: 10.1109/ISS1.2017.8389299.

[6] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," 2019 IEEE Students Conference on Engineering and Systems (SCES), Allahabad, India, 2019, pp. 1-6, doi: 10.1109/SCES46477.2019.8977211.

[7] B. Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella and S. V. Alluri, "Real-time Acoustic based Depression Detection using Machine Learning Techniques," 2020 International Conference on

 

ISSN (online): 2581-3048
Volume 8, Issue 2, pp 128-132, February-2024
https://doi.org/10.47001/IRJIET/2024.802018

Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-6, doi: 10.1109/icETITE47903.2020.394.

[8] M. Hooda, A. R. Saxena, D. Madhulika and B. Yadav, "A Study and Comparison of Prediction Algorithms for Depression Detection among Millennials: A Machine Learning Approach," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, 2017, pp. 779-783, doi: 10.1109/CTCEEC.2017.8455078. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 01 | Jan 2021 www.irjet.net p-ISSN: 2395-0072 © 2021, IRJET | Impact Factor value: 7.529 | ISO 9001:2008 Certified Journal | Page 1839.

[9] A.Noureen, U. Qamar and M. Ali, "Semantic analysis of social media and associated psychotic behavior," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, 2017, pp. 1621-1630, doi: 10.1109/FSKD.2017.8393009.

[10] P. Gupta and B. Kaushik, "Suicidal Tendency on Social Media: A Case Study," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 273-276, doi: 10.1109/COMITCon.2019.8862236.

[11] D Nethra Pingala Suthishni, KS Senthil Kumar, "A review on machine learning based security approaches in intrusion detection system," 2022, 9th International Conference on Computing for Sustainable Global Development (INDIACom) pages -341-348 Publisher IEEE.

[12] Dr. S.Gnanapriya, D.Umanandhini "Feature Selection in Genetic Algorithm Based Sentiment Classification Used Five Layered Artificial Neural Network for Cross Domain Opinion Mining", International Research Journal of Innovations in Engineering and Technology (IRJIET), ISSN (online): 2581-3048, Volume 7, Issue 2, pp 80-85, February 2023 https://doi.org/10.47001/IRJIET/2023.702011.

**Citation of this Article:**

Srinju M, Prof. P. Gopika, "Depression Prediction Using LSTM and SVM" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 2, pp 128-132, February 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.802018

*******

 