Computers in Human Behavior Reports 2 (2020) 100036

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-reports

An analysis of mental health of social media users using unsupervised approach

Deepali Joshi [a,*], Dr.Manasi Patwardhan [b]

[a] Department of Technology, University of Pune, Ganeshkhind, Pune, India
[b] Tata Research Development and Design Centre, Hadapsar Industrial Estate, Hadapsar, Pune, India

ARTICLE INFO

Keywords:
Mental health of social media users
Psychological analysis of tweets
Analysis of behavior change
Social media users' posts reveal mental health
Tweets and posts becoming tools for psychiatrists

ABSTRACT

With the shift of population toward digital lifestyle, it is becoming increasingly far easier to express opinions, behavior and mindset online – instantly and openly due to majorly following three very important reasons: firstly the online disinhibition effect/anonymity, second is the psychological distance and the third the emotional contagion (Lieberman and Schroeder, 2020). The myriads of data originating from social media platforms provide a major insight into the life of people. This insight underlines the mental health and emotional conditions of the users, chiefly the young population. The challenge is to identify the users who are showing signs of succumbing to mental illness at its onset (prodrome period). In this proposed method, we applied unsupervised algorithms on the data, signaling behavior change for psychological analysis and identified the probability of users showing at-risk behavior. By at-risk behavior, we mean users who are on the verge of acquiring some mental illness. In this study, we analyzed posts and tweets from the social media platform namely *Twitter* and developed an unsupervised model to classify users based on the scale of change in their behavior. Our model has achieved 76.12% accuracy.

1. Introduction

With the increase in the number of social media providers and decrease in charges of the internet services, the digital media is permeating almost everyone's life, directly or indirectly, chiefly the young population and the senior citizens (Luo and Hancock, 2020). The number of users on Twitter alone has reached to 17 Million by July 2020 in India (https://www.statista.com/, 2426). According to the BBC news, US has made it mandatory to disclose the social media accounts of all new visa applicants from Mar 29, 2018. This shows the importance how far the social media has spread and at the same time has assumed a role to gauge the behavior and state of mind of the users (Bauer, 2017) and can be used for other purposes like to track the users in case of emergency.

1.1. Limitation to studies on patients' health

Traditional method of analyzing the patients with mental issues suffers from a number of challenges. The constraints on available mental health services, diagnostic cost, and inadequate professionals in India (Thirunavukarasu and Thirunavukarasu, 2010) make it difficult to deal with the mental health problems. The treatment process of those who suffer from mental health issues in India is not direct or conclusive.

1. The patients are required to be monitored in a controlled environment where the exact problem cannot be diagnosed by a psychologist/psychiatrist alone (http://indianmhs.nimhans.).
2. Health experts record the clinicians' narrative about how they observed patients and other findings that lead to the diagnosis and treatment plans for the patients (Seung, 2014). Hence, clinicians are not in a position to "paint the complete picture" of the patients' experience of mental health problem, as the patients generally answer interrogative questions in such a way that they themselves perceive to be viewed favorably by others (Chen et al., 2017).
3. Healthcare service providers record observations of patients during meetings only. As a result, critical changes in patients' behavior is not recognized immediately, in some cases it is missed out completely because of delay in reporting.

These situations prevent real-time recoding of the changing behavior of the patients over the period of time. On the contrary, it is very important to record real-time behavior of the patients to conclude the diagnosis for appropriate treatment.

1.2. Social media – a promising platform for behavior analysis

The situation in the rural areas is even critical given the fact that there

* Corresponding author.
E-mail addresses: deepalijayantjoshi@gmail.com (D. Joshi), manasi.patwardhan@tcs.com (Dr.Manasi Patwardhan).

https://doi.org/10.1016/j.chbr.2020.100036
Received 20 May 2020; Received in revised form 30 August 2020; Accepted 22 October 2020
Available online xxxx

Nomenclature

PTSD	Post Traumatic Stress Disorder
NLTK	Natural Language Toolkit
BS4	Beautiful Soup 4

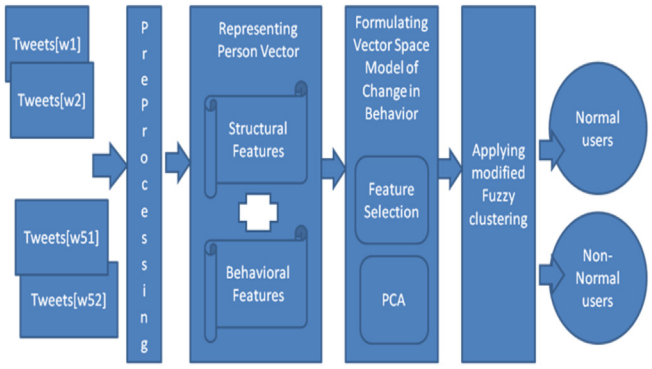

Fig. 1. System diagram.

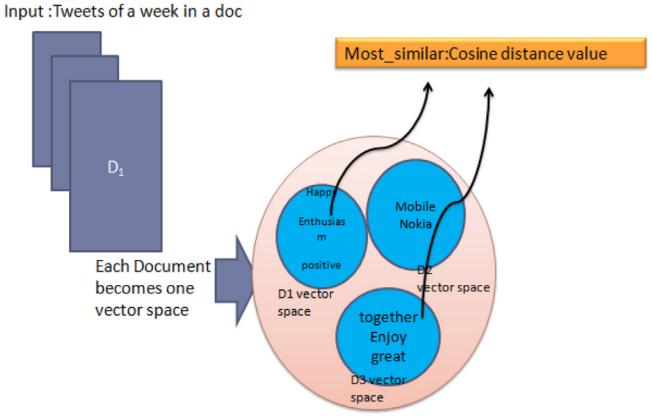

Fig. 2. Word embedding visualization.

is huge dearth of medical facilities. This is where social media comes to the rescue to the medical experts (Shatte et al., 2020).

Social media can become a platform to analyze a person's behavior (Liu et al., 2019). It can show the natural tendency toward the changing behavior of the users like how a person reacts to the current situation in his society, locality, state, country, community, linguistic group, business or professional group, and even in his family (Lieberman and Schroeder, 2020). There have been several cases in the last few years that gathered much limelight in the news that people posted messages of their own suicide and they committed suicide soon after, or they spread hate messages that led to fights and communal rights (Memon et al., 2018).

Such messages can raise an alarm when the things tend to go in a wrong direction. Not only that it can also help to communicate with the users who show going through similar experiences and can be stopped or rescued at the right time.

Studying social media platforms, like Twitter in the present study, can help us understand what matters most to the users or patients (Weerasinghe et al., 2018). Therefore, this type of large-scale user-generated content offers an opportunity to understand mechanisms underlying mental health conditions at an unprecedented level. For instance, studies of children and adolescents have already shown that high daily usage of social networking sites (Beattie et al., 2019; Seung, 2014) may be independently associated with low self-rating of mental health and experiences of higher levels of psychological distress and suicidal tendencies.

We aim to classify user behavior as normal vs. at-risk (prodrome/sub clinical) behavior based on features that define this inclination toward at-risk behavior. It is important to classify or predict the user's negative mental health traits into the aforementioned categories. Our long-term goal would be to provide assistance to the people suffering from mental health illness, and also to aid clinical researchers, epidemiologists and policy makers to understand community mental health through the communications taking place on social media.

2. Traditional methods to detect mental illness

Traditional methods of mental illness detection are based on questionnaires (Seung, 2014) and face-to-face interaction of the patients with

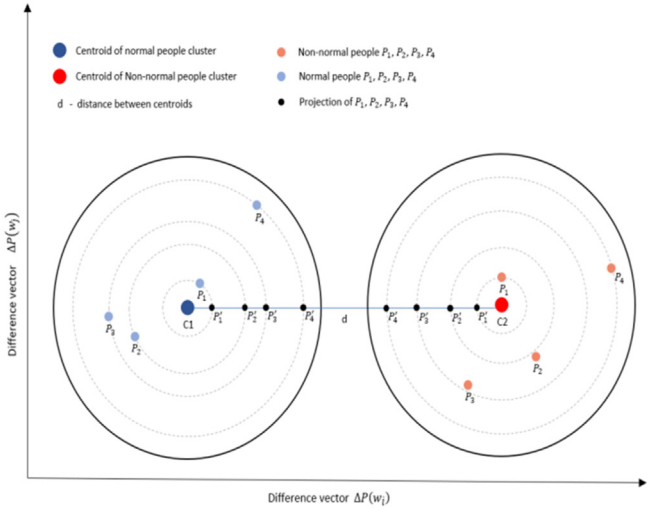

Fig. 4. Extended K-means membership calculation.

User ID	W_1	W_2	W_3	W_4	W_5	W_6	W_7
User 1	<SF><BF>	<SF><BF>	<SF><BF>	.....	.....	.....	<SF><BF>
User 2	<SF><BF>	<SF><BF>	<SF><BF>	.....	.....	.....	<SF><BF>
User 3	<SF><BF>	<SF><BF>	<SF><BF>	.....	.....	.....	<SF><BF>
User 4	<SF><BF>	<SF><BF>	<SF><BF>	.....	.....	.....	<SF><BF>
.....	<SF><BF>	<SF><BF>	<SF><BF>	.....	.....	.....	<SF><BF>
User 200	<SF><BF>	<SF><BF>	<SF><BF>	.....	.....	.....	<SF><BF>

Fig. 3. Data set visualization.

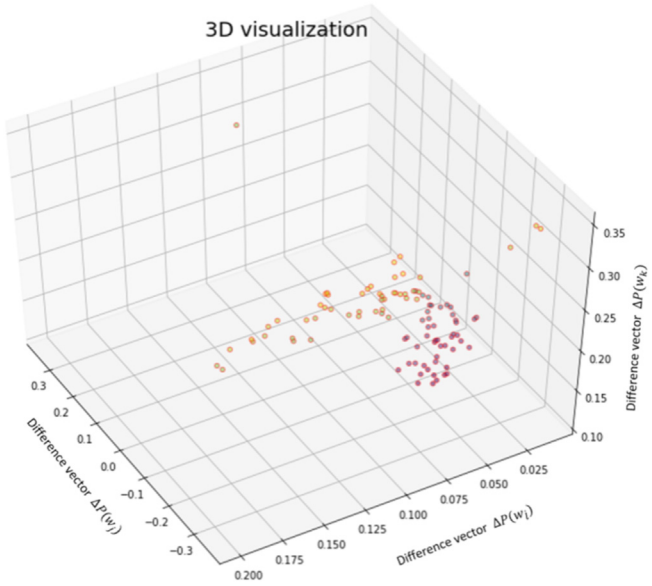

Fig. 5. Visualization of fuzzy C-means clustering.

the psychological consultants. Though it is a very effective method, very few patients are benefitted as there is scarcity of experts in India. On an average, there is only one psychologist to every 4 lakh patients in India (Garg et al., 2019). The present study for detecting mentally ill patients through social media behavior focuses mainly on the supervised approach wherein they have a labeled data set of patients who have the diseases and those who do not and use a classifier to do the task. In (O'Banion and Birnbaum, 2013; Ding et al., 2017; Islam et al., 2018; Pennacchiotti and Popescu, 2011) the authors gather social media data through crowd sourcing, build a SVM Classifier to predict the accuracy of depression from the social media posts. They have also used Linguistic Enquiry and Word Count (LIWC) (36. http://www.liwc.net/) lexicon in their implementation to score the words based on the emotions they carry. Multiple other approaches have used other classifiers such as naïve bayes (Joshi and Gaikwad, 2016), decision trees (Saravia et al., 2016), random forest (Joshi et al., 2018) and so on. The problem with these approaches is that they consider all the data as a snapshot and do not consider the time ordering when these posts are made. In (De Choudhury et al., 2013a) the features aggregated with respect to time depict overall behavior of a person but ideally the most recent posts should carry more weight as compared to older posts as suggested in (Fang and Hua Tai, 2014). Handling real-time data stream (Kumar et al., 2019) to identify suicidal or very intense posts are also being developed which try to provide instant help to the users. These systems generally follow an anomaly detection method to detect hazardous posts.

However, the problem with the supervised models is that it is highly domain dependent such as data set containing depression words, PTSD words, schizophrenia, and such expressions. The data is gathered either using crowd sourcing (De Choudhury et al., 2013a; O'Dea et al., 2015) or using regular expressions containing depressive words (Neuman et al., 2012). The data set thus obtained is very limited and fails to capture the features during the onset of the disease but later these features appear when the user is diagnosed with some mental illness. Getting a labeled data for all kinds of mental patients showing inclination toward some at-risk behavior is very difficult and cumbersome.

Therefore, we propose an unsupervised approach to detect users showing signals of succumbing to any kind of mental illness by tracing their behavior over a long period of time.

3. Proposed methodology

In the proposed methodology, there are three different algorithms suggested and implemented. First is our basic machine learning-based unsupervised model that we proposed in (Deepali, 2018) and tested with our synthetic data set and we have shown its result on real data in

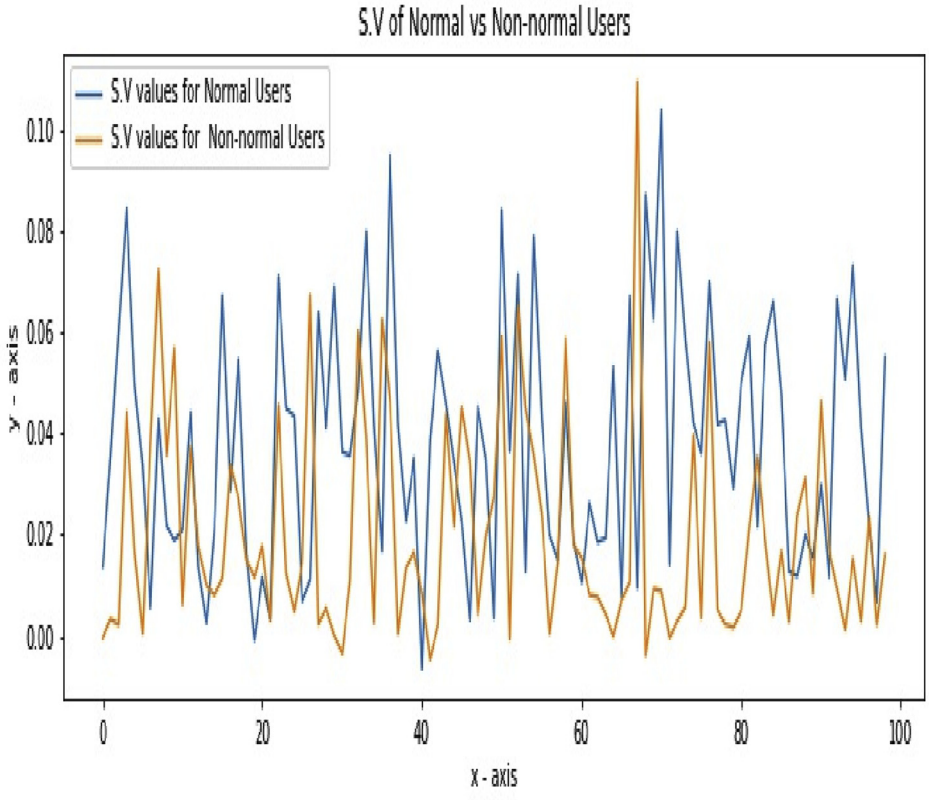

Fig. 6a. Visualization of structural features.

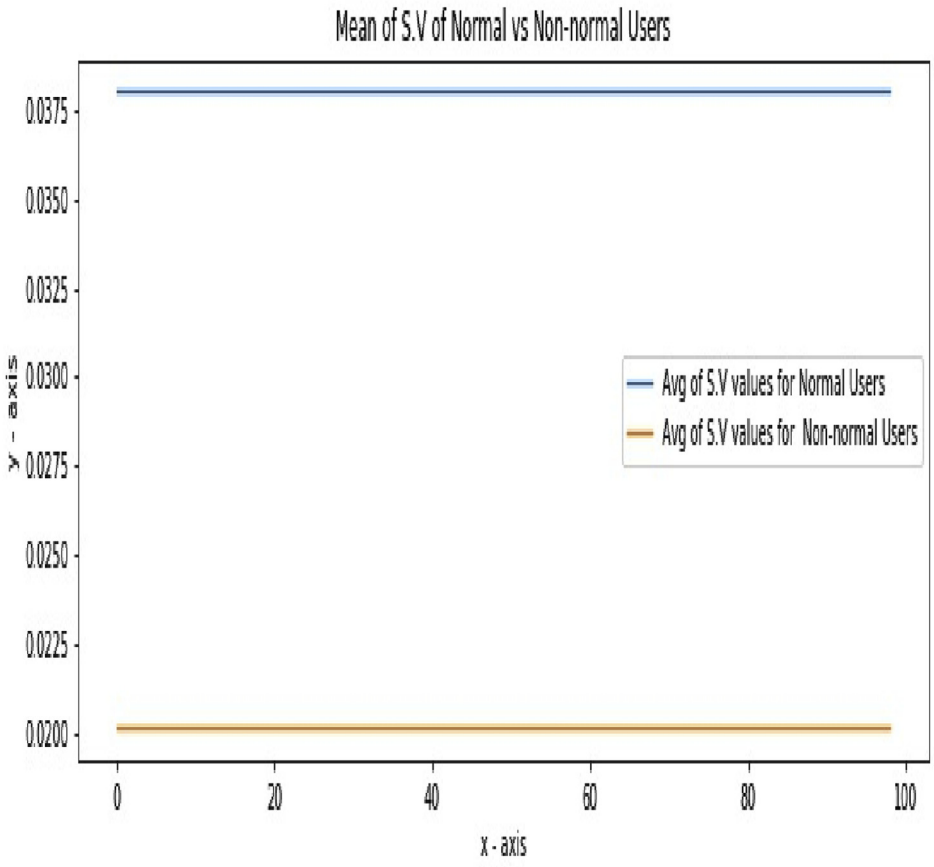

Fig. 6b. At-riskand normal users on structural features.

this paper. Also, we have done some modifications to our basic algorithm to incorporate fuzziness in it and detect the at-risk behavior in a probabilistic manner. Another modification that we have done on the basic algorithm is that we have used Word2Vec embedding for automatic feature extraction (Rong, 2016).

3.1. *Unsupervised approach using clustering*

In this section, we present a methodology that we have termed as unsupervised approach. This approach aims to achieve the goal of identifying at-risk behavior of users by noticing a drastically changed pattern in the user behavior (Fig. 1).

The first step is to process the text in the tweets by Natural Language Processing (NLP). The NLP has to be done very meticulously because of the various challenges such as availability of text in multiple languages, inclusion of slangs and day-to-day jargons, use of links and other media content, preserving context of the previous data with the current data, different encoding format of the data, and so on.

These issues must be dealt with during the data acquisition, preprocessing and vectorization activities and hence are one of the primary goals before we advance toward training the language model. For our study, we need two groups of people; (i) normal people and (ii) at-risk people—people who have been diagnosed with PTSD, anxiety, bipolar disorder, depression etc.

3.2. *Data set*

For the data gathering process, we have collected the data using two methods. The first method is to gather data using **keywords** and **hashtags**. Twitter API provides a method using which one can gather tweets that contain specific keywords or hashtags. We have used hashtags and keywords such as #**anxiety**, #**bpd**, #**bipolar**, #**depression**, #**stress** etc. These words being related to mental health give us relevant tweets. From such tweets, we identified the users who frequently make such tweets by using these keywords or hashtags and collect all the tweets of that particular user for inclusion in our data set. The second method is to gather the data by manually visiting and reading the tweets of the users who declare themselves to be depressed or suffer from mental illness in their description using regular expressions. By reading tweets of such users we decide if they are someone who are depressed and then we collected all their tweets for further reading. Then we followed the network of such users to find more users who might be depressed and studied their tweets closely. We have also considered the data set of those users who have registered themselves with the Depression Forums. The psychologists studied such tweets and validated their at-risk behavior.

For the second group of users – the normal users – we collected data of random users. We compared the data of normal users with the at-risk users. We took special care to select users who have posted at least 200 tweets. This is required because if the number of tweets are less than 200, we do not have sufficient data left to represent the users after the data goes through the filtering and extraction processes.

3.3. *Data volume*

Gathering good amount of data is useful to analyze the problems the users face and study the past issues that led to the mental health condition.

The data that we collected was of over 80,000 tweets. This data was in a very raw format and had to be filtered and preprocessed in order to make it useable.

The tweet data that we collected for each person was of 52 snapshots spanning over 52 weeks. The timestamp of the tweet proves to be important for our analysis. We used isocalendar() function in python to bifurcate the tweets in weeks. The isocalendar() function from the datetime() package in python returns the week number for a particular date that served as our timestamp.

Moreover, the data is full of social media jargons, which most of the online users use. Furthermore, we have used python modules such as

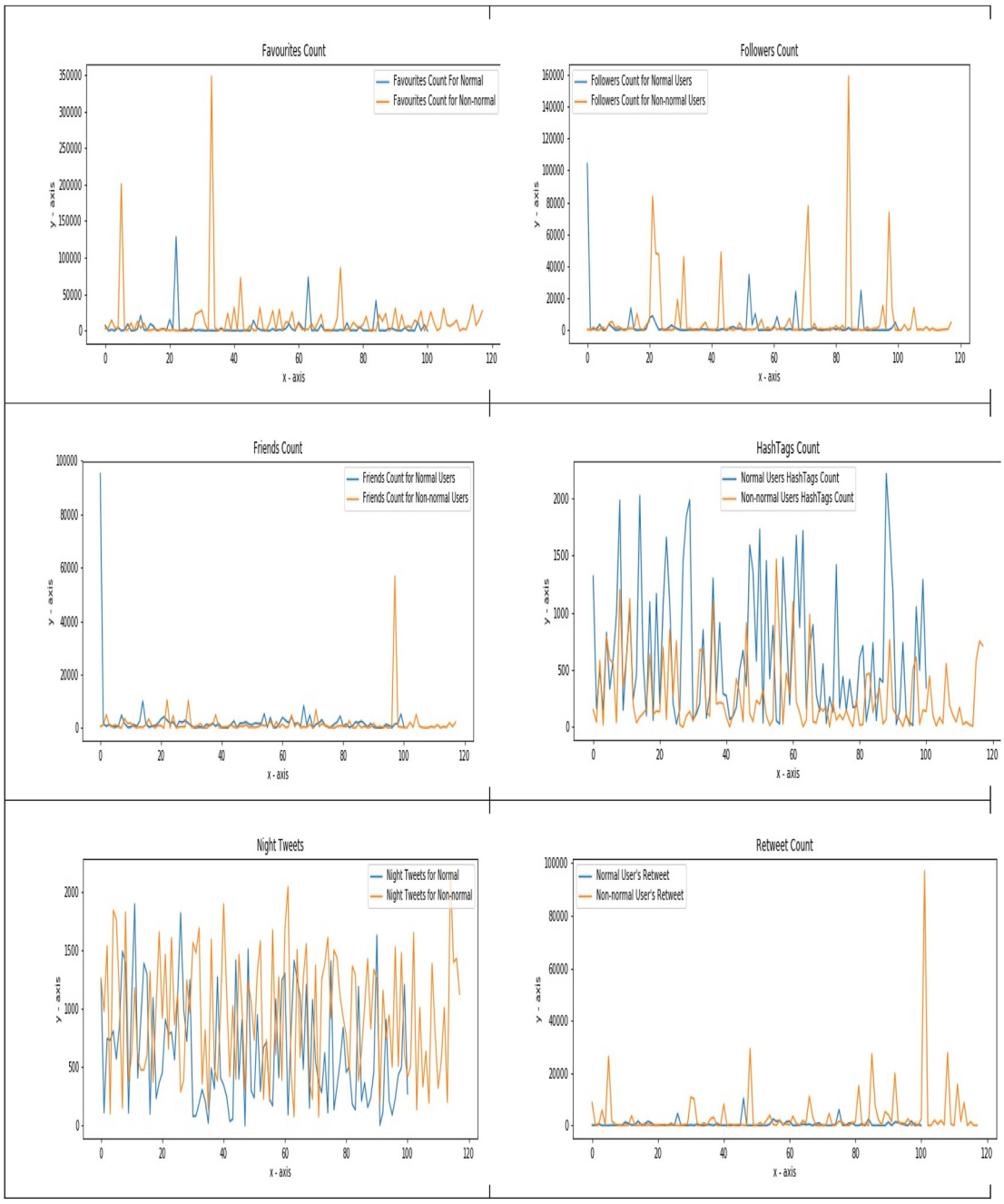

Fig. 7. Visualization of behavioral features.

NLTK and **gensim** to preprocess the data and remove stop-words (mundane words), which do not really affect the input fed to the model.

The data can be multilingual but currently our data is in English language only.

3.4. Data preprocessing

Data preprocessing is a very important step in cleaning the data. Various steps such as tokenizing and converting data to lowercase, removing mundane words that do not really affect the meaning of the sentence, removing special characters and symbols, tagging of part-of-speech (POS), and encoding characters were done.

We used NLTK module in python to do the initial preprocessing of data. The module is used for normal tokenization (NLTK et al., Bird) of the tweets. Tokenizing the data set splits up the sentences into tokens and then the stemming reduces the tokens into a single type, normally a root word; for example, the word "laughing" can be reduced to "laugh". As such, the stemming process reduces redundant words in a document. Abbreviations such as OMG, WTF, ASAP were treated as individual tokens, following the tokenization process. We replaced the emoticons with a suitable emotion carrying word using emoticon dictionary (Bracewell, 2008). We also identified the informal intensifiers, slang language, and social media jargons such as all-caps like 'I want to KILL myself!!!' and character repetitions like 'I've got a mortgage!! happyyyyyy' during the preprocessing. All-caps words were converted to lower case. This normalization helps to improve the performance of the POS tagging.

3.5. Feature engineering

In this section, we explain the process of extracting the desired

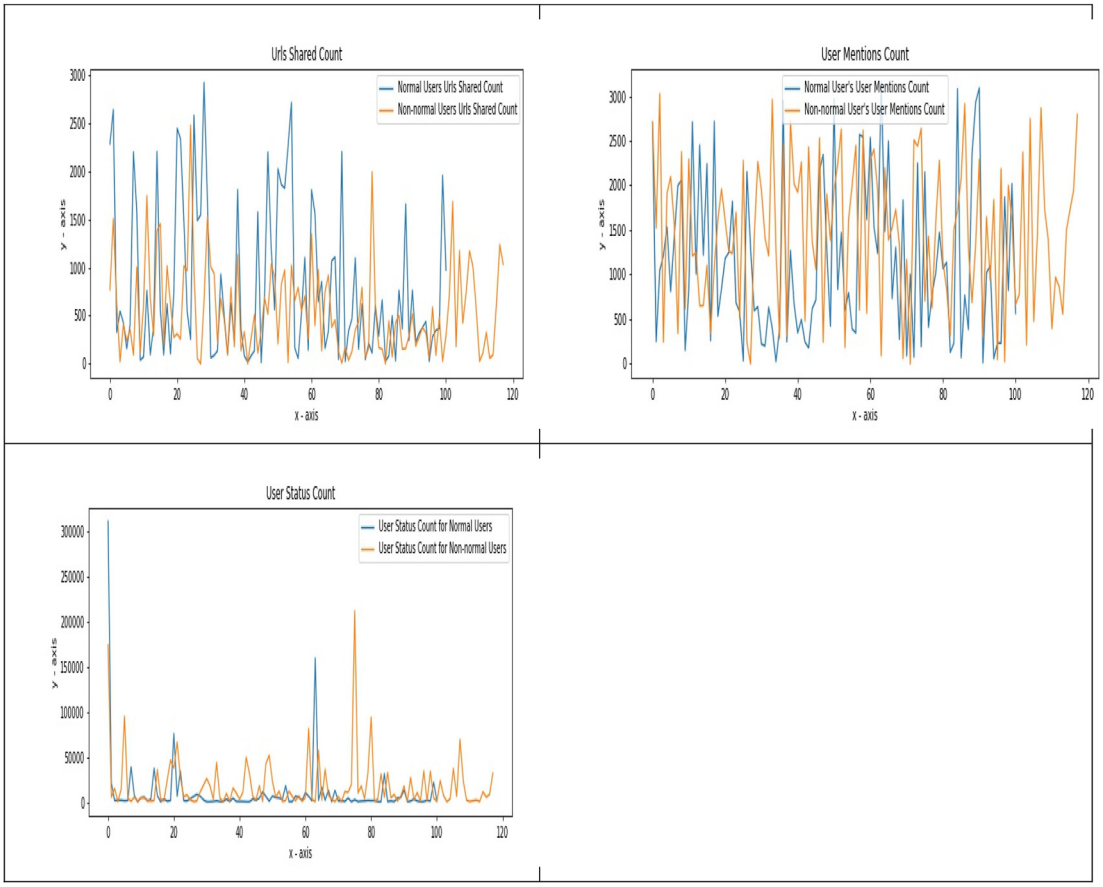

Fig. 7. (*continued*).

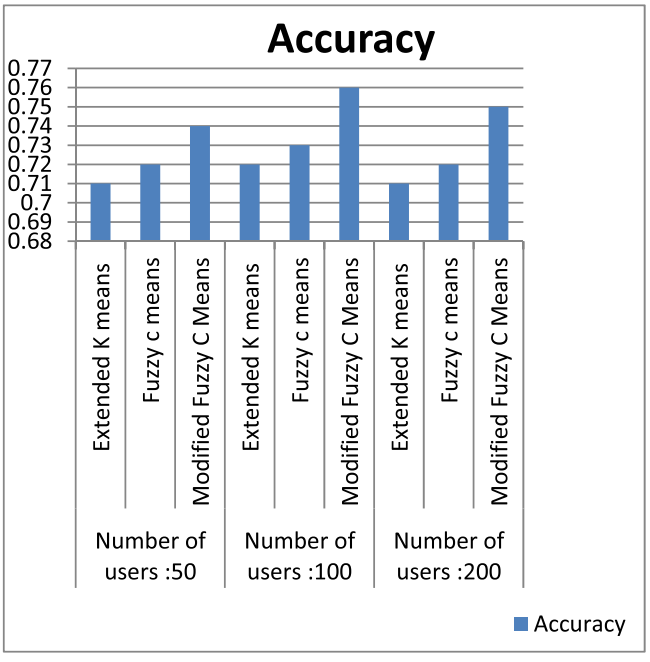

Fig. 8. Chart showing performance of proposed algorithms.

features in order to represent our users from their behavioral perspective. The features have been categorized into two types: **structural features** and **behavioral features**. Structural features consist of all the words in the tweets, which we get after preprocessing the texts written in the tweets and behavioral features consist of the features depicting the engagement and responsiveness of the users. We selectively considered the features that are relevant to the users' behavior and mood such as number of friends, followers, favorites, time of the tweet, replies to people, mentions of other users and so on. We selected 14 such features (Deepali, 2018) that are relevant to the user activity. The structural and behavioral features were aggregated over a period of one week.

3.6. *Vectorization using Word2Vec model*

Word2Vec (Rong, 2014) is a group of related models that are used to produce word embedding. These models are shallow 2-layered networks that are used to capture linguistic contexts of words. We need to supply corpus of texts to the model, which in turn produces vectors for each word of typically having large dimensions, such that the words that have similar contexts are located close to each other in the vector space.

In this project, we used Word2Vec model for training and then to get vectors for each word in the tweet. For the training, we first fed it with a lot of tweets and generated context. Here the Word2Vec model creates the vectors according to the context in which different words are used in the tweets. This makes it easier to find the words that are related to each other, the words that are opposite, and the words used in similar context. We also applied an additional filter that removes all the words that have occurred less than five times in all the tweets from the Word2Vec model. Hence, we do not consider words for which we do not have enough information or context. This completes the training of the Word2Vec model.

Therefore, when we analyzed a tweet, we split the tweet into a list of words that gives a vector for each word. We used all the vectors of the words and calculated an average vector. This vector represents a vector for the whole tweet. We calculated the average vector for all tweets of the users.

For example, let us consider this sentence.

"This is a very nice looking phone loaded with a lot of features."

On the basis of Word2Vec analysis, we can list the words as follows.

["nice", "looking", "phone", "loaded", "lot", "features"].

We receive a **100 dimensional vector** generated by the Word2Vec model for each word in the list. The word embedding visualization can be seen in Fig. 2.

This example shows that how useful the Word2Vec model is in preserving the context of the sentences. In this method, a timeline of a person is built having 52 weeks data, where each week is the consolidated record of all the tweets made in that week. This data is concatenated by the behavioral features and a person vector is represented by 52 dimension vector in a vector space model.

4. Mathematical representation

In this section, we demonstrate the mathematical representation of the given problem by modeling a person in VSM and by devising an algorithm for at-riskbehavior detection.

4.1. Classification of users in vector space model

As shown in Fig. 3, a person is represented in the Vector Space Model (VSM) by the structural features <SF> and behavioral features <BF>. These features are aggregated week wise and computed over a period of one year. There is an inherent time dimension attached with these features, which signifies sequentiality in the pattern of our data.

- Week is represented by timestamp: W_1 …. W_n
- W_1 is the first timestamp and W_n is the last timestamp.
- Where, W_i contains consolidated tweets posted during ith week.
- SF > Vectorized tweets made by user in W_i.
- BF > Behavioral features aggregated over one week.

4.2. Detection of behavior change

Detection of at-risk or substantial change in behavior pattern of a person is our major goal that signifies the traits depicting inclination toward mental illness. In our previous study (Joshi et al., 2018; Deepali, 2018), we proposed a novice algorithm to detect at-risk change in behavior. The limitation of the study is that we were not able to prove the accuracy with the real data as it was not labeled. Therefore, we got the data labeled from a professional psychologist. This data helped us in calculating the accuracy of our model on real data.

We also introduced an improved fuzzy clustering approach and extended version of k-means algorithm. Fuzzification gives us the advantage of measuring the probability of a person belonging to at-risk vs normal category.

We have week wise numeric representation of the user. This means for an individual user, we now have *n* numeric values for *n* number of weeks' data.

$$\Delta P_{i(W_2)} = P_{i(W_2)} - P_{i(W_1)}$$

Subsequently, in order to detect change in behavior of a user, we calculate change as:

$$\begin{aligned} \Delta P_{i(W_2)} &= P_{i(W_2)} - P_{i(W_1)} \\ \Delta P_{i(W_3)} &= P_{i(W_3)} - P_{i(W_2)} \\ \Delta P_{i(W_i)} &= P_{i(W_i)} - P_{i(W_i-1)} \end{aligned}$$

4.2.1. Feature vector used for clustering

Thus, we have set of difference between successive weeks of individual user represented by. $\Delta P_{i(W_i-1)}$

In order to consider important features, we applied Principal Component Analysis (Ding and Li, 2007) that reduces the dimensionality of the data and gives better results. We have also considered a maximum of 30 displacement vectors and got better results than considering all 51-week difference vectors for each user. We have explained the details in the experimentation section. The person's vector is displayed in a new vector space model that is as follows:

$$\Delta P_i = \{\Delta P_{i(W_2)}, \Delta P_{i(W_3)}, \ldots\ldots\ldots, \Delta P_{i(W_n)}\}$$

After PCA/Feature selection, we applied modified fuzzy c-means (Windham, 1981) and extended k-means algorithm to find two clusters namely of at-risk and normal people. As we have the data labels, we can easily validate our approach and calculate the accuracy of our model.

In the next section, we have explained our basic algorithm and its two modifications.

4.3. Approaches to detect behavior change

Following are the three proposed approaches to detect at-risk behavior of users by calculating the change in behavior of the users in the subsequent weeks.

CASE 1. Our proposed prodromal(at-risk) detection algorithm

The algorithm of change in behavior was observed as follows:

1. For time instance t = 1,…,t-1, t, t +1,……,T

   Features persons in a Vector Space Model Mt

   Where Mt = {$P1$,2 ,…..,Pi ,….,Pp }
2. For i = 1, 2, …., p Pi ={$S1$,$S2$,….,Sj,…..,Sm,$B1$,$B2$,….,Bk ,….,Bl }

   Where Sj for j = 1,…. m is a set of structural features for a week's duration forming time instance t

   Bk for k = 1… l is a set of behavioral features for a week's duration forming time instance t

   Total number of dimensions for Pi are n = m + l
3. (for t ≠ 1) Compute ΔMt + 1 = Mt + 1 - Mt
4. For i = 1, … , p
5. Δ $Pi(t+1)$ = $Pi(t+1)$ - $Pi(t)$
6. Calculate feature vectors difference successively that represents change in behavior week wise.
7. If Δ $(t+1)$ > threshold defined by k-means clustering on ΔMt + 1, the person has change in behavior at time instance t.

CASE 2. Modified c-means algorithm

It uses the idea from the field of fuzzy logic and fuzzy set theory in which objects are allowed to belong to more than one cluster with some weight. When clusters are well separated, a crisp classification of objects into clusters makes sense. But in many cases, clusters are not well separated as it is in our case. If we follow a crisp classification, a borderline object ends up being assigned to a cluster in an arbitrary manner that we need to address.

Assume a set of n objects,

$X = \{x_1, x_2, \ldots. x_n\}$

Where xi is a d-dimensional point.

A fuzzy clustering is a collection of k clusters, C1, C2 up to Ck, and a partition matrix W = wi, j belongs to [0, 1], for i = 1 upto n and j = 1 upto k, where each element $w_{i,j}$ is a weight that represents the degree of membership of object i in cluster C_j .

Following are the two restrictions to achieve pseudo partition:

All weights for a given point **xi** must add up to 1.

Each cluster Cj contains, with non-zero weight, at least one point, but does not contain, with a weight of one, all the points.

The fuzzy c-means algorithm is as follows:

1. Select an initial fuzzy pseudo-partition and assign values to all $w_{i,j}$
2. Repeat this for all the data points.
3. Compute the centroid of each cluster using the fuzzy partition.
4. Update the fuzzy partition, i.e, the $w_{i,j}$
5. Don't change until the centroids.

4.3.1. Modified c-means algorithm analysis

We modified the fuzzy c-means algorithm by altering the distance measure to cosine distance instead of Euclidean distance. In cosine similarity, we can measure the similarity of the direction instead of magnitude. Therefore, it works well with our problem in hand. It is the most widely used distance measure with respect to text documents (Huang, 2008; De Choudhury et al., 2013b, 2014).

CASE 3. Extended k-means algorithm

K-means generates partitions and each data point can be assigned in one of the clusters. It is also called as hard clustering. The drawback of the k-means algorithm is that it does not give us the strength of belonging to a cluster. This means if a data point is very close to the mean and other point is at the cluster boundary, both are treated equally but the belongingness of the prior is obviously more than the former, which is not taken care by k-means algorithm. Therefore, we have implemented extended k-means algorithm that not only does clustering but also tells us the strength of belongingness of each data point toward a cluster.

The extended k-means clustering algorithm that we propose is as follows:

1. Select the number of seed points and randomly select those many data points as centroids.
2. Assign all the data points to the nearest centroid based on the Euclidean distance.
3. Calculate new seed points by finding the mean of all the points.
4. Repeat until there is no change in the centroid.
5. Considering the centroid to be 100% belonging to the cluster, the closest data points of the other cluster is considered to be 0% belonging. All the other points distance from the centroid is calculated by taking their circulated projection on the horizontal plane and their belongingness is calculated.

Fig. 4 shows the supposition made for the calculation of percentage membership values for all users. Doted circles show the projection of the individual data points in cluster on the straight-line (d) connecting centroids. This straight-line is the Euclidean distance between centroids (C1, C2) of both clusters.

D is derived by the Euclidian distance formula.

Here, we have the Euclidian distance of each data point from its centroid, which is (x). Where d is the distance between centroids. Centroids are 100% absolute belonging member of cluster, so by considering d as the maximum distance percentage of the membership, values for all the users are calculated as.

$$P_i(m) = Percentgeofmembership = 100 - \left(\frac{100 \times x}{d}\right)$$

where,

i = Person belongs to cluster (1,2,3,4 …. n)

x = Euclidean distance between centroid and its cluster members.

For example,

In first cluster, distance from C1 to $P_1^{'}$ is x and C1 to C2 is d.

Thus, we get two clusters mainly containing the normal and at-risk clusters. We can validate the results as we already have labeled data set with us. The visualization of the clusters can be seen in Fig. 5. The three dimensions are the principal components of the difference vectors that we have calculated as shown in section 3.6. We can observe in the clustering result that the difference vector for at-risk users is less than the normal users that signifies that the at-risk people tend to a redundant behavior, repeat the same things several times that is captured by the less magnitude of the difference vector.

5. Results & discussion

In this section, we start by depicting the most contributing features that distinguish normal users from at-risk users. Few of them are given below.

Fig. 6a depicts the visualization of structural features. The x-axis is the number of users and the y-axis is the structural feature encoding score we get after applying the Word2Vec Model.

Fig. 6b shows the difference between average of at-risk and normal users using only structural features.

Graphs showing behavioral feature differences are given below (Fig. 7).

These above behavioral features show the significant contribution these behavioral features make in distinguishing normal users from at-risk users.

We can observe that the at-risk users tweet more at night as compared to normal users, favorite count, re-tweet count, and follower count of at-risk users is more as compared to normal users. At-risk users also put lot of status and user mentions.

In this section, the results of the machine learning algorithms using newly proposed unsupervised method is stated and depicted visually to comprehend the results. We have implemented the variations of the clustering algorithms to different sets of users to check the performance of the algorithm if the data increases. The performance of the algorithm is quite consistent for different number of users. Modified fuzzy c-means algorithm is the best performer of all. We have also experimented considering only structural features, only behavioral features before applying clustering but we get the best results when we use both types of features (Fig. 8).

In the case of true positives where our model identifies correct at-risk users, their data would be forwarded to the health care department

where a clinical psychologist will closely observe their data manually and verify the finding and also give them required counselling and treatment. In the case of false positives where our model predicts one normal person as at-risk person they will get filtered out when their data will be verified by clinical psychologist and wont get any kind of disturbance in his daily routine life.

Declaration of competing interest

No conflict of Interest.

Acknowledgement

This project is funded by Department of Science and Technology, New Delhi under SYST Scheme. Ref No: SP/YO/060/2016. We are extremely thankful for the authorities for their support.

We are very grateful to Mr. Aditya Shukla (Applied psychologist & Founder, Chief Author, cognitiontoday.com) and Dr. Vinay Vaidya (Professor, Department of Technology) for their valuable support and continued guidance in this Project.

References

http://www.liwc.net/.

Bauer, Una (2017). Online oversharing and impersonal engagement. *Performance Studies international, 23*. Overflow.

Beattie, T. S., Prakash, R., Mazzuca, A., et al. (2019). Prevalence and correlates of psychological distress among 13–14 year old adolescent girls in North Karnataka, South India: a cross-sectional study. *BMC Publ. Health, 19*, 48. https://doi.org/10.1186/s12889-018-6355-z

Bracewell, David B. (2008). Semi-automatic creation of an emotion dictionary using wordnet and its evaluation. In *2008 IEEE Conference on Cybernetics and Intelligent Systems*. IEEE.

Chen, Wu, et al. (2017). Personality differences in online and offline self-disclosure preference among adolescents: a person-oriented approach. *Pers. Indiv. Differ., 105*, 175–178.

De Choudhury, Munmun, et al. (2013a). *Predicting Depression via Social Media*. ICWSM.

De Choudhury, Munmun, Counts, Scott, & Horvitz, Eric (May 2013). *Social Media as a Measurement Tool of Depression in Populations*. WebSci.

De Choudhury, Munmun, Counts, Scott, Horvitz, Eric J., & Hoff, Aaron (Feb 2014). *Characterizing and Predicting Postpartum Depression from Shared Facebook Data*. Social Technologies and Well-Being.

Deepali, J (2018). Joshi Modeling and Detecting Change in User Behavior through His Social Media Posting Using Cluster Analysis CODS'17. *Proceedings of the Fourth ACM IKDD, Conferences on Data Sciences*.

Ding, Chris, & Li, Tao (2007). Adaptive dimension reduction using discriminant analysis and k-means clustering. In *Proceedings of the 24th International Conference on Machine Learning*.

Ding, Tao, Bickel, Warren K., & Pan, Shimei (2017). Multi-view unsupervised user feature embedding for social media-based substance use prediction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Fang, Ying-En, & Hua Tai, chih (2014). A Mental Disorder Early Warning Approach by observing depression symptom in social diary. In *IEEE Conference on Systems,Man,Cybernetics*. San Diego,CA,USA.

Garg, Kabir, Naveen Kumar, C., & Chandra, Prabha S. (2019). Number of psychiatrists in India: baby steps forward, but a long way to go. *Indian J. Psychiatr., 61*(1), 104.

http://indianmhs.nimhans.ac.in/Docs/Summary.pdf.

https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.

Huang, Anna (2008). Similarity measures for text document clustering. In *Proceedings of the Sixth new zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand* (vol. 4).

Islam, Md Rafiqul, et al. (2018). Depression detection from social network data using machine learning techniques. *Health Inf. Sci. Syst., 6*(1), 8.

Joshi, Deepali J., Makhija, Mohit, Nabar, Yash, Nehete, Ninad, & Patwardhan, ManasiS. (2018). Mental health analysis using deep learning for feature extraction. January 11–13, 2018, Goa, India .©. In *CoDS-COMAD '18*. Association for Computing Machinery.ACM (ISBN).

Kumar, Akshi, Sharma, Aditi, & Arora, Anshika (2019). *Anxious Depression Prediction in Real-Time Social Data*. Available at: SSRN 3383359.

Lieberman, Alicea, & Schroeder, Juliana (2020). Two social lives: how differences between online and offline interaction influence social outcomes. *Current opinion in psychology, 31*, 16–21.

Liu, Xingyun, et al. (2019). Proactive Suicide Prevention Online (PSPO): machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. *J. Med. Internet Res., 21*(5), Article e11705.

Luo, Mufan, & Hancock, Jeffrey T. (2020). Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current Opinion in Psychology, 31*, 110–115.

Memon, Aksha M., et al. (2018). The role of online social networking on deliberate self-harm and suicidality in adolescents: a systematized review of literature. *Indian J. Psychiatr., 60*(4), 384.

Neuman, Yair, et al. (2012). Proactive screening for depression through metaphorical and automatic text analysis. *Artif. Intell. Med., 56*(1), 19–25.

NLTK: Edward Looper, Steven Bird. NLTK: the Natural Language Toolkit.

O'Banion, Shawn, & Birnbaum, Larry (2013). Using explicit linguistic expressions of preference in social media to predict voting behavior. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

O'Dea, Bridianne, et al. (2015). Detecting suicidality on twitter. *Internet Interventions, 2.2*, 183–188.

Pennacchiotti, Marco, & Popescu, Ana-Maria (2011). A machine learning approach to twitter user classification. In *Fifth International AAAI Conference on Weblogs and Social Media*.

Joshi, Deepali J., & Gaikwad, Govin (2016). Multiclass mood classification on twitter using lexicon dictionary and machine learning algorithms. In *International Conference on Inventive Computation Technologies (ICICT)*.

Rong, Xin (2014). *word2vec Parameter Learning Explained*. arXiv preprint arXiv: 1411.2738.

Rong, X. (2016). *word2vec Parameter Learning Explained*. arXiv:1411.2738vol. 4.

Saravia, Elvis, Chang, Chun-Hao, De Lorenzo, Renaud Jollet, & Chen, Yi-Shin (2016). MIDAS: mental illness detection and analysis via social media. In *IEEE/ACM ASONAM*. August 18-21, 2016, San Francisco, CA, USA.

Seung, W. (2014). Choi, benjamin schalet, Karon F. Cook,David Cella, "Establishing a common metric for depressive symptoms: linking the BDI-II,CES-D, and PHQ-9 to PROMIS depression. *Psychological Assesment, 26*, 513–527.

Shatte, Adrian BR., et al. (2020). Social media markers to identify fathers at-risk of postpartum depression: a machine learning approach. *Cyberpsychol., Behav. Soc. Netw.*

Thirunavukarasu, M., & Thirunavukarasu, P. (2010). Training and national deficit of psychiatrists in India–A critical analysis. *Indian J. Psychiatr., 52*(Suppl. 1), S83.

Weerasinghe, Janith, Morales, Kediel, & Greenstadt, Rachel (2018). Analyzing machine learning models that predict mental illnesses from social media text. *Studies, 3*, 5.

Windham, Michael P. (1981). Cluster validity for fuzzy clustering algorithms. *Fuzzy Set Syst., 5*(2), 177–185.