SOCIAL NETWORK MENTAL DISORDERS DETECTION VIA ONLINE SOCIAL MEDIA MINING

Ms.S.Srividhya, (AP/CSE), Gobi Raj.B,
Jeevanantham.M, Prem Anand.R,
Dept of CSE, Erode Sengunthar Engineering College,
Erode, Tamil Nadu.

ABSTRACT

Social networking's rapid rise in popularity causes harmful utilization. Recently, it has been highlighted that there are a rising number of social network mental diseases (SNMDs), including cyber-relationship addiction, information overload, and net compulsion. Today, the majority of mental disorder symptoms are passively monitored, which delays clinical intervention. Users of this project contend that researching online social activity offers a chance to actively identify SNMDs at an early stage. Because mental state cannot be immediately inferred from online social activity logs, it is difficult to identify SNMDs. This fresh and creative method for detecting SNMDs does not rely on the subject's own disclosure of such mental variables via psychological questionnaires. The recordings of tweets are used to diagnose depression. after the correct pre-processing procedures. Additionally, this effort suggested a depression detection classification model that uses K-Nearest Neighbour and a Support Vector Machine-based classification to address the aforementioned issue. The language for project development is Python 3.7.

Keywords: *Social Media, Support Vector Machine, K- Nearest Neighbor, Mental Disorder*.

---

1. INTRODUCTION

Everyone has experienced sadness on occasion. Depression, though, is considerably different from this situation. Depression is a psychological disease that requires pharmacological treatment. the Depression's definition: The World in Data website states that depressive disorders typically come in varied degrees of severity. These disorders are classified as mild, moderate, or severe by the WHO's International Classification of Diseases (ICD-10). The definitions were adopted by the Institute for Health Metrics and Evaluation (IHME) and were broken down into three categories: moderate depression, persistent depression (dysthymia), and major depressive illness (severe). Some of the following symptoms are present in all forms of depression.:

(a) diminution of focus and attention
(b) decrease in confidence and self-esteem
(c) shame or unworthiness notions (even in amild type of episode)
(d) negative predictions about the future
(e) Self-harm or suicide thoughts or actions
(f) Sleep disturbance
(g) Appetite diminish

One of the main factors contributing to disability worldwide is depression. In underdeveloped nations, approximately 70% of those who suffer from mental illnesses go untreated, and every year, almost a million people commit suicide. Additionally, one in thirteen people worldwide, according to the World Health Organization (WHO), experience anxiety. According to the WHO, anxiety disorders are the most prevalent mental illnesses in the world, with social phobia, major depressive disorders, and particular phobias being the most prevalent types.

1.1 Depression on Social Media: Platforms for social media ingratiate themselves into peoples' daily lives. They represent our private lives. On social media, people want to express their happiness, excitement, and despair. Researchers use these platforms to find the root causes of depression. Twitter is aware of when you are suffering from or depressed, and there is a chance that it will develop an artificial intelligence model that will be able to scan your Twitter feeds and determine whether you are at risk of depression or will receive notifications from third parties, such as ones alerting you to the need for help.

Most depressed people desire to be alone, which makes them hopeful that there will be a cure for their condition. They believe that by talking about their emotions and illnesses, they might find relief and put an end to their worries and anxiety. It has been discovered that this social sharing and communication really makes people more worried and anxious. Facebook is one of the social networking sites with the quickest growth rate. More than two billion people have accounts on social networking websites worldwide. Additionally, India has more than 270 million Facebook users, making it the country with the largest Facebook audience..

Facebook is the website with the most internet access because the greatest number of users are used as a platform for socializing and presenting. Since they were first introduced, social networking sites (SNS) have garnered bold and significant attention. With the use of the Internet, social networking sites may be easily accessed and used for communication. The majority of people use the internet to gather information and rely on social networking sites to find the latest information and diverse viewpoints on a range of topics that involve sharing, emotions, subjectivity, assessments, evaluation techniques, observations, influences, point of views, and ideas using various typesof materials and marketing.

1.2 Social Network Sites

Instagram, Snapchat, and YouTube all lag behind Facebook in terms of daily access and user popularity. J. Clement discovered that people using Facebook most frequently are between the ages of 25 and 40. Facebook's monthly activity rate has been demonstrated to reflect how frequently users log on. This study has suggested scrutinizing the user's genuine behavior and way to bring about their differences/level of depression from the SNS users due to a number of issues. Results from 945 individuals (including 780 Facebook members and 165 non-members) revealed that Facebook members scored significantly higher than non-members on traits such stress depression, selfishness, lack of regard for the elderly, and opennessto others.

1.3 Depression

One of the most important challenges today is depression. The main cause of the global mental health problem is depression. Millions of people develop depression problems every year, yet only a small percentage of them take professional treatment seriously. Depression causes people to be unable to participate in their work with enthusiasm and to find less delight in routine everyday activities. When suffering from this mental condition, a person is afraid to expose themselves. In terms of both personal and public health, depression is on the rise. One of the best ways to solve the issue is to thoroughly investigate the individual's conduct. The information about these behaviors can be found by checking social networking sites about their regular activities. Social networking is a good way to learn about someone's feelings, behavior, emotions, and points of view, among other things. Theemotional level of the user can vary greatly.

1.4 Social Media

Additionally, compared to non-Facebook users, Facebook users appear to exhibit more depression symptoms. The bulk of Facebook users are college students. According to the study, those who have been diagnosed with mental illness choose to isolate themselves from others and turn to social media in order to express their thoughts and feelings in order to find relief. There are other techniques for analyzing tweet sentiment.

Early depression detection could be a key step in treating the condition and providing assistance to those who are suffering from horrible mental illnesses. developing a method to identify depression in tweets There are many techniques for sentiment analysis in machine learning, including Bayesian classifiers, neural networks, decision-based systems, sample-based techniques, and support vector machines. These writers attempted to use sentiment analysis using the potent Bayes Theorem from probability theory after reading several publications about various machine learning and artificial intelligence strategies for detecting depression on social media. The model, which was created in Python, determines whether or not a given tweet is depressing. K-Nearest Neighbor and Support Vector Machine-based categorization are both used in this research. Based on the gathered tweets, this projectsuggested a novel depression detection model.

2. METHODOLOGY

Users contend that this research offers the chance to actively identify SNMDs at an early stage by mining online social activity. Because the mental status cannot be immediately detected through online social activity logs, it is difficult to identify SNMDs. Here, we classify the records from the obtained tweet data set using the TF-IDF algorithm.

After classifying the given training and testing data records, the accuracy score is determined. This method, which is brand-new and ground-breaking in the field of SNMD identification, does not rely on respondents to psychological questionnaires self-reporting such mental elements. Following the appropriate preprocessing methods,depression detection using twitter records is carried out.

- More records in the dataset mean longer processingtimes.
- The deletion of Unicode characters is not done.
- Additionally, TF-IDF values for sentences withUnicode values are prepared.
- Accuracy score is not significantly higher for therecords in the dataset.

The suggested system employs every system approach currently in use. Unicode characters have been eliminated. This project suggested a depression detection classification model that uses K-Nearest Neighbor and a Support Vector Machine-based classification to handle the problem. Both KNN and SVM-based classification accuracy scores and confusion matrices are created, and it is discovered that they outperform the current system.

The following modules are present in the project.

a) Dataset collection
b) Find depression based on TF-IDFSVM classification
c) KNN classification SVM/KNN classification

2.1 DATASET COLLECTION

The tweet dataset from Twitter, which includes the characteristics message and label (1/0), is used in this module. Records with null values are removed during preprocessing. The paragraphs have been edited to remove Unicode characters. Detecting Depression in Tweets dataset download link (https://github.com/viritaromero/).

2.2 FIND DEPRESSION BASED ON TF-IDF

Message and label columns from sentiment tweet CSV (comma separated values) records are used in this module. The words in the text are separated, and term frequency and inverse document frequency are determined. The classification of testing data is then done based on the words (mainly occurred) in both training. The model is then forecasted using test data after being trained with training data. Of those, the majority of entries are accurately identified as having depression or not.

2.3 SVM CLASSIFICATION

20% of the data in the specified data set are used as test data in this module, and the remaining 80% are used as training data. Numbers are assigned to the text (categorical) columns. The model is then trained using training data, and forecaster using test data. With accuracy scores and confusion matrices, the majority of the records are categorized as having depression or not.

2.4 KNN CLASSIFICATION

In this module, 80% of the data in the given data set are used as training data, and 20% are used as test data. The text (categorical) columns are transformed into numbers. The model is then forecast using test data after being trained with training data. Most of the records are categorized as having depression or not using an accuracy score and confusion matrix.

3. EXPERIMENT RESULTS AND FINDINGS

The proposed system has following advantages.

a) Even if the dataset has more records, processing time will still be reduced.
b) Characters from Unicode are eliminated.
c) KNN classification is used to increase precision.
d) SVM will perform better in terms of accuracy.

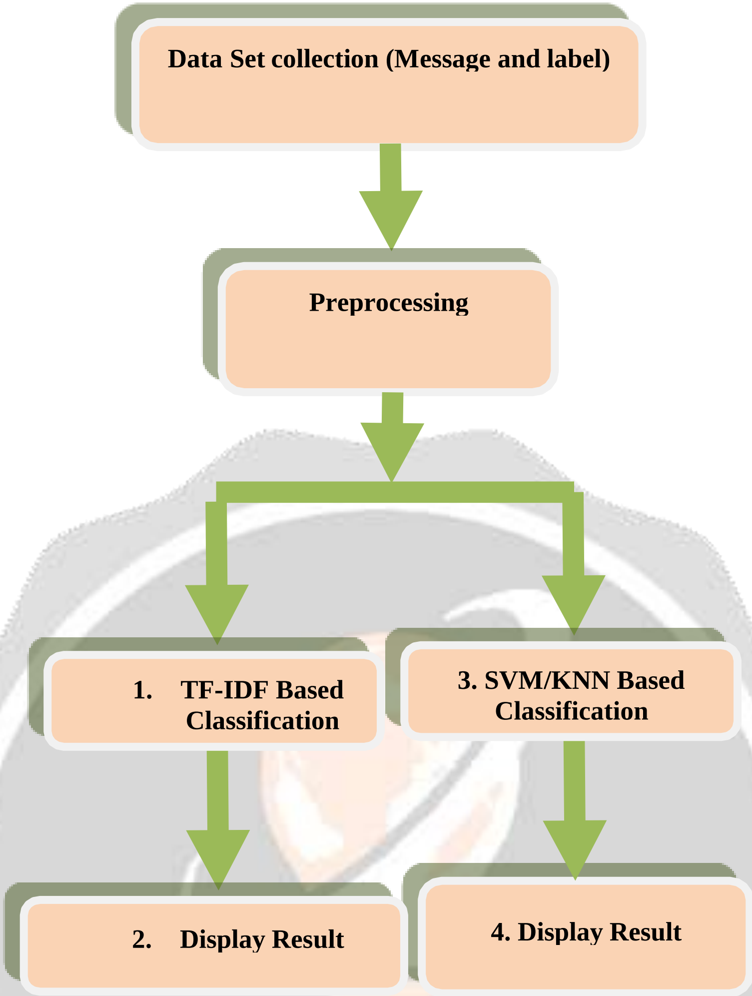


FIG 3.1 ARCHITECTURE DIAGRAM

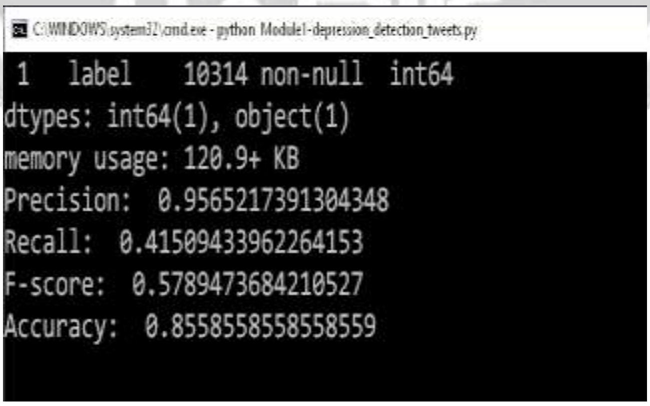
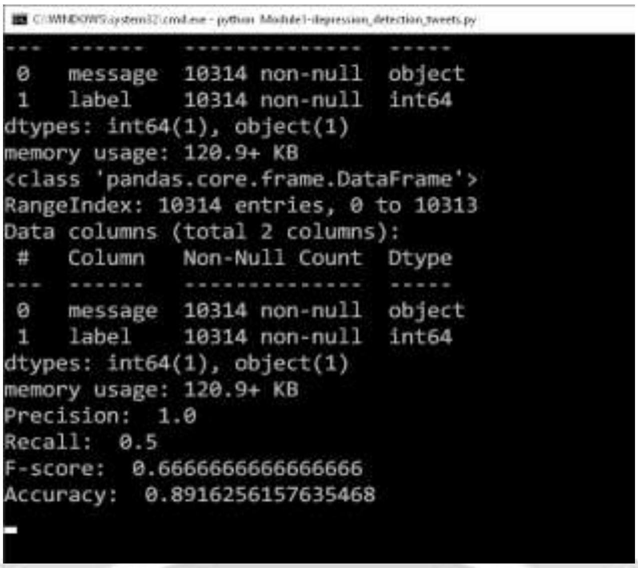


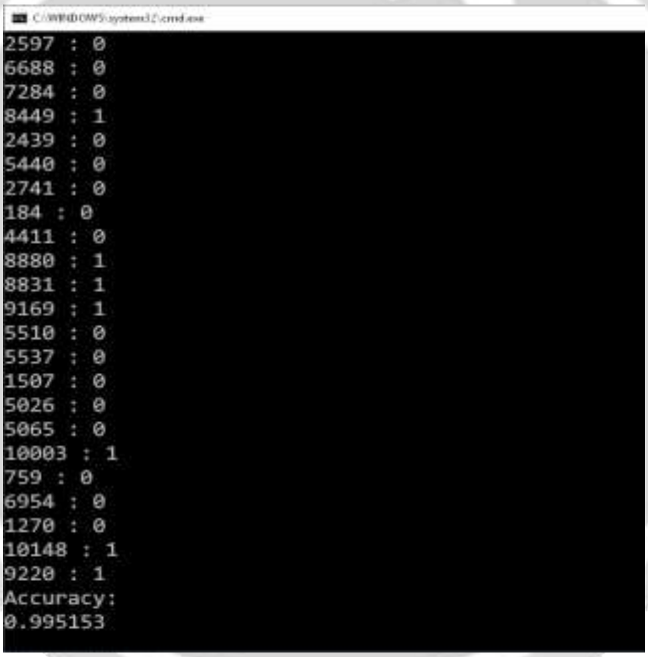
FIG 3.2 TF-IDF ACCURACY SCORE (1)



C:\WINDOWS\system32\cmd.exe - python Module1-depression_detection_tweets.py

```
---  ------   --------------  -----
 0   message  10314 non-null  object
 1   label    10314 non-null  int64
dtypes: int64(1), object(1)
memory usage: 120.9+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10314 entries, 0 to 10313
Data columns (total 2 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   message  10314 non-null  object
 1   label    10314 non-null  int64
dtypes: int64(1), object(1)
memory usage: 120.9+ KB
Precision:  1.0
Recall:  0.5
F-score:  0.6666666666666666
Accuracy:  0.8916256157635468
```

FIG 3.3 ACCURACY SCORE (2)



C:\WINDOWS\system32\cmd.exe

```
2597 : 0
6688 : 0
7284 : 0
8449 : 1
2439 : 0
5440 : 0
2741 : 0
184 : 0
4411 : 0
8880 : 1
8831 : 1
9169 : 1
5510 : 0
5537 : 0
1507 : 0
5026 : 0
5065 : 0
10003 : 1
759 : 0
6954 : 0
1270 : 0
10148 : 1
9220 : 1
Accuracy:
0.995153
```

FIG 3.4 SVM ACCURACY SCORE

```
Accuracy Score: 0.9958
              precision    recall  f1-score   supp

           0       1.00      1.00      1.00      2
           1       0.99      0.99      0.99

    accuracy                           1.00      3
   macro avg       0.99      1.00      0.99      3
weighted avg       1.00      1.00      1.00      3
```

FIG 3.5 KNN ACCURACY SCORE

4. CONCLUSION

In this study, depression detection is carried out in the existing system using word frequency inverse document frequency-based classification. The framework for new classification techniques that is proposed eliminates Unicode characters. For testing records, accuracy score, confusion matrix, and depression category are identified and displayed. K-Nearest Neighbor and a Support Vector Machine-based classification are utilized in this case to solve the classification problem. A depression detection model was presented in this project based on the tweets gathered using the TF-IDF, SVM, and KNN approaches. The suggested approach employs KNN and SVM classification to produce improved classification outcomes. The Python Language 3.7is being used to create the project.

5. REFERENCES

[1] T. Huijie Lin, "Psychological stress detection from cross- media microblog data using deep sparse neural network, conference: 2014 ieee international conference on multimedia and expo (icme).

[2] Andrey Bogomolov , "Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits" October 2014. http://dx.doi.org/10.1145/2647868.2654933.

[3] G. Bauer and P. Lukowicz. Can smartphones detect stress-related changes in the behaviour of individuals? In Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE. InternationalConference on, pages 423–426. IEEE, 2012.

[4] A. Bogomolov, B. Lepri, and F. Pianesi. Happiness recognition from mobile phone data. Proceedings of the 2013International Conference on Social Computing (SocialCom 2013), pages 790–795, 2013.

[5] N. Bolger and E. A. Schilling. Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. Journal of Personality, 59(3):355–386, 1991.

[6] G. E. P. Box and D. R. Cox. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B (Methodological), 26(2):211–252, 1964.

[7] L. Breiman. Random forests. Mach. Learn., 45(1):5–32, Oct. 2001.

[8] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Mining large-scale smartphone data for personality studies. PersonalUbiquitous Comput., 17(3):433–450, Mar. 2013.

[9] J. Cohen. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1):37–46, Apr. 1960.