International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Machine Learning Framework for Detection of Psychological Disorders at OSN

Punam B. Nalinde, Anita Shinde

Abstract: ***Now a days attractiveness of social networking sites indications to the problematic habit. For this reason, researchers devised stress detection systems based psychological disorders in social networks. In this work, we propose a system of psychological disorders detection (PDD) that can provide online social behaviour extraction. It offers an opportunity to identify disorder at an early stage. These PDD system are made a different and advanced for the preparation of disorder detection. Propose system a machine learning approach that is detection of psychological disorders in social networks and social interaction features from social network data for detect with precision possible cases of disorders detection. We perform an analysis of the characteristics, and we also apply machine learning classifier in large-scale data sets and analyse features of psychological mental disorders. After classification results show that user are in stress or not, will be detected by PDD system is used to recommend hospitals on a map and at the same time admin will send mail of precaution list to user for users healthy and happy in life. The proposed method could help in developing a social network diagnostic tool for stress detection. It is useful in the diagnosis of psychological disorder detection in social platforms.***

Index Terms**: *PDD (Psychological Disorder Detection), OSN (Online Social Network), SNMD (Social Network Mental Disorder) Classifier, feature extraction.***

I. INTODUCTION

With the explosive growth in popularity of social networks, messaging applications and online social networks (OSN) they have become part of the daily life of many peoples. Psychological stress is turning into a risk to individual's well-being these days. The research on the extraction of social networks focuses on the discovery. The knowledge behind the data to improve people's life while OSNs seem to expand the capacity of their users increasing social contacts can actually diminish Interpersonal interactions face to face in the real world [1],[18].Psychological stress is a leading cause of several psychophysiological disorders. For example, it increases the likelihood of depression, stroke, heart attack and cardiac arrest [17]. User social interactions on social networks contain valuable prompts for pressure identification. It is not easy to detect user psychological disorder in an early time to protect user.

Manuscript published on 30 September 2019.
*****Correspondence Author(s)
Miss. Punam B. Nalinde, ME Student, Computer Department, MMCOE College, Pune, India.
Prof. Anita Shinde, Associate Professor, Computer Department, MMCOE College, Pune, India.

In proposed system, we find that users disorder state is closely related to friends in social media, and we work a facebook dataset from real-world social platforms to systematically study the correlation of users disorder states and social interactions. Psychological confusion itself is non-clinical and common in our life, extreme and perpetual issue can be rather harmful to people physical and mental health. User social interactions on social networks contain valuable prompts for pressure identification. The advance of informal organizations like Twitter, Facebook a consistently expanding number of individuals will share their consistently occasions and mind-sets, and connect with companions through the interpersonal organizations. We can group utilizing machine learning structure. Because of use both Facebook post content attributes and social communications to improve Psychological confusion discovery. Utilizing some related and previous work, we systemize a group of features as attributes to construct the detective models we proposed.

The contribution of this paper are as follows:

- Today online disorders are usually processed late. To actively identify possible cases of disorders, we propose an innovative approach, disorder detection practice, extracting data records from OSN users as an early detection system.
- We develop a machine learning framework to be detected Psychological disorders detection (PDD). We also design and analyse many important features to identify OSN disorders, such as disinhibition, parasocality, self-revelation, etc. The proposed framework can be implemented to provide a timely alert for potential patients.
- To improve the accuracy using advanced naive bayes algorithm and it is simple to recognize. The proposed framework provides a better results and make predictions for these reasons alone you should take a closer look at the algorithm.

II. LITERATURE SURVEY

Chun-Hao et al. [2] have introduce a research on mental disorders on social media networks. Author used predictive models for data collection process, using subconscious crowdsourcing. This study, B. Saha et al. [1], the proposed technique can help to classifying online mental health related communities. Author are extract two type of feature 1) STL (single task learning 2) LIWC (linguistic inquiry and word count) features from online social media for depression patients to analyze their outlines.

B. Viswanath et al. [6] have introduced about the study of evolution of users activities on facebook. In the social network of Facebook to capture this notion. Author find that links in the activity network tend to come and go. Quickly over time, and the strength of the bonds exhibits a reduce the decreasing tendency of the activity as a connection of a social network centuries. For example, only 30 percent of Facebook user pairs interact. It is interesting to note that also find this, even if the connections of the activity network many properties of graph theory change rapidly over time. The network of activities remains unchanged.H. Lin, et al. [8] introduced about a programmed pressure identication technique from cross-media smaller scale blog information. They implement three-level system for stress discovery from cross-media small scale blog information. This paper used by joining a Deep Sparse Neural Network to consolidate distinctive highlights from cross-media small scale blog information, the structure is very practical and effective for stress identication. This structure, the proposed technique can help to consequently identify mental worry from informal communities. The future extension intend to research the social relationships in mental worry to additionally enhance the identication execution.Liqiang Nie et al. [9] have introduced to contemplate about spanning the vocabulary hole between wellbeing searchers and medicinal services information with a worldwide learning approach. This paper restorative wording task plan to connect the vocabulary hole between wellbeing searchers and human services information. Author includes two segments, nearby mining and worldwide learning. Extensive assessments on a genuine world dataset show that our plan can create promising execution when contrasted with the predominant coding techniques. They examine how to adaptably arrange the unstructured therapeutic substance into client needs-mindful cosmology by utilizing the suggested restorative phrasings.Chi Wang et al. [11] have introduced and learns about the impact augmentation issue, which plans to locate a little subset of hubs (clients) in an interpersonal organization that could amplify the spread of impact. In this paper propose a Pairwise Factor Graph (PFG) model to formalize the issue in probabilistic model, and they broaden it by consolidating the time data, which results in the Dynamic Factor Graph (DFG) mode. In this study proposed methodology can viably and the dynamic social impacts. Author are improve the parallelization of our calculation should be possible in future work to scale it up further.Lexing Xie et al. [13] have introduced about Picture labels and world information: taking in label relations from visual semantic sources considers the utilization of regular words to depict pictures. This paper proposed labeling calculation sums up to inconspicuous labels, and is additionally enhanced joining tag-connection highlights got by means of ICR. Author used methods to all the more likely 1) consolidate multi-word terms. 2) out-of-vocabulary words. 3) Propelled NLP procedures for taking in word relations from freestyle content. 4) Assessment of idle idea connection proposal. 5) Anticipating the sort of relations.Yuan Zhang et al. [14] introduced a novel issue of feeling forecast in informal communities. This paper used a technique alluded to as Mood cast for demonstrating and anticipating feeling elements in the informal community. The proposed methodology can successfully demonstrate every clients feeling status and the expectation execution is superior to a few gauge strategies for feeling forecast. It is utilized to because of the predetermined number of members. This framework trial results on two diverse genuine interpersonal organizations exhibit that the proposed methodology can successfully demonstrate every clients feeling status and the forecast execution is superior to a few standard techniques for feeling expectation.Hong-Han Shuai et al. [20] introduced a tensor model have been used because a tensor can naturally constitute multi source data. The latest technology SNMDD based Tensor model STM is presented which allows incorporating the characteristics of SNMDs. Furnished with a tensor model semi supervised learning has been constructed to categorize each user by utilizing TSVM (Transductive Support Vending Machine). Chih-Hua Tai et al. [15] have introduced a method to measure and detect mental disorder through his blog posts on the Internet. In this paper author used LDA (Latent Dirichlet Allocation) to find out top frequency words and use SentiWordNet to calculate the emotion score of the user. In proposed method, words others than sentiment words were removed first. Hong-HanShuai et al. [16], Ahmad Rauf Subhani et al. [18] introduced about the Social Network Mental Disorder Detection (SNMDD) model. This paper introduced data mining features of SNMDD. Author used methods to all the more likely 1) Cyber Relationship (CR), 2) Information Overload (IO), 3) Web addiction. In cyber relationship obsession which includes the obsession with social media surfing to converse and share private information to the public where online relationships became more significant than friends and family circles. In Information Overload comprises obsessive scanning of user status tweets posts which leads minimal in person interaction and vast reduction in work level. In Web addiction which includes obsessive online social gaming and gambling which affects ones profession.

TABLE I. Comparative Study of Approaches Used for Detecting Depression

Sr No.	Author	Paper Name	Method	Result
1	Huije Lin et al,(2010)	Psychological stress detection from Cross-Media Microblog data using deep sparse neural network	Use cross-media microblog data with Deep sparse neural network	Accuracy-SVM-85.97, SAE-89.68, LAE-90.55
2	Jie Tang et al,(2011)	Quantitative study of Individual emotional states in social networks	SVM and Naïve Bayes with Mood cast	Accuracy-62.17
3	Munmun De Choudhury et al.,(2013)	Social Media as a Measurement tool of Depression in population	SVM classifier, Center for Epidemiologic studies Depression scale(CES-D), social media depression index(SMDI), Principal Component Analysis(PCA),	SMDI can nearly reflect CDC characterized insights on depression.

4	Munmun De Choudhury et al.,(2014)	Characterizing and Predicting Postpartum Depression from Shared Facebook Data	Patient Health Questionnair (PHQ9), LIWC	Postpartum Depression was best predicted.
5	Budhaditya Saha et al,(2015)	A Framework for classification online mental health related communities with an interest in depression	Combine LIWC + topic, LDA	Accuracy-0.876
6	Elvis Saravia et.al.,(2016)	MIDAS: Mental illness detection and analysis via social media	Center for Epidermologic studies depression scale, TF-IDF,PLF, Sentiment 140API, Random Forest Classifier: a main learning mode	They built an online system that extracts the features of a user by considering two mental disorders, and their system gives the minimal results, that can be used in future to predict the user behaviour more efficiently
7	Keumhee et.al.,(2016)	Identifying Depressive users in Twitter using Multimodel analysis	SVM based Learning, Built a Lexicon by using Visual Sentiment Ontoloy and Sentistrength dictionaries, LIWC, K-means Clustering Latent Fusion	The results shows that a multimodel that is developed has high accuracy as compared to the existing methods, and can efficiently predicts the user's mood.
8	Maryam Mohammed Aldarwish et.al.,(2017)	Predicting Depression Levels using Social Media Posts	BDI-II Questionnaires, create a depression model using Rapid Miner, SVM and Naïve Bayes classifiers	The performance of the model is calculated and they got the best precision and minimal accuracy and recall
9	Adrian Benton et, al.,(2017)	Multitask Learning for Mental Health Conditions with Limited Social Media data.	Multitask Learning approach(MTL), Logistic Regression, Feed forward multilayer perceptron Single task Learning(STL),	Results shows that the proposed model performs better compared to LR models.
10	Hong-Han et, al,(2018)	A Comprehensive study on social network mental disorder detection via online social media mining	Use SNMD with SNMD-based Tensor model(STM)	Accuracy-CR-80.5, NC-77.6, Io-82.9

Table 1. Summaries the different approaches used by the researchers to predict the Mental state of users, most of the approaches are based on the machine learning algorithms were used to analyses the mental states of the users based on their input information.

III. PROPOSED METHODOLOGY

In proposed system we develop new approaches for detecting psychological disorder cases of OSN (online social network) users. We claim that mining social network data of individuals, as an opposite another to the predictable psychological approach, delivers an outstanding opportunity to actively identify those cases at an early stage. In this paper, we develop a machine learning framework for detecting PDD (psychological disorder detection) users, that is to say Social Network Mental Disorder (SNMD).This work aims to build a framework for detecting psychological disorders in social media users. We pursue to accomplish our complete method through:

1) Collection of Data
2) Cleaning and preprocessing of Data.
3) Extracting Features

Through manipulating machine learning techniques with the ground truth found via the current analytical repetition in Psychology, we extract and analyse several features of different categories from OSNs, including parasocial relationships, online and offline interaction ratio, social capital, disinhibiting, self-disclosure, and bursting temporal behaviour. These features capture important factors or serve as proxies for disorder detection.

A. Algorithms

Preprocessing Algorithms:

It is a technique that is used to translate the raw data into a clean data set. Each and every time the data is collected from different sources it is collected in fresh format which is not achievable for the study. For achieving better results from the realistic model in Machine Learning developments the format of the data has to be in a proper manner. So in data preprocessing is required because of the presence of unformatted real world data. In this algorithm, we will discuss the some steps involved in text processing.

1. Stop word Removal

In this process stop words are words that are mostly common in a text body and thus considered as rather un-informative (e.g., so, and, or, the...).

One approach to stop word removal is to search against a language-specific stop word dictionary. Another approach is to create a stop list by sorting all words in the entire text body by regularity. This stop list after conversion into a set of non-redundant words is then used to remove all those words from the input documents that are classified between the top n words in this stop list.

The algorithm is implemented as below given steps.

1) In document text is tokenized and separable words are put in array.
2) A single stop word is read from stop word list.
3) The stop word is matched to goal text in form of array using sequential search technique.
4) If it equals, the word in array is removed, and the evaluation is continued checkout length of array.
5) After removal of stop word totally, an additional stop word is read from stop word list and again algorithm go to step 2. The algorithm runs continuously up to all the stop words are compared.
6) Follow-on text empty of stop words is displayed, also mandatory data like stop word removed, no. of stop words removed from goal text, total count of words in target text, count of words in resultant text, separate stop word count originate in goal text is displayed.

2. Tokenization

In tokenization defines the common process of breaking down a text body into separate features that help as input for various natural language processing algorithms. Follows, tokenization processing steps:

1. Segmenting act of breaking up a sequence of series into parts such as phrases, words, keywords, symbols and other features called tokens
2. Tokens or words are separated by whitespace, punctuation marks or line breaks.
3. White space or punctuation marks may or may not be included depending on the need.
4. The tokens become the input for another process like parsing and text mining.

3. Stemming

Stemming is a procedure where words are reduced to a root by removing inflection through reducing unnecessary characters, usually a suffix. The results can be used to identify relationships and commonalities across large datasets.

1) A stemming algorithm is a process of linguistic normalization, in which the variant forms of a word are reduced to a common form for e.g. - 1. Played-play 2. Clustering– cluster
2) After stemming find eventually is that you can be improving performance of the language, while producing a parallel degradation of performance in another area.

Classification Algorithm:

This section will introduce some of the main concepts and procedures that are needed to apply the classification tasks. We have to used Naive Bayes algorithm is easy to build and particularly useful for very large data sets. Naive Bayes machine learning algorithm used for classification of personal and social features.

Steps:

1. Given training dataset D which consists of documents belonging to different class say Class A and Class B
2. Calculate the prior probability of class A=number of objects of class A/total number of objects
   Calculate the prior probability of class B=number of objects of class B/total number of objects.
3. Find NI, the total no of frequency of each class
   Na=the total no of frequency of class A
   Nb=the total no of frequency of class B
4. Find conditional probability of keyword occurrence given a class:
   P (value 1/Class A) =ni (A)
   P (value 1/Class B) =ni (B)
   P (value 2/Class A) =ni (A)
   P (value 2/Class B) =ni (B)
   ..
   ..
   ..
   P (value n/Class B) =ni (B)
5. Avoid zero frequency problems by applying uniform distribution.
6. Classify Document C based on the probability p(C/W) a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value 2/Class A). P(value n /Class A)
   b. Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class B). P (value n /Class B)
7. Assign document to class that has higher probability.

IV. SYSTEM ARCHITECTURE

In a system architecture we can detect user are in stress or not due to interaction social network. Facebook and twitter are examples of social network. In social network people become more willing to share moods on facebook. User can different posts on a facebook. There are two types of information that we can use as the initial inputs, i.e., Personal features, OSN (Online Social Network) features as well as social attention factors (being liked, commented,) of a single facebook post as shown in Fig. 1. User level posting behavior as summarized from a user monthly facebook postings, post time, post type; social interaction extracted from a user's social interactions with friends.

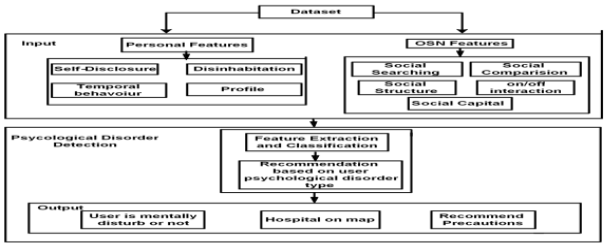

Fig. 1. System Architecture

Our system architecture, is divided in to the following main components: collecting data, preprocessing, features extraction and classification. We propose Psychological Disorder Detection (PDD) framework as shown in Figure 1.

In proposed system implementing an algorithm for identifying the psychological disorder. Classification algorithm- naive bayes classification is a general process related to categorization, the process in which objects are recognized, differentiated and understood.

V. EVALUATION

A. Experimental Settings

For evaluations, we considered the features of precision, recall, F-Measure and accuracy. Following all the detailed classifications are given below, where

TP represents obsessed user that are indicated as depressed, FP represents normal user that are indicated as depressed, TN represents normal users that are indicated as healthy, and FN represents obsessed user that are indicated as healthy.

1. Precision represents the ratio of the amount of truly obsessed user to those being identified as depressed, i.e.,

$$\text{Precision} = \frac{TP}{TP+FP}$$

2. Recall represents the percentage of obsessed user being successfully identified, i.e.,

$$\text{Recall} = \frac{TP}{TP+FN}$$

3. Accuracy is the percentage of the correct decision, i.e.,

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

4. F-Measure takes the evaluation point of harmonic mean of precision and recall, i.e.,

$$\text{F-Measure} = \frac{2TP}{2TP+FP+FN}$$

B. Experimental Results

In experimental results, we calculate our system using only text content and feature. We observed two types of features 1) OSN feature and 2) Personal feature detection result can also be improved by using Naive Bayes model. We first calculate our method by precision, recall, F-Measure and accuracy. We can observe disorder detection score of the unhappiness and healthy people with the online social networks data. For each subject, we collect user online facebook post as the input testing data, and extract the keywords for calculating the disorder detection. As for the next steps, we apply Naïve Bayes classifier to evaluate the psychological status of social networking user. After classification user are in stress, will be detected by PDD system is used to recommend hospitals on a map and at the same time admin will send mail for precaution. Fig. 2 shows the best performance with the Naïve Bayes classifier. Therefore, the best performances are in the precision (50.0%), recall (56.5%), and while the performances in the F-Measure (53.1%), accuracy (79.9%).

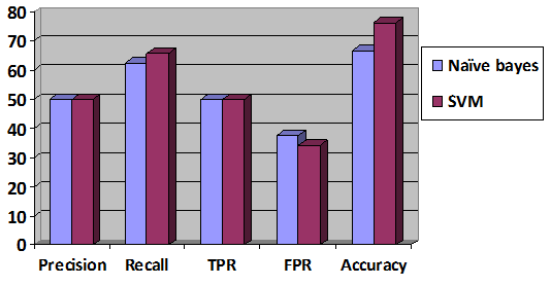

Fig. 2 shows the proposed method performance.

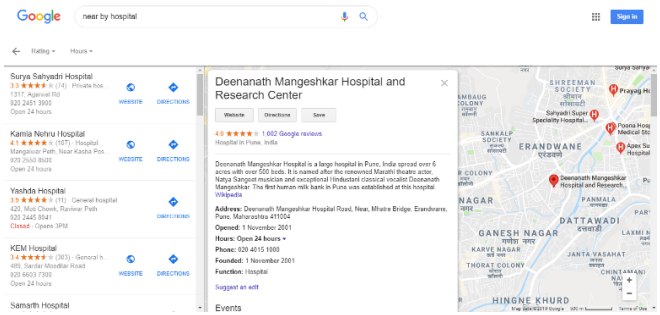

Fig. 3 Recommend hospitals on a map

VI. CONCLUSION

In this paper, we have studied a novel problem of psychological disorder detection in social networks. We propose a PDD framework which is Psychological Disorder Detection framework. This framework recognize mental confusion opportunity for proactive consideration. In this way we displayed a structure for recognizing user's psychological disorder states from user's web based life information. Using this framework, the proposed method can help to utilizing facebook post content just as user's social collaborations and utilizing genuine online networking information as the premise, we considered the relationship between user's psychological disorder states. In addition to this application, we can show the hospitals for further treatment on graph which will locate the shortest path from users current location to the hospital and recommend the user's for health precaution.

ACKNOWLEDGMENT

This work is supported in stress detection due to interaction of social networks. We would like to thanks Department of Computer, MMCOE, for support and encouragement for this project.

REFERENCES

1. B.Saha, T. Nguyen, D. Phung, and S. Venkatesh. "A framework for classifying online mental health-related communities with an interest in depression". *IEEE Journal of Biomedical and Health Informatics*, 2016.
2. Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen "Subconscious Crowdsourcing: A Feasible Data Collection Mechanism for Mental Disorder Detection on Social Media". *IEEE/ACM International Conference on Advance in Social Networks Analysis and Mining (ASONAM)* 2016.
3. 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.
4. Kohei Hayashi Takanori Maehara Masashi Toyoda Ken-ichi Kawarabayashi. "Real-Time Top-R Topic Detection on Twitter with Topic Hijack Filtering" 2015
5. Munmun De Choudhury, Scott Counts, Eric Horvitz, "Social Media as a Measurement Tool of Depression in Populations," *WebSci, May* 2013
6. H. Lin, J. Jia, Q. Guo, Y. Xue, J. Huang, L. Cai, and L. Feng. "Psychological stress detection from cross-media microblog data using deep sparse neural network." *In proceedings of IEEE International Conference on Multimedia Expo*, 2014.

7. Liqiang Nie, Yi-Liang Zhao, Mohammad Akbari, Jialie Shen, and TatSeng Chua. "Bridging the vocabulary gap between health seekers and healthcare knowledge". *Knowledge and Data Engineering, IEEE Transactions on*, 27(2):396409, 2015.
8. Elvis Saravia, 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
9. Keumhee Kang, Chanhee-Yoon, Eun Yi Kim (2016). "Identifying Depressive Users in Twitter Using Multimodal Analysis," *Big Comp*, 2016
10. Yuan Zhang, Jie Tang, Jimeng Sun, Yiran Chen, and Jinghai Rao. "Mood cast: Emotion prediction via dynamic continuous factor graph model." *2013 IEEE 13th International Conference on Data Mining*, pages 11931198, 2010.
11. Chih-Hua Tai, Zheng-Han Tan, Yung-Sheng Lin, Yue-Shan Chang. "Mental Disorder Detection and Measurement using Latent Dirichlet Allocation and SentiWordNet" *2015 IEEE International Conference on Systems, Man, and Cybernetics*. 2015.
12. Hong-Han Shuai, Chih-Ya Shen, De-Nian Yang, Yi-Feng Lan, WangChein Lee and Phlips S .Yu "Mining Online Social Data for Detecting Social Network Mental Disorders" *Proc. Int. Conf. World Wide Web*, 2016
13. Maryam Mohammed Aldarwish, Hafiz Farooq Ahmed, "Predicting depression levels using social media posts," *13th International Symposium on Autonomous Decentralized Systems*, 2017.
14. Ahmad Rauf Subhani, Wajid Mumtaz, Mohamad Naufal b Mohamad Saad, Nidal Kamel and Aamir Saeed Mali, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels" *2017 IEEE Acces.*
15. 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.
16. Hong-HanShuai, Chih-Ya Shen, De-NianYang, Senior Member, IEEE, Yi-Feng Lan, Wang-Chien Lee, Philip S. Yu. "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining" *2018 IEEE Transaction on knowledge and data engineering*, vol. 28, no. 4, April 2018.
17. Maryam Mohammed Aldarwish, Hafiz Farooq Ahmed, "Predicting depression levels using social media posts," 13th International Symposium on Autonomous Decentralized Systems, 2017.
18. Ahmad Rauf Subhani, Wajid Mumtaz, Mohamad Naufal b Mohamad Saad, Nidal Kamel and Aamir Saeed Mali, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels" 2017 IEEE Acces.
19. 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.
20. Hong-HanShuai, Chih-Ya Shen, De-NianYang, Senior Member, IEEE, Yi-Feng Lan, Wang-Chien Lee, Philip S. Yu. "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining" 2018 IEEE Transaction on knowledge and data engineering, vol. 28, no. 4, April 2018.

AUTHORS PROFILE

Miss. Punam B. Nalinde, pursuing ME degree in the department of Computer Engineering, from Pune University, post graduating in the year 2019. His research interests includes machine learning and data mining.

Prof. Anita Shinde, Received M.E degree in the department of Computer Engineering and PhD Pursuing. Her research interests includes data mining, Information Retrieval and IOT. She is an Assistant Professor for the Department of Computer Engineering, Marathwada Mitra Mandal college of Engineering, Karvenagar Pune.