Stress Detection Methodology based on Social Media Network: A Proposed Design

S. M. Chaware, Chaitanya Makashir, Chinmayi Athavale, Manali Athavale, Tejas Baraskar

Abstract: Mental disorders can be recognized by how a person behaves, feels, perceives, or thinks over a period of a lifetime. Nowadays, a large number of people are feeling stressed with the rapid pace of life. Stress and depression may lead to mental disorders. Work pressure, working environment, people we interact, schedule of the day, food habits, etc. are some of the major reasons behind building stress among the people. Thus, stress can be detected through some conventional medical symptoms such as headache, rapid heartbeats, feeling low energy, chest pain, frequent colds, infections, etc. The stress also may reflect in normal behavior while carrying out day-to-day activities. Individuals may share their day-to-day activities and interact with friends on social media. Thus, it may be possible to detect stress through social network data. There are many ways to detect stress levels. Some of the instruments are used to detect stress while there is a medical test to know the stress level. Also, there are apps that analyze the behavior of the person to detect stress. Many researchers had tried to use machine learning techniques including the use of various algorithms such as Decision Tree, Naïve Bayes, Random Forest, etc. which gives a lower accuracy of 70% on average. In this paper, we are using a closeness of stress levels with social media data shared by many users. In our proposed system design, Facebook posts are being accessed using a token. Further, we recommend the use of machine learning algorithms such as Conventional Neural Network (CNN) to extract Facebook posts, Transductive Support Vector Machine (TSVM) to classify posts and K-Nearest Neighbors (KNN) to recommend nearby hospitals. With the help of these algorithms, we predict the stress level of the person as positive, negative. Thus, we are expecting more accuracy to detect the stress along with the preventive recommendation. We have proposed a methodology to detect stress because severe stress may lead to self-harming activities and also it may affect the lives of people around us. Thus, stress detection has become extremely important and we are expecting that our proposed model may detect it with more accuracy.

Keywords: Social Media, Mental Disorder, Conventional Neural Network, Transductive Support Vector Machine, K-Nearest Neighbors, Facebook

I. INTRODUCTION

Mental disorders are threatening people's health.

Revised Manuscript Received on January 30, 2020.
* Correspondence Author
S. M. Chaware*, Department of Computer Engineering, Savitribai Phule Pune University, Pune, (Maharashtra), India.
Chaitanya Makashir, Department of Computer Engineering, Savitribai Phule Pune University, Pune, (Maharashtra), India.
Chinmayi Athavale, Department of Computer Engineering, Savitribai Phule Pune University, Pune, (Maharashtra), India.
Manali Athavale, Department of Computer Engineering, Savitribai Phule Pune University, Pune, (Maharashtra), India.
Tejas Baraskar, Department of Computer Engineering, Savitribai Phule Pune University, Pune, (Maharashtra), India.

They are considered to be a major factor of change the mood of a user and the user goes into a depression.

Nowadays users can be stressed due to social interactions of social networks. The rapid increase of mental disorders or stress has become a great challenge to human health and quality of life. It is difficult to timely detect mental disorders or stress for proactive care. Thus, there is significantly important to detect mental disorder before it turns into severe problems.

Our proposed design join hands to detect stress to avoid further consequences such as going into depression, self-harming acts, etc. Once stress is detected, people can take the help of stress management methodologies such as meditation, 'smile and laugh', reading motivational books, etc. A person can also follow proper treatment suggested by doctors, consultants. But for this, there is a need to suggest nearby hospitals so that a person gets help as quickly as possible.

There are also some techniques that are implemented to detect the mental state of mind using different machine learning algorithms. For this, real-world social media data has been analyzed. But algorithms like Decision Tree, Naïve Bayes, Random Forest failed to achieve expected accuracy. These algorithms gave an approximate accuracy of 70%.

II. LITERATURE REVIEW

Nowadays people are constantly using social media to reflect their lives over the internet. Social media platforms like Facebook, Twitter, Snapchat, Instagram, LinkedIn, Tumblr, Pinterest, etc. engage people more than one-to-one human interactions. Though social media has provided a platform to facilitate the sharing of thoughts, feelings, career interests, etc. on the internet, unfortunately, it's overuse leads to addiction to social media and stress.

Research says that symptoms of mental disorder can be noticed from interactions over social media so that delays in treatment can be avoided. The emphasis is on Cyber-Relationship addiction, Net compulsion, Information overload to detect social network mental disorders [1]. Features like social relationships, self-disclosure or self-esteem, loneliness, bursting temporal behavior, etc. are analyzed. To build the SNMD-based Tensor Model, the Transductive Support Vector Machine (TSVM) is used that gave an accuracy of 84.3%. Mining online social behavior provides an opportunity to detect mental disorders based on features extracted from data logs of online social networks [1].The main emphasis of previous studies is on the classification of emotions of tweets, posts gathered from social media platforms like Twitter, Facebook. This is because these platforms are the most frequently used platforms. Preprocessing includes classification of a dataset into a training dataset and testing dataset to carry out tokenization further [2].

Next to it, pre-processing of tweets is done which includes removing handles, removing URLs, timings of tweets, #hashtag, etc. [2]. Support Vector Machine (SVM) and Decision tree algorithms are implemented to obtain positive or negative results.

Thus, the accuracy of 82% for SVM is obtained. Comparative study of different classification algorithms such as Naïve Bayes, Decision Tree, Random Forest has done for sentiment analysis of the social networks [3]. The pre-processing phase involves many steps such as removing noise, remove duplicate tweets, remove punctuation, remove stop words, tokenization, join words. A Hybrid model [CNN + CNN (word to vector) + RNN (word to vector)] has obtained the highest accuracy of 83.6% [3]. Results are based on Accuracy, Sensitivity, Specificity of a particular classifier. But RNN-LSTM also performs nearly well as a Hybrid model.

RNN-LSTM gains an accuracy of 82.3% which is quite nearer to the Hybrid model. Researchers have performed an analysis on a dataset containing standard Arabic dialets to identify risky behavior and depressive state. The study focuses on depressive behavior and self-harm activities in the Middle East and North Africa [4]. Also, various natural language processing (NLP) tools and techniques have been used to extract linguistic patterns and perform sentiment analysis on tweets. The methodology includes the use of a machine learning approach such as Support Vector Machine (SVM), Deep Neural Network (DNN). This identifies risky behavior and also provides insights to explain human behavior. But it fails to specify security issues and related solutions. Previous studies show that social media network defines various aspects of social behavior. Thus, researchers have also focused on tweets related to Asthma, Cancer, Diabetes and identified generic terms that refer to target diseases. The study includes a classification of extracted data and evaluates results such as 'which disease has higher posts than others?' [5]. It considers high-frequency keywords including attack, allergy for Asthma, breast, awareness for Cancer, type, sugar for Diabetes. For the construction of these keywords, GPS coordinates and unique queries are used that include lexicon and filter. It focuses on the most commonly identified diseases like Asthma, Diabetes, Cancer but does not focus on mental health and their disorders which are rare to identify. But results are not validated against benchmarks. The focus is on particular geographical areas and hence the same methodology may not be applicable for other geographical areas [5].

For stress detection, researchers also have proposed text sentiment analysis general framework which includes various stages for opinion mining [6]. In this, after data collection and acquisition phase, tokenization is performed. Feature extraction is done for sentiment detection which includes opinions, views, beliefs. These sentiments are classified and the result is visualized in the form of graphs. Different algorithms such as Decision Tree, KNN, SVM, Naïve Bayes, Random Forest, ANN are discussed in the context of sentiment analysis. It provides a taxonomy of sentiment analysis tasks and introduces to different levels of sentiment analysis and its techniques[6]. Though the proposed framework is generalized, it is only compatible with machine learning methods. Thus, it is not able to provide different frameworks required for other methods. Also, it is found that different approaches and algorithms have direct influences on the overall classification [6].

Analysts have also proposed a simple tree stage framework [7]. The first stage involves the extraction of embedded noun phrases. The second stage involves identifying the medical concepts from these noun phrases based on their specificity. Then detected medical concepts are normalized to terminologies. This idea bridges the vocabulary gap between health seekers and healthcare knowledge. The accuracy provided by the support vector machine is up to 85%. It needs unstructured medical content to be more flexibly organized. Scientists have constructed a three-level framework to detect psychological stress by obtaining a set of low-level features from the tweets [8]. In the first level, low-level features such as texts, images, comments, retweets, and favorites are extracted. Middle-level is defined and built based on psychological and art theories. The last level designs Deep Sparse Neural Network that helps to learn the stress categories. From social networks, the proposed system helps to automatically detect psychological stress [8]. The proposed normalized approach gives an accuracy of 82%. As it uses cross-media microblog data, it is efficient. But the system has not used Facebook data for detecting psychological stress.

Common Limitations found are:

- Security is not taken into consideration, which is one of the important issues. Exposing private data is a threat. Hence there must be a focus on data security and privacy as well.
- While dealing with a sensitive topic like stress, the accuracy of stress detection plays an important role. It decides how accurately your model detects stress. The algorithm used like Decision Tree has very little accuracy i.e. 67%.
- The focus of previously proposed systems is on mental state detection and on the detection of diseases like Asthma, Diabetes, Cancer. There is no focus on recommending preventive or remedial measures to the user who is using the system.

Also, referred news articles gave more information about the reasons behind stress, how stress affects, etc. [10], [13]. Most of the times stress leads to depression. Stress has a negative impact on our daily routine, way of behavior [11][12]. Thus, stress detection is important to avoid further consequences and also provide measures for proactive care [14][15].

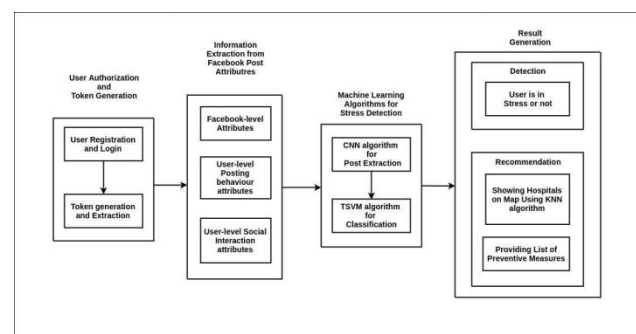### III. PROPOSED DESIGN

Fig. 1. Prosed Design

On social networking sites like Facebook or Twitter, users interact with other people and upload different posts on such sites. This data generated from social media sites can help in detecting whether the person is in stress or not. Thus, the proposed framework includes the detection of stress through analysis of social media data as well as provision for proactive care by recommending hospitals and preventive measures.

The proposed framework in fig. 4.1 represents a stress detection methodology based on the social media network. Thus, the framework includes the extraction of Facebook posts and analysis of those posts to detect the stress level.

The proposed design incorporates four modules.

1. User Authorization and Token Generation

User has to be an authorized member of the system. User must have a Facebook account already so that he/she can generate and extract token.

I. User Registration and Login

User has to register in the system only once. User has to log in the system. Then the user is supposed to insert token by following some steps.

II. Token Generation and Token Extraction

User has to visit Facebook's developer site. Hereafter login, the user has to create an app if he/she is visiting for the first time; otherwise, the user can directly extract token. Then, the user has to extract token by selecting certain permissions. Thereafter, the user has to copy token and paste the token at a provided place in the GUI of the system. For every next time, the user has to extract token while monitoring the mental state. To restrict access to any unauthorized person, the user has to logout so that the data gathered will be erased.

2. Information Extraction

There are three types of information that we can use as the initial inputs. Those are Facebook-level attributes, user-level posting behavior attributes, and user-level social interaction attributes.

a. Facebook-level attributes describe the linguistic i.e. positive and negative words and visual content like brightness, cool color, dull color, as well as social attention factors (being liked, commented) of a single Facebook post.

b. User-level posting behavior attributes are summarized from user's Facebook postings, post time, post type.

c. User-level social interaction attributes are extracted from a user's social interactions with friends. In particular, social interaction attributes can further be divided into

(i) Social interaction content attributes like words and emotions can be extracted from the content of the user's social interactions with friends.

(ii) Social interaction structure attributes can be extracted from the structures of the user's social interactions with friends.

3. Machine Learning Algorithms for Stress Detection

This system mainly uses machine learning algorithms such as Convolutional Neural Network (CNN), Transductive Support Vector Machine (TSVM).

I. CNN algorithm for post-extraction

CNN algorithm is used to extract Facebook posts.

II. TSVM algorithm for classification

TSVM algorithm is used to classify obtained Facebook posts into two categories namely, positive and negative. If words like happy, good, pretty, etc are present then that describes the positive state of the post. Similarly, if words like bad, sad, not good, etc. exist then those words represent the negative state of that respective post. This is how gathered posts are classified for further analysis based on lists of positive and negative words. This is how the framework detects whether the user is in stress or not.

4. Result Generation

This stage includes two parts which are Detection and Recommendation. This module plays a significant role as users can view results about their mental state here.

I. Detection

Here, the user views his/her result that whether he is in stress or not as a result. The result is available graphically as well as post-wise.

Recommendation

This module is to be designed in such a way that the system can send a list of preventive measures to the user via Email. Depending upon the shortest path from the current location of the user to the hospital, the system recommends the nearest hospitals to the user on the map. K- Nearest Neighbour (KNN) algorithm is to be used for recommendation.

IV. CONCLUSION

Psychological stress has become a threat to people's health. It is non-trivial to detect stress timely for proactive care. Stress generally leads to irrational irritation, a distraction from work, depression, etc.. This stress may arise from the work environment, pressure of study, over-accessing of social media. Also, nowadays we are used to social media more than before. Thus, we can say that we interact more on social media. This enables us to process a huge amount of data and detect whether the user is stressed or not. Many researchers have used various machine learning algorithms to process huge data generated from social media to detect stress. From Table- I, we found that the TSVM algorithm best classifies data than machine learning algorithms like Naïve Bayes, Random Forest, Decision Tree, etc..

Table- I: Comparison of machine learning algorithms for stress detection

Sequence No	Algorithm used with Accuracy	
	Algorithm	*Accuracy*
1	TSVM	84.2%
2	SVM	82%
4	Naïve Bayes	77.5
5	Random Forest	73.8
6	Decision Tree	72.5
3	Adaboosted D-Tree	67%

Therefore, we have presented a design for detecting users' psychological stress states from users' social media data,

leveraging Facebook post's content as well as users' social interactions. On the basis of real-world social media data, it is found that users' psychological stress states have a correlation with their social interaction behaviors. Thus, the facility of recommendations of preventive measures via mail is included in the proposed design. The proposed design also includes the facility of showing the hospitals for further treatment on a graph that locate the shortest path from the current location of the user to that hospital. Finally, more accuracy is expected to be achieved using machine learning algorithms and proposed techniques.

ACKNOWLEDGMENT

REFERENCES

1. Hong-Han Shuai, Chih-Ya Shen, De-Nian Yang, Yi-Feng Carol Lan, Wang-Chien Lee, Philip S. Yu, Ming-Syan Chen, "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining", IEEE Transactions on Knowledge and Data Engineering, Pages: 1212 – 1225, Year: 2018, Volume: 30, Issue: 7, Journal Article, Publisher: IEEE
2. Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, "Sentiment Analysis of Tweets Using Machine Learning Approach", 2018 Eleventh International Conference on Contemporary Computing (IC3), Pages:1-3, Year: 2018, Conference Paper, Publisher: IEEE
3. Mohammed H. Abd El-Jawad, Rania Hodhod, Yasser M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning", 2018 14th International Computer Engineering Conference (ICENCO), Pages:174-176, Year: 2018, Conference Paper, Publisher: IEEE
4. Wajdi Zaghouani, "A Large-Scale Social Media Corpus for the Detection of Youth Depression (Project Note)", Procedia Computer Science, Volume 142, 2018, Pages: 347-351
5. Alvaro Esperanca, Zina Ben Miled, Malika Mahoui, "Social Media Sensing Framework for Population Health", 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Pages:0298-0304, Year: 2019, Conference Paper, Publisher: IEEE
6. Hoong-Cheng Soong, Norazira Binti A Jalil, Ramesh Kumar Ayyasamy, Rehan Akbar, "The Essential of Sentiment Analysis and Opinion Mining in Social Media: Introduction and Survey of Recent Approaches and Techniques", 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Pages:272-277, Year: 2019, Conference Paper, Publisher: IEEE
7. B. K. Bhavitha, A. P. Rodrigues, N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis", 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2017,pp.216-221.
8. Liqiang Nie, Yi-Liang Zhao, Mohammad Akbari, Jialie Shen, Tat-Seng Chua, "Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge", IEEE Transactions on Knowledge and Data Engineering, Pages:396 – 409, Year: 2015. Volume: 27, Issue: 2, Journal Article, Publisher: IEEE
9. Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Jie Huang, Lianhong Cai, Ling Feng, "Psychological stress detection from cross-media microblog data using Deep Sparse Neural Network", 2014 IEEE International Conference on Multimedia and Expo (ICME), Pages:1-6, Year: 2014, Conference Paper, Publisher: IEEE.
10. Web References: https://developers.facebook.com/
11. 'Mental illness is devastating lives! Do we really care?', Times Of India, 10 October 2019, Accessed 25 August 2019, Retrieved from https://timesofindia.indiatimes.com/life-style/health-fitness/de-stress/mental-illness-is-devastating-lives-do-we-really-care/articleshow/71503619.cms
12. 'Young people suffer from mental illness', Maharashtra Times, 27 June 2019, Accessed 12 September 2019, Retrieved from https://maharashtratimes.indiatimes.com/maharashtra/kolhapurwesternmaharashtranews/kolhapur/indias-most-of-youngster-indepression/articleshow/69964319.cms
13. 'Stress causes increased psoriasisa', Maharashtra Times, 7 January 2019, Accessed 2 October 2019, Retrieved from https://maharashtratimes.indiatimes.com/maharashtra/punenews/stresscausesincreasedpsoriasis/articleshow/71825166.cms
14. 'Double disappointment among girls due to social media', Maharashtra Times, 7 January 2019, Accessed 25 August 2019, Retrieved from https://maharashtratimes.indiatimes.com/international/internationalnews/dabblestressbecauseofsocialmedia/articleshow/67414567.cms
15. '89 percent of India's population suffering from stress; most don't feel comfortable talking to medical professionals', The Economic Times, 10 July 2018, Accessed 25 August 2019, Retrieved from https://economictimes.indiatimes.com/magazines/panache/89-per-cent-of-indias-population-suffering-from-stress-most-dont-feelcomfortabletalkingtomedicalprofessionals/articleshow/64926633.cms

AUTHORS PROFILE

S. M. Chaware, He has pursued Ph.D. from NMIMS University in 2012 and has 41 publications to his credit in SCI/SCOPUS/WOS International journals and conferences. His areas of interest are Big data Analytics, IoT and Security. He is a life member of CSI and ISTE and also, he is in editorial and review board member of many International Journals.

Chaitanya Makashir, He is pursuing a bachelor of computer engineering from Savitribai Phule Pune University. His areas of interest are the Internet Of Things, Artificial Intelligence, Machine Learning. He has done some projects in the field of machine learning. He is a member of CSI and ISTE.

Chinmayi Athavale, She has recently published a paper titled 'Obstacle Detection & Gate Automation at Railway Crossings' in UGC approved journal which emphasizes working in the IoT domain. Her areas of interest are IoT, Machine Learning, Artificial Intelligence.

Manali Athavale, She has recently published a paper titled 'Obstacle Detection & Gate Automation at Railway Crossings'in UGC approved journal which emphasizes working in the IoT domain. Her areas of interest are IoT, Artificial Intelligence, Machine Learning.

Tejas Baraskar, He is pursuing a bachelor of computer engineering from Savitribai Phule Pune University. He has recently published a paper titled 'IoT Based Home Automation' in the IJRASET journal. His areas of interest are the Internet Of Things, Machine Learning, Artificial Intelligence.