



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Major Depression Disorder Detection Using Sentiment Analysis in Machine Learning

Ruthvik Gunti¹, Divya Jampani², Krishnaveni Gandham³, Ganga Reddy Baddam⁴, D.Himagiri⁵

UG Students, Dept. of C.S.E., J.B.I.E.T, Hyderabad, Telangana, India^{1,2,3,4}

Assistant Professor, Dept. of C.S.E., J.B.I.E.T, Hyderabad, Telangana, India⁵

ABSTRACT: Major Depression Disorder (MDD), also known simply as depression, is among the most prevalent psychiatric disorders globally. As described in the World Health Organization's Comprehensive Mental Health Action Plan 2013-2020, depression alone affects more than 300 million people worldwide and is one of the largest single causes of disability worldwide, particularly for women. Depression currently accounts for 4.3% of the global burden of disease, and it is expected to be the leading cause of disease burden in high-income countries by 2030. The Institute of Medicine Committee on the Prevention of Mental Disorders identified depression as the most preventable disorder, and several studies have demonstrated that early recognition and treatment of depression can improve the negative impacts of the disorder. Therefore, it is vital to provide an early identification of subjects suffering from depression to intervene as soon as possible and minimize the impact on public health by potentially reducing the escalation of the disease. However, provisions and services for the early detection and treatment of depression and other mental health disorder remain limited.

KEYWORDS: Tokenization, Stemming, Depression detection, Machine learning, Social media.

I. INTRODUCTION

Major Depression Disorder (MDD), also known simply as depression, is among the most prevalent psychiatric disorders globally. As described in the World Health Organization's Comprehensive Mental Health Action Plan 2013-2020, depression alone affects more than 300 million people worldwide and is one of the largest single causes of disability worldwide, particularly for women. Depression currently accounts for 4.3% of the global burden of disease, and it is expected to be the leading cause of disease burden in high-income countries by 2030.

The Institute of Medicine Committee on the Prevention of Mental Disorders identified depression as the most preventable disorder, and several studies have demonstrated that early recognition and treatment of depression can improve the negative impacts of the disorder. Therefore, it is vital to provide an early identification of subjects suffering from depression to intervene as soon as possible and minimize the impact on public health by potentially reducing the escalation of the disease. However, provisions and services for the early detection and treatment of depression and other mental health disorders remain limited. Although there are also some validated laboratory tests to diagnose depression, such as Beck Depression Inventory-II, Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale, Hospital Anxiety and Depression Scale, Patient Health Questionnaire-9, and Hamilton Rating Scale for Depression most diagnoses are formed on the basis of self- or family reports.

Social networks such as Twitter, Facebook, Instagram and Reddit, have become part of our daily lives as media through which to data banks for marketers and researchers who can analyze user metrics, shared content and related information to identify preferences and tastes as well as other attitudes and behaviors. In fact social networks have proved to be used by people to interact with peers because of their support and ability to understand someone's experience, while maintaining a comfortable emotional distance.

II. EXISTING SYSTEM

There are several ways to detect depression like Beck Depression Inventory-II, Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale, Hospital Anxiety and Depression Scale, Patient Health Questionnaire-9, and Hamilton Rating Scale for Depression. But most of the above methods were formed on the basis of self- or family reports. Visiting your regular doctor is also a good place to start. The existing methods heavily rely on self realization and self recognition of the symptoms. Unfortunately, about a half the people who have depression never get it diagnosed or treated. Recognizing symptoms is the key for the existing methods of depression detection. Some of

the symptoms include trouble concentrating, remembering details, making decisions, fatigue, feeling of guilt, worthlessness, helplessness, insomnia, restlessness, etc.

Disadvantages:

Recognizing symptoms on your own while going through such a mental state can be extremely difficult and sometimes impossible. As the existing methods of depression detection heavily rely on self identification of symptoms, several individuals may be unable to realize that they are exhibiting these symptoms and suffering from depression.

III.PROPOSED METHODOLOGY

As social networks have become an important aspect of each and everyone of us, one can infer many things through an individual’s social media profile. The proposed method uses the social network Twitter’s data and trains a model that classifies a given tweet as “Depressed” or “Not Depressed”. In machine learning, there are many ways for sentiment analysis such as decision-based systems, Bayesian Classifiers, support vector machines, neural networks and sample-based methods. The proposed system uses the Bayesian classifier. It applies sentiment analysis and uses the tf-idf and bow classifiers Bayes theorem.

Advantages:

- It uses experience to classify.
- The person suffering from depression need not work on identifying the symptoms by himself. His/her social media profile is enough to classify.
- The trained model can be deployed and integrated with any existing system.

IV.RESULTS

Test cases

S.NO	TEST CASE	Scenario	EXPECTED OUTPUT	ACTUAL OUTPUT	RESULT
1	Depressed Tweet	Feeling Down	True	True	True Positive
2	Depressed Tweet	Extreme Sadness,lackof energy, hopelessness	True	True	True Positive
3	Depressed Tweet	My depression will not let me work	True	True	True Positive
4	Non Depressed Tweet	Lovely how me and my lovely partner is talking about what we want	False	False	False Negative
5	Non Depressed Tweet	It is the little things that make me smile	False	False	False Negative
6	Non Depressed Tweet	Super happy that tomorrow is Friday	False	False	False Negative


```
[ ] sc_tf_idf = TweetClassifier(trainData, 'tf-idf')
sc_tf_idf.train()
preds_tf_idf = sc_tf_idf.predict(testData['message'])
metrics(testData['label'], preds_tf_idf)

[ ] Precision: 0.9629629629629629
Recall: 0.49056603773584906
F-score: 0.65
Accuracy: 0.8709677419354839
```

Fig.3.TF-IDF Performance

```
[ ] sc_bow = TweetClassifier(trainData, 'bow')
sc_bow.train()
preds_bow = sc_bow.predict(testData['message'])
metrics(testData['label'], preds_bow)

[ ] Precision: 0.9
Recall: 0.16981132075471697
F-score: 0.2857142857142857
Accuracy: 0.7926267281105991
```

Fig.4.Bag-of-Words Performance

```
▾ Depressive tweets

[ ] pm = process_message('Hi hello depression and anxiety are the worst')
sc_bow.classify(pm)

[ ] True

[ ] pm = process_message('My depression will not let me work out')
sc_bow.classify(pm)

[ ] True

[ ] pm = process_message('I am extremely depressed and I have anxiety')
sc_bow.classify(pm)

[ ] True
```

Fig.5.Bag-of-Words Prediction(Depressive)

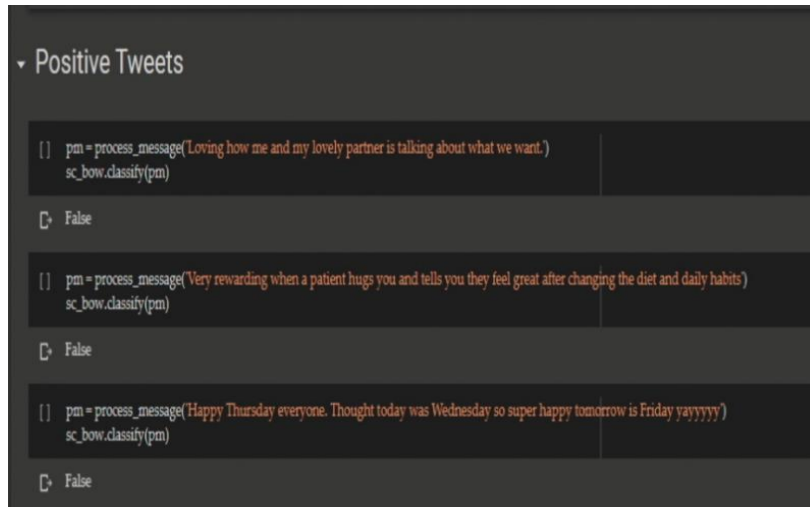


Fig.6.Bag-of-Words Prediction(Positive)

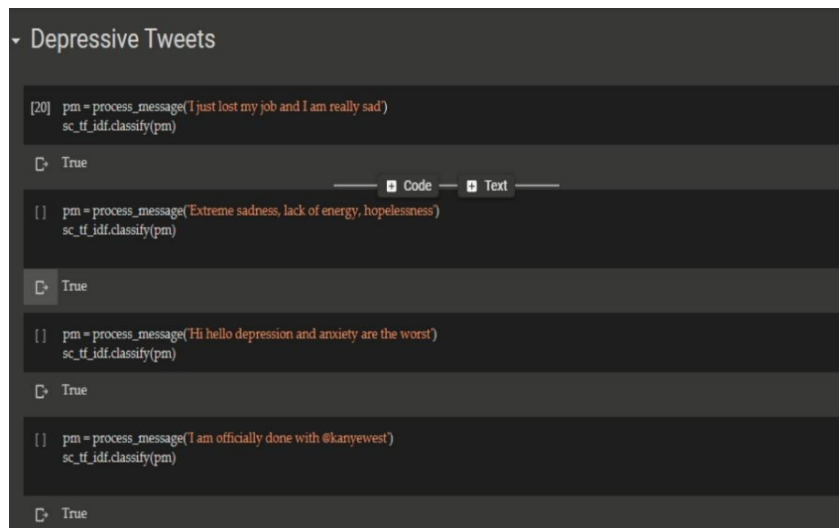


Fig.7.TF-IDF Prediction(Depressive)

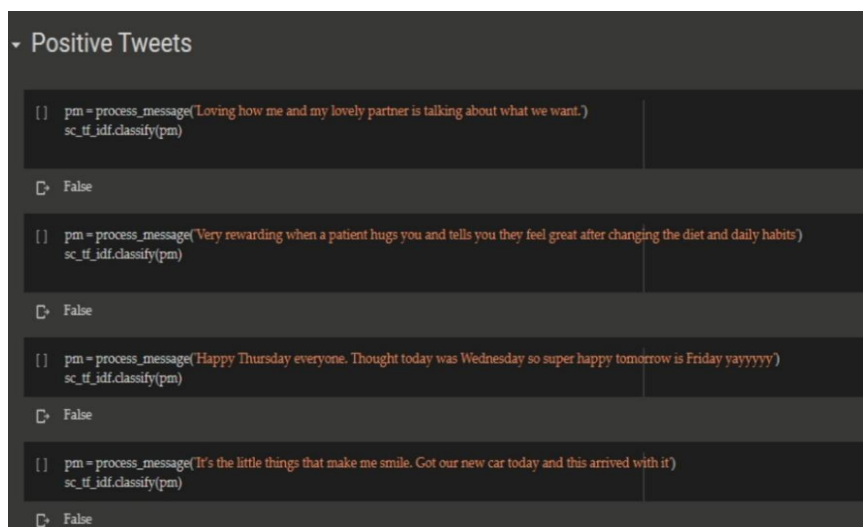


Fig.8.TF-IDF Prediction(Positive)



V.CONCLUSION

This presented a machine learning model to detect depressive tweets by classifying them as depressive or not. Two types of classifiers were used to assess their performances and select the one with better performance. The word cloud analysis helps us in understanding the type of words that have a significant impact in determining if a tweet is depressive or not. After assessing the performances of both the classifiers, the TF-IDF classifier proved to be giving better results when compared to the Bag-of-words classifier.

REFERENCES

- [1] Early Depression Detection: Social Network Analysis and Random Forest Techniques by Fidel Cacheda, Diego Fernandez, Francisco J Novoa and Victor Carneiro.
- [2] Monitoring Tweets for Depression to detect At-Risk Users by Zunaira Jamil, Diana Inkpen, Prasadith Buddhitha, Kenton White.
- [3] Identifying Depression on Twitter by Moin Nadeem, Mike Horn, Glenn Coppersmith, Johns Hopkins University and Dr. Sandip Sen.
- [4] Anxious Depression Prediction in Real-Time Social Data by Akshi Kumar, Aditi Sharma, Anshika Aurora.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details