Detection of Depression in Social Media via Twitter Using Machine learning Approach

Sangeeta R.Kamite
Department of computer science & Engineering
P.E.S.College of Engineering, Aurangabad
Maharashtra, India
Sangeetakamite94@gmail.com

Dr. V. B. Kamble
Department of computer science & Engineering
P.E.S.College of Engineering, Aurangabad
Maharashtra, India
vbkamble@yahoo.com

Abstract— **In recent year's social media platforms plays a crucial role in our daily life. The content which share by users on social media replicate the user's internal life. Folks wish to share daily life incidence like happy moments, joyful memories, and unhappiness on social media. These social network sites widely used for scientist to spot the principles of depression and observe them. Social media provides a new chance to remodel early depression intervention services mostly in young adults. The aim is to develop a project in which system is capable and analyze the syntactical markers related to onset and perpetual symptoms of depression. By using this system it is possible to develop an algorithm which predicts depression in effective manner to put out the approach that syntactical markers used in tweets are often utilized to frame statistical model which will effectively find and also predict depression in such a way that can integrate and extend ancient method of identification.**

Keywords— Social media, Depression, Twitter, Machine learning

I. INTRODUCTION

It is possible to construct a deep learning model which can give overview of persons mental health far early in time as compare to the traditional approaches by analyzing statistical markers in social media post .Lots of people nearly about millions are tackling with the problem of depression and only some of them receive an competent treatment. Depression is a preeminent explanation for incapacity across the world and approximately 810,000 individuals in a year committed suicide which is one of the major reasons for death among 15-29 year olds. Resultant treatment for depression square measure usually delayed, imprecise or lost entirely [4]. Social media provide new chance to remodel early depression intercession services mostly in 15-29 year olds adult. every second countless number of tweets are posted on twitter which approximately cross over 400,000 tweets sent per minute, nearly about two hundreds billion of tweets posted annually and approx five hundred millions per day[1]. 75% of population uses any of the variety of the social networking apps like Facebook and twitter because depression is associate degree health problem therefore typically needs the awareness about symptoms. Social media post content give an expensive supply of information and data that may be accustomed train associate degree economical system.

A. Depression Detection

The aim is to develop a project in which system is a able to capturing and analyze the syntactical markers related to onset and perpetual symptoms of depression. [6]. Proceed that the concept of syntactical markers present in posted tweets is wont to developed applied math models which may sight and even predict depression symptoms in a ways can competent and extend ancient methods to designation. To build associate degree algorithmic rule which will analyze Tweets reflect

Self-assessed depressive options can make it doable for people to investigate social media posts for syntactical identification that symptoms deteriorating psychological state way before ancient methods presently use [5]. Identify syntactical markers in social media posts permits for a abidance that may blandish ancient methods and give far earlier prevision of depressive markers. While several existing studies of depression-related social media posts used Facebook, some centered on Twitter data. However, there are certain differences between the 2 platforms that cause unambiguously completely contrary content, and Twitter data is unambiguously worthy [3]. however Facebook users often post below their real names to friends and members of the family, Twitter users usually use real name and are additional doubtless to be linked with users they need ne'er met. This permits for an additional anonymous means that of communication, which can offer a less fake account of Associate in Nursing individual's thoughts and experiences. Twitter conjointly have a broad population of users agency might be usually doubtless to hold off in mental health connected analysis additionally as a unsubtle contrast of publically out there info, providing the simplest way to beat a number of the constraints of ancient strategies of information assortment. Changes in language and activity are systematically correlate to the difference in behavior on social media. Emotional language and syntactical options utilized in social media posts are well-tried to impression that characterized major depression. Users mostly use social networking podiums such as Twitter express to their feelings and views in written kind. Data content of posts on twitter square measure significantly helpful for analysis as a result of their most typically created within the course of daily life activities and special events. This gives an expensive and coherent suggestion that for getting activity to individuals thinking, mood, socialization, communication, activities.

978-1-7281-5970-6/20/$31.00 ©2020 IEEE

II. LITERATURE PROSPECT

The reluctance of syntactical analysis within the obstruction of mental state can't be immoderate. By researching a person's words, it shows a transparent and valuable window into person status [3]. Even the best research of social networking will give North American nation with major access into people views and feelings and cause considerably larger understanding and treatment of mental state.

In[14] Munmun De Choudhury et al.,(2013) tends to use social networking sites as a calculative tool of depression in number of people for measuring depression by using SVM classifier method. In[13] Munmun De Choudhury et al.,(2014) tends to Characterizing and Predicting Postpartum Depression from Shared Facebook Data for Detect, Characterize and Predict Postpartum Depression by using patient health questionnaire. In[12]Peter Burmap et al.,(2015) tends to Machine Classification and Analysis of Suicide- Related Communication on Twitter by using principal component analysis. In[11]Elvis Saravia et.al.,(2016) tends to detection of Mental illness and analysis via social media to Predicting Depression by using random forest classification. In[10]Keumhee et.al.,(2016) tends to Identifying Depressive users in Twitter using Multimodal analysis which Extracts Tweets from Twitter that indicate Depression..By using K-means clustering latent fusion. By using Naïve Baye's classifiers In[8]Adrian Benton et,al.,(2017)tends to use Multitask Learning for Mental Health Conditions with Limited Social Media data, for Predicting Depression by using Multitask Learning approach .

A. *Naive baye's algorithm*

It is a classifying technique based on Baye's Theorem consist an assumption of independence among predictors. In simple terms, a Naive Baye's classifier states that the there is no interconnection between two different features.

III. PROPOSED SYSTEM

We build the model for detecting depression via twitter by using machine learning approach. In which statistical model is train to execute the particular coursework without using definite instructions instead use the patterns and interference, similar to the Artificial Intelligence. Machine learning contains various algorithms which is used to develop a mathematical model which is work on the sample data provide called as training data which used to decisions without explicitly execute the coursework. There are various applications in which machine learning algorithm is used such as service personalization, Natural language processing, sentiment analysis, computer vision, speech recognition and many more. In the proposed model we used machine learning algorithm i.e Random Forest algorithm for the detection of depression from twitter data. Random Forest algorithm is a supervised machine learning algorithm as the name indicates it builds the forest of number of trees. As much the maximum number of tress in the forest gives the maximum accuracy result.

A. *Why Random Forest Algorithm*

1) Missing value exclusively handle by the random forest algorithm.
2) Also able to construct the random forest classifier for categorical values also
3) Random forest algorithm reduces the problem of over fitting.
4) Classification and regression task both can be solve by using same random forest algorithm.
5) Feature engineering mainly used random forest algorithm.
6) For identification of the most important features from available features of the training dataset.

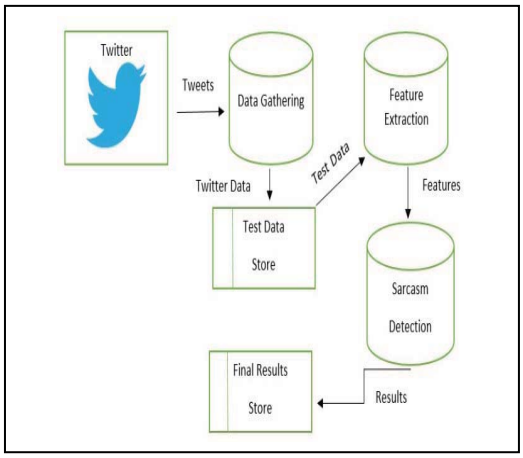

Fig 1.Architecture of Proposed Model

B. *METHODOLOGY.*

1) Fig.1 explicates the architecture of proposed model.
2) In proposed model we gather data related to depression level by using twitter users social activity
3) To this we analyze the online activity of the twitter user when person make tweets on twitter.
4) Data collected from user's online social activity and store in a test data store where data is store for the further processing.
5) After storing the data next step is to test the data in which process of feature extraction is performed.
6) Feature extraction efficaciously turn down the data which must be processed, while original data set is still accurately and completely describe.
7) Properly optimizes feature extraction is the key to effective model construction.
8) After feature extraction sarcasm detection is performed and finally result is stored.

Below fig.2 explicate the working of the proposed model in which various operations are performed such as data collection, data preprocessing, and sentiment analysis.

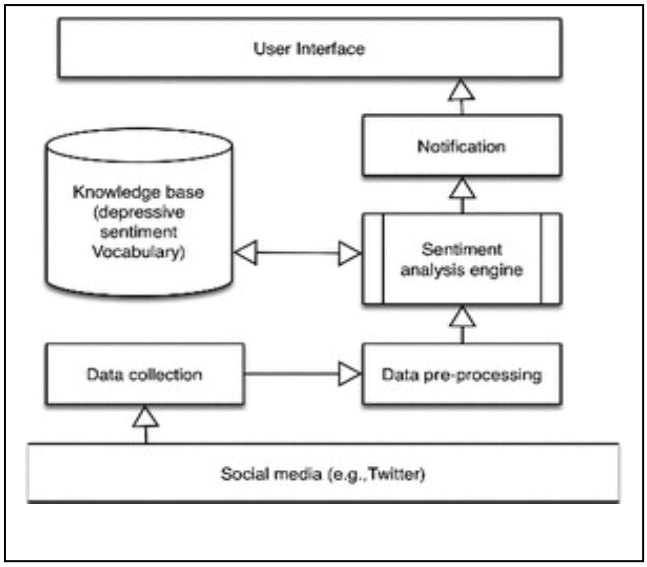

Fig.2 Working of the Proposed Model

Following table contains some examples of positive tweets and depressive tweets.

TABLE I. EXAMPLE OF POSITIVE AND DEPRESSIVE TWEETS

Positive tweets	Depressive Tweets
Always bring your own sunshine	Depression sucks. I'm in the middle of it
Sharing happiness daily	I hate myself too
Begin each day with a smile and a grateful heart.	It was my fault I let you control me..
Have a blessed day	It's so loud in my head
Find people who will make you better	I am not depressed ,everyone just pretending to be happy

C. *Result*

Experimental outcome

TABLE II. COMPARISON OF EXPERIMENTAL OUTCOME OF EXISTING AND PROPOSED SYSTEM.

	Existing system	Proposed system
Algorithm used	Naive bye's algorithm	Random forest algorithm
F1 Score (Accuracy)	0.9487394953	0.998949579838

Test accuracy is measure by the F1 score. Our model gives the better result when the F1 score is maximum. Mathematically, it can be expressed as:

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}} \qquad (1)$$

F1 Measure calculates the relativity between recall and precision.

Precision: calculate by dividing number of actual positive results by the number of positive tweets which is classifier calculated.

Recall: calculate by dividing number of actual positive tweets by the number of all related samples (all samples that should have been identified as positive)

IV. FUTURE SCOPE

In this paper as we are working on predicting tweets sentiments but it can be more accurate and better with the large data. The bit hard part was searching the annotated dataset and cleaning it for the analysis.

In future, I want use deep learning techniques for the more accurate result. I also want to analyze the depression level of the individual by predicting the depressive as well as the positive or neutral tweets of the user. I want to use multiple factors to analyze in order to make more accurate results.

V. CONCLUSIONS

The social media based mental condition of an individual can be analyzes by testing the relation between depression and specific syntactical markers through an algorithmic program. Which can help to know the sign of depression taking place with individual? The depressive behavior or condition can be diagnosed or find out through the use of syntactical markers as a tool by which the depressive character can be easily predict by the tweets even without the use of advanced model. This can be finding out just by some collective, clear and visual analysis of tweets. This project expands the scope of diagnosis of depression from social media.

Acknowledgement

I heartily thanks to the publisher, researcher and professor for guiding me for the project. The college authority also helped me by providing me a great support and the entire needed infrastructure which was required for the guidance for the project. Also a great thanks to my loving family and friends for supporting and believing in me.

References

[1] Mowery D, Bryan C, Conway M Feature ," Studies to Inform the: Classification of Depressive Symptoms from Twitter Data for Population Health.", 2017 arXiv:1701.08229.

[2] G, Dredze M, Harman,'' C: Quantifying mental health signals in Twitter. Workshop on Computational Linguistics and Clinical Psychology." ACL; 2014:51-60.

[3] Zigmond AS, Snaith RP," The hospital anxity and depression scale.ActaPsychiatrscand 1983,"67:361-370.

[4] Zhung, W. W. K.," A self-rating depression scale.Archives of General Psychiatry 12," 1 (1965), 63–70

[5] Mohr DC, Zhang M, Schueller S," Personal sensing: understanding mental health using ubiquitous sensors and machine learning." Ann Rev ClinPsychol 2017, 13.

[6] Tsugawa, S., Mogi, Y., Kikuchi, Y., Kishino, F., Fujita,K., Itoh, Y., and Ohsaki, H,"On estimating depressivetendency of twitter users from their tweet data." InProceedings of the 2nd International Workshop onAmbient Information Technologies (AMBIT'12) (Mar.2013), 29–32.

[7] Yoshihiko Suhara, YinzhanXu, Alex Sandy Pentland," Deep Mood: Forcasting depressed mood based on selfreported histories via Recurrent Neural Networks," International World Wide Web Committee, April 2017.

[8] Adrian Benton, Margaret Mitchell, Dirk Hovy," Multitask Learning for mental health condition with limited social media," Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol 1, PP 152-162,April 2017.

[9] Maryam Mohammed Aldarwish, HafizFarooq Ahmed," Predicting depression levels using social media posts," 13th International Symposium on Autonomous Decentralized Systems, 2017

[10] Keumhee Kang, Chanhee-Yoon, Eun Yi Kim(2016). "Identifying Depressive Users in Twitter Using Multimodal Analysis," BigComp, 2016

[11] ElvisSaravia, Chun Hao Chang, Renaud Jollet De Lorenzo, Yi- Shin Chen,"MIDAS: Mental Illness Detection and Analysis via Social Media," International Conference on Advances in Social Networks Analysis and Mining,Vol, August 2016.

[12] Peter Burmap, Gualtiero Colombo, Jonathan Scourfield, "Machine Classification and Analysis of Suicide- Related Communication on Twitter," 2015.

[13] Munmun De Choudhury, Scott Counts, Eric J. Horvitz, Aaron Hoff," Characterizing and predicting postpartum depression from shared Facebook data," Social Technologies and Well-Being, Feb 2014.

[14] Munmun De Choudhury, Scott Counts, Eric Horvitz, "Social Media as a Measurement Tool of Depression in Populations, "WebSci, May 2013.