An Approach to Identify the Depressed People using Tweets

[1]D Venkataraman and [2]Reshma B Nair

ABSTRACT

Depression exists as one of the most common form of psychological disorder. It is seen in most of the individuals at some point of time in life. People who are depressed will feel sad, anxious, hopeless, sleepless etc. It creates an impact on both physical and mental health. The outcomes of depression can turn out to be severe if it is left untreated. It can lead an individual to risky behaviour such as drug or alcohol addiction. It can also ruin the relationships, affect the life in the work environment and also make it difficult for an individual to return back to the usual lifestyle. In this paper we provide a brief study on the data that is obtained from twitter and analyze whether the person is depressed or not.

Index Terms: Depression, outcomes, serious, twitter.

I. INTRODUCTION

Depression is a common but serious illness. Depression affects how people feel about themselves. They start losing interest in work and things they normally enjoy. They may lack energy, have difficulty sleeping or sleep more than usual. Some people find it hard to concentrate. Depression makes life more difficult to manage from day to day. Not every person who is experiencing depression will have these symptoms. Depression is a syndrome of many symptoms and it is not just feeling sad. The people with depression generally don't seek medication. But recovery can be done when treatment is provided at an earlier stage. The consequences of depression are long lasting and wide ranging.

II. CAUSES OF DEPRESSION

A number of things or reasons can be associated to lead a person to depression. It might to a recent event that has happened or due to some personal problems. Some of the common reasons for depression are as follows [1].

(A) Family situation

When the family has a poor financial background or when the family cannot financially support a person for studies then it can lead an individual to depression. But it is not assured that all those suffering from financial difficulties will suffer from depression.

(B) Individual personalities

A person may feel depressed depending upon his or her personalities. If a person has a lot of negative thoughts, sensitive to criticisms, or generally have a tendency to worry a lot then they are likely to get depressed.

---

[*] Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore.

[**] Amrita Vishwa Vidyapeetham, Amrita University, India.

E-*mail: [1]d_venkat@amrita.edu.in, [2]reshma4nair@gmail.com*

(C) Drug and alcohol use

The usage of drug and alcohol can lead to depression and also can be an outcome of depression. Due to a stressful life, nearly half of the population is addicted to alcohol.

(D) Work environment

Due to the pressure in the work environment or if a person is dissatisfied with the job, then he or she is likely to get depressed.

A combination of different factors can lead a person to depression and the factors differ from one individual to another. The above mentioned are some of the common factors of depression and there exists several other factors that leads to depression [1].

III. EXISTING WORK

Priya Nambisan *et al.* [3] states that nowadays social media has turned out to be general platform for sharing one's thoughts and ideas. The tweets are collected and then analyzed to detect the depressed people and also reveal their behaviour

It is analyzed that critical significance of repetitive thoughts and ruminating behaviour is seen with people who have tweeted about depression.

The analysis is performed between two groups. One group consisting of depressed people and the other group consisting of people who have not tweeted about depression and it is known as the control group. The findings of the study includes

1. Users with depression tweet more amount of time regarding symptoms of depression than the people without depression.
2. The depressed group more symptoms about depression than the control group
3. The symptomatic characteristics are associated more with the depressed group.

It is found that the depressed people exhibit the same kind of ruminating behaviour such as sleep, pain, suicide etc. Hence this data will help to identify the depressed people and also use the tweets in a much efficient and effective way.

The system proposed in [4] is used to do the real-time analysis of variations in emotions. The data for analysis is collected from twitter. The regional variation in emotions is analyzed rather than the individual variations. The analysis is done in three steps

1. Information on diurnal variations in mood
2. Link data with prevalence data from countries globally to determine if patterns are consistent with these rates
3. Whether "events of significance" are associated with changes in expressed emotion.

12 weeks of emotional tweets were collected and analyzed in order to find out the global variations in emotions

Some of the limitations are

1. The period of data collection is relatively low
2. The individual emotional tweets is not analyzed

Lina Alfantoukh *et al.* [5] states that social network has turned out to be a common and one of the most convenient platforms for sharing one's data. Hence the data that is obtained from the social media will help to perform analysis with greater accuracy.

The traditional method of collecting data through questionnaires and surveys will not provide the data for analysis in real-time.

The main contributions of this work include:

1. Collect Data from online social networking, specifically from Twitter.

2. Demonstrate different approaches for collecting the data from Twitter.

3. Evaluate the two approaches and examine the possibility of combining two approaches.

The data is collected by two methods

- Collecting by group
- Collecting by filtering

The combination of both the methods can be used in order to improve the efficiency.

IV. PROPOSED SYSTEM

The architectural representation gives an idea on how the data collected from twitter and identification of depressed users.

The data collected from twitter API can be stored on database. The stop word removal has to be done in order to segregate only the required information. After performing the required modifications the necessary analysis can be provided to find out the depressed users.
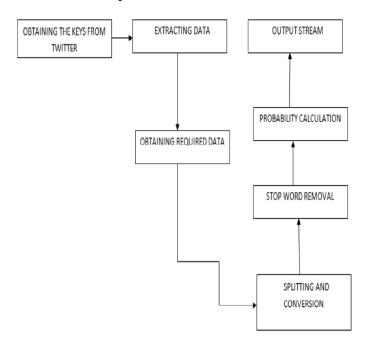

Figure 1: Steps in analysis of twitter data

V. DATA COLLECTION FROM TWITTER

In order to access the twitter data, the following steps has to be followed.

- Register in twitter to access API
- Get the consumer token and consumer secret
- Enter the credentials

- Get the access token and access secret
- Request the tweets using the access token and secret.

(A) Tweets of depressed users

The real time tweets of depressed people can be obtained by using python code. Python is a widely used, general purpose high level language. Its syntax allows programmers to write in fewer lines of code than by using C++ or Java.
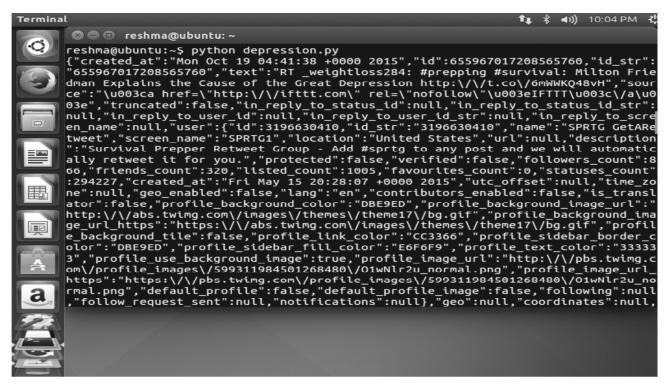
(B) Twitter data for keyword depression

Figure 2: Twitter data for keyword depression

Figure 3: Twitter data for keyword depression

The above unstructured tweets can be converted into structured tweets and then map reduced in order to segregate the tweets containing the depressed information alone.

VI. SPLITTING THE TWEET AND CONVERSION TO LOWER CASE

The entire tweet is being divided into single word for analyzing. From that the stop word removal is then performed.

Figure 4: Splitting the tweet

The conversion to lower case is done so that when the keyword is compared, it has to match with the stop words text file as well as the predefined set of positive and negative words and also the additional white spaces is being removed.

Figure 5: Conversion to lower case

VII. STOP WORD REMOVAL

The stop words removal is done in order to remove the unwanted text from the tweet. These are the text that are of less importance in the tweet and they do not convey much meaning. An example for stop words is a, an, about etc… the stop words depends upon the context.

Figure 6: Stop word removal

VIII. CALCULATING THE PROBABILITY

A predefined set of positive and negative keywords is used. The keywords of the tweet are compared with these keywords and if they are found to be positive or negative, then their count is being incremented.
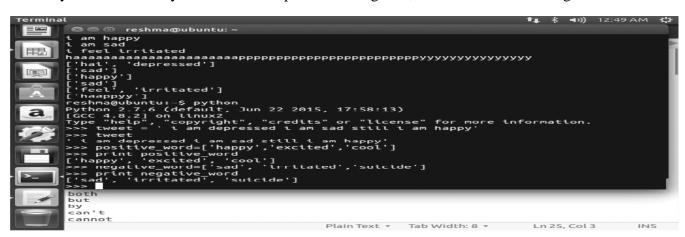

Figure 7: Positive and negative words

The final output is that it determines whether the person who has tweeted is a depressed person or not. This is found from the probability that is calculated using naïve bayes classifier. If the probability of the negative count is greater, then the person is a depressed person else the person is not a depressed person. The probability is found as follows

Negative_counter = total number of negative words/ total number of words

Positive_counter = total number of positive words/total number of words

Thus it finally determines whether the person is depressed or not.

IX. CONCLUSION

Depression can affect an individual to a greater extent if it is left untreated. The reason for depression also changes from one person to another. It can be due to the family situations, the work environment, or due to the drug and the alcohol use etc.
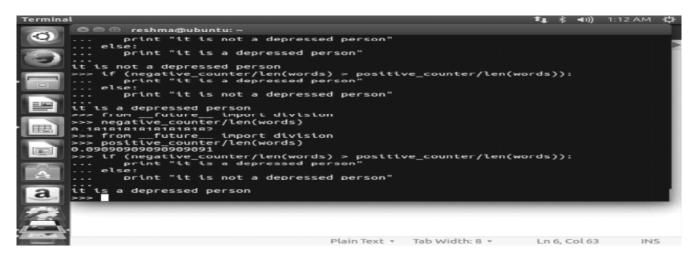
Figure 8: Final Analysis

Therefore in order to recover a person from depression, it has to be detected at an earlier stage and proper treatment has to be provided. The probability determined will specify whether the person is depressed or not.

REFERENCES

[1] https://www.beyondblue.org.au/the-facts/depression/what-causes-depression

[2] http://www.helpguide.org/articles/depression/depression-signs-and-symptoms.htm

[3] Priya Nambisan, Zhihui Luo, Akshat Kapoor, Timothy B Patrick , Ron A Cisler . "Social Media, Big Data and Public Health Informatics: Ruminating behavior of depression revealed through Twitter" 48th Hawaii International Conference on System Sciences, 2015.

[4] Mark E. Larsen, Tjeerd W. Boonstra, Philip J. Batterham, Bridianne O'Dea, Cecile Paris, and Helen Christensen. "We Feel: Mapping Emotion on Twitter" IEEE Journal of biomedical and health informatics, Vol. 19, no. 4, july 2015.

[5] Lina Alfantoukh, Arjan Durresi. "Techniques for Collecting data in Social Networks", International Conference on Network-Based Information Systems, 2014.

[6] Helen Christensen , Philip J. Batterham , Bridianne O'Dea. "E-Health Interventions for Suicide Prevention" *International Journal of Environmental Research and Public Health,* Vol.11, pp.8193-8212, 2014.

[7] Ritter, A., Cherry, C., Dolan, B. "Unsupervised modeling of Twitter conversations". In Proceedings of HLTNAACL, 2010.