International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

An automatic adolescent psychological pressure detection system using advanced data mining techniques

Miss.Alaknanda Avinash Deshmukh and Prof. M.D. Ingle

Department of Computer Engineering JSPM's Jayawantrao Sawant College of Engineering Savitribai Phule Pune University Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Sentiment analysis alludes to the application for processing natural language, content investigation, computational etymology to deliberately perceive, remove, evaluate, and learn full of feeling states and emotions. Twitter, being one among a few well known web based life stages, is where individuals frequently decide to express their feelings and notions about personal life and other things.Advanced data mining techniques could give powerful algorithms and frameworks to a target appraisal and observing of mental issue and, specifically of adolescent psychological pressure. In this paper, the application of sentiment analysis methodologies to psychological pressure detection and prediction are discussed. In addition, a fundamental plan of an incorporated multimodal framework for psychological pressure checking, that incorporates estimation investigation and full of feeling processing strategies, is proposed. In particular, the paper traces the fundamental issues and moves comparative with the structure of such a framework.

Keywords: *Psychological pressure, text mining, sentiment analysis, social media, machine learning.*

Introduction

The number of active online social network (OSN) users has grown considerably, and some studies indicate there will be 2.95 billion users by the end of 2020 [1]. This high number of users, on OSN, is mainly due to the increase of the number of mobile devices, such as smartphones and tablets, connected to the Internet. Currently, OSN have become a rich and universal means of opinion expression, feelings, and they reflect the bad habits or wellness practices of each user. In recent years, the analysis of the messages posted on OSN have been used by many applications [2], [3] in the industry of health care informatics. The sentiments and emotions, expressed on the messages posted on OSN, provide clues to different aspects of the behavior of users; for instance, sentences containing wordswith negative meaning may indicate sadness, stress, or dissatisfaction [4]. Conversely, it can be inferred that if a person is in a positive mood state, this person can be more self-confident and emotionally stable [5]. users have various practices on OSN, if the assessment force estimation of posted sentences stay at low levels, or in the event that it every now and again changes from high to low levels and the other way around, these realities can show a few passionate unsettling influence, for example, sadness or stress[6] occasions saw that clients compose short sentences when they are encountering a time of sorrow. Likewise, these clients utilize the principal individual pronoun in their sentences and experience the ill effects of incessant a sleeping disorder. Therefore, their behavior can be reflected in the sentences posted on OSN. The presence of certain words in the sentences can be monitored and analyzed to identify users at a high risk of attempting suicide and an appropriate intervention can take place. Mental pressure is one of the most prevalent mental disorders in all regions and cultures around the world [10]. Unfortunately, depression recognition rate remains low. Most of the studies about health systems use sensor devices to detect mental disorders. In , the proposed trained[1] classifier, which is trained using electroencephalogram signals, is able to detect stress with an average accuracy of 80.45% using 4- fold cross validation.

In, authors use heart rate variability data to propose a classification model that considers different stress levels, baseline, mild stress and severe stress, reaching accuracy values of 74%, 81%, 82%, respectively. In this context, the main goal of this work is to introduce an mental pressure detection that uses an approach named automatic adolescent psychological pressure, which aggregates an ontology collection for health scenarios, which is not addressed in other proposed system designed to improve emotional health. The proposed system also includes the sentiment analysis approach and an emotional health monitoring system. The monitoring system filters sentences from an OSN that allows to identify potential users with mental pressure conditions. To

accomplish this task, an objective method based on an naive bayes is used to detectpotential psychological disorders. Later, a KBRS is activated to send happy, calm, relaxing, or motivational messages to these users. These messages have different intensity levels depending on the sentiment intensity of the sentences posted on an OSN, which is determined by an enhanced sentiment analysis metric.

Review of Literature

Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, Alex (Sandy) Pentland- In our paper, propose an elective methodology giving proof that day by day stress can be dependably perceived dependent on social measurements, got from the client's cell phone movement and from extra markers, for example, the climate conditions (information relating to transient properties of the earth) and the character attributes (information concerning perpetual attitudes of people). Our multifactorial factual model, which is individual free, acquires the precision score of 72.28% for a 2-class day by day pressure acknowledgment issue. The model is productive to actualize for a large portion of sight and sound applications because of exceptionally decreased low dimensional element space (32d). Besides, we recognize and talk about the markers which have solid prescient force. Budhaditya Saha, Thin Nguyen, Dinh Phung, Svetha Venkatesh - Mental behavior deeply affects people, families, and by expansion, society in general. Interpersonal organizations permit people with mental scatters to speak with others sufferers by means of online networks, giving a priceless asset to examines on literary indications of mental medical issues. Mental clutters regularly happen in blends, e.g., a patient with a nervousness issue may likewise create sadness M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri - In this paper, author propose an integrated social media content analysis platform that leverages three levels of features, i.e., user-generated content, social graph connections, and user profile activities, to analyze and detect anomalous behaviors that deviate significantly from the norm in large-scale social networks. Several types of analyses have been conducted for a better understanding of the different user behaviors in the detection of highly adaptive malicious users. Huijie Lin, Jia Jia, Jiezhon Qiu, Yongfeng Zhang, Lexing Xie, Jie Tang, Ling Feng, and Tat-Seng Chua - In this paper, find that clients stress state is firmly identified with that of his/her companions in internet based life, and we utilize an enormous scale dataset from genuine social stages to methodicallly contemplate the relationship of clients' pressure states and social cooperations. We initially characterize a lot of pressure related printed, visual, and social properties from different angles, and afterward propose a novel half and half model - a factor diagram model joined with Convolutional Neural Network to use tweet substance and social cooperation data for stress recognition. Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen - In this paper, at building prescient models that influence language and standards of conduct, utilized especially in online networking, to decide if a client is experiencing two instances of mental issue. These prescient models are made conceivable by utilizing a novel information assortment process, authored as Subconscious Crowdsourcing, which assists with gathering a quicker and progressively solid dataset of patients. Our tests recommend that extricating explicit language examples and social association highlights from dependable patient datasets can significantly add to advance examination and recognition of mental issue. A. E. U. Berbano, H. N. V. Pengson, C. G. V. Razon, K. C. G. Tungcul, and S. V. Prado - The paper exhibits further research on neural designing that centers around the grouping of enthusiastic, mental, physical and no worry using Electroencephalography (EEG) signal examination. Stress is one of the main sources of a few wellbeing related issues and infections. Accordingly, it gets fundamental for individuals to screen their pressure. The human body obtains and reacts to worry in various manners coming about to two orders of pressure to be specific, mental and passionate pressure. Customary techniques in characterizing pressure, for example, through polls and self-appraisal tests are said to be abstract since they depend on close to home judgment. Subsequently, right now, is characterized through a target measure which is EEG signal investigation. The highlights of the EEG accounts are then pre-handled, extricated, and chose utilizing Discrete Wavelet Transform (DWT). These highlights are then ussed as contributions to characterize pressure utilizing Artificial Neural Network (ANN) and approved utilizing K-overlay Cross Validation Method. In conclusion, the outcomes from the product helped technique is contrasted with the consequences of the customary strategy. I.-R. Glavan, A. Mirica, and B. Firtescu - Social media tools are wide spread in web communication and are gaining popularity in the communication process between public institutions and citizens. This examination leads an investigation on how online life is utilized by Official Statistical Institutes to cooperate with residents and disperse data.

A direct relapse system is performed to look at which web based life stages (Twitter or Facebook) is a progressively successful apparatus in the correspondence procedure in the official measurements region. Our investigation proposes that Twitter is a more incredible asset than Facebook in upgrading the connection between legitimate measurements and residents, consenting to a few different examinations. Next, played out an investigation on Twitter organize attributes talking about ||authentic insights|| utilizing NodeXL that

uncovered the unexploited capability of this system by legitimate factual organizations.

Bimal Viswanath Alan Mislove Meeyoung Cha Krishna P. Gummadi – In this paper, study the development of movement between clients in the Facebook informal organization to catch this thought. Additionally find that joins in the action arrange will in general go back and forth quickly after some time, and the quality of ties displays a general diminishing pattern of movement as the interpersonal organization interface ages. For instance, just 30% of Facebook client sets cooperate reliably starting with one month then onto the next. Strikingly, and locate that despite the fact that the connections of the movement organize change quickly after some time, many chart theoretic properties of the action arrange stay unaltered. III. PROPOSED METHODOLOGY In the proposed systemic approach, we formulate the task as a classification problem to detect four types of detection of psychological pressure in social networks using the sentiment analysis and machine learning framework. An innovative solution to monitor and detect potential users with emotional disorders, according to the classification of sentences with psychological pressure contents.

A. Architecture

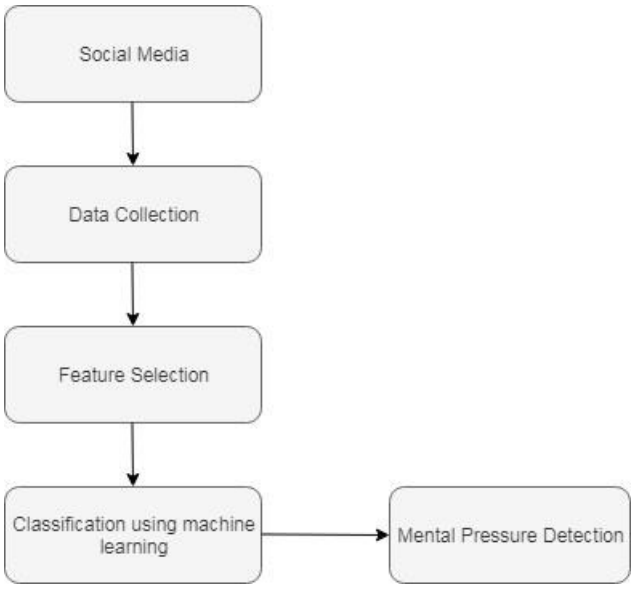

Fig. 1. Proposed System Architecture

• Social Media Data Collection - In data collection, social post are collected from popular social media websites using APIs. The datasets have different numbers of posts in each mood.

• Data Preprocessing - Data preprocessing removes redun- dancy and ambiguity inherit in the data and transforms the posts into sentences to facilitate sentence-level clas- sification. First, sentences are extracted by identifying the delimiters (e.g. dot, exclamation or question mark). Next, redundant information, e.g. duplicate sentences, is removed. Finally, ambiguous, vague or misspelled terms are corrected.

• Feature Selection - We take into account a variety of syntactic, semantic, and contextual features derived from the social media text.

• Data classification - A machine learning algorithm clas- sifies each aspect in a social media post and find mental pressure by considering all aspects and emotions and their linkages to sentiment words. machine learning algorithms are very efficient and helpful.

B. Algorithm

Preprocessing Algorithms

1. Stop word Removal-This technique removes stop words like is, are,they,but etc.
Initialize i,j
for i=1 to no of words in documents for j=1 no of words in stopword list
if
Words(i)==Stopwords(j) then eliminate words(i) end if end for
2. Tokenization-This technique removes Special character and images.
Initialize feature vector bg feature =[0,0..0] for token in text.tokenize() do if token in dict then token idx=getindex(dict,token) bg feature[token idx]++
else continue end if end for
3. Stemming– Removes suffix and prefix and Find Original words for e.g.- 1. played – play 2.Clustering - cluster
The word w Input = Normalize(input) if normalizeValidate(input) then return input;
for each rule in rules do if input match with rule then
Stem = ExtractStem(input,rules) if not TestStemLength(Rule)
then end for return input

Classification Algorithm:

Step 1: Convert the data set into a frequency.
Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.
Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.
For example:
Problem: Players will play if weather is sunny. Is this statement is correct?
We can solve it using above discussed method of posterior probability.
P(Yes — Sunny) = P(Sunny — Yes) * P(Yes) / P (Sunny) Here we have P (Sunny —Yes) = 3/9 = 0.33, P(Sunny) =
5/14 = 0.36, P(Yes)= 9/14 = 0.64
Now, P (Yes — Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems 4) Performance Evaluation: The evaluation can be done based on following factors:
i) Performance matrices such as TPR FPR Precision Recall etc.
ii) Impact of spam to Non-spam ratio iii) Impact of Different Sampling method iv) Investigation of time related data

C. Mathematical Equations

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) =count/ni (A) P (value 1/Class B) =count/ni (B) P (value 2/Class A) =count/ni (A) P (value 2/Class B) =count/ni (B)
. . . .
. . . .
. . . .
P (value n/Class B) =count/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.

D. Dataset

We use Facebook real-time dataset.

How to extract Facebook data:
1. Go to https://developers.facebook.com/ website2. More
3. Tools
4. Graph API Explorer
5. Login to Facebook
6. After Login click on button Get Access then select the
Facebook permission which we want
7. After getting permission get access token and copy that access token in above Token Field 8. After entering the token user get Facebook post

Results and Discussion

Experimental evaluation is done to compare the proposed system with the existing system for evaluating the performance. The simulation platform used is built using Java framework (version jdk 8) on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application.

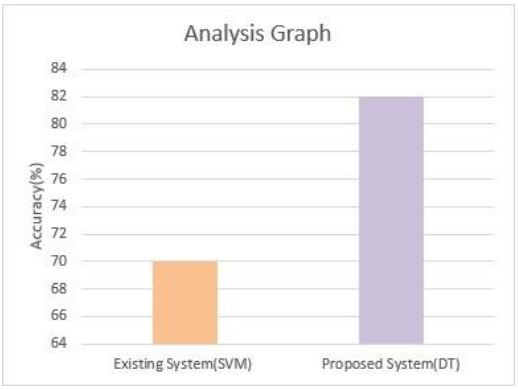

Fig. 2. Graph

Table 1: Comparative Result

Sr. No.	Existing System	Proposed System
1	70%	82%

References

[1]. M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri, ―Leveraging analysis of user behavior to identify malicious activities in large-scale social networks,‖ IEEE Transactions on Industrial Informatics, vol. 14, no. 2, pp. 799–813, Feb 2018.
[2]. H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T. S. Chua, ―Detecting stress based on social interactions in social networks,‖ IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 1820–1833, Sept 2017.
[3]. Budhaditya Saha, Thin Nguyen, Dinh Phung, Svetha Venkatesh‖ A Framework for Classifying Online Mental Health Related Communities with an Interest in Depression‖ IEEE 2016.
[4]. Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen‖ Subconscious Crowd- sourcing: A Feasible Data Collection Mechanism for Mental Disorder Detection on Social Media‖ 2016 IEEE/ACM
[5]. Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, Alex (Sandy) Pentland,‖ Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits‖ IEEE Conference 2015
[6]. Bimal Viswanath Alan Mislove Meeyoung Cha Krishna *† †‡ †* P. Gum- madi,‖ On the Evolution of User Interaction in Facebook‖ ACM 2011 [7] I.-R. Glavan, A. Mirica, and B. Firtescu, ―The use of social media for communication.‖ Official Statistics at European Level. Romanian Statistical Review, vol. 4, pp. 37–48, Dec. 2016.

[7]. E. U. Berbano, H. N. V. Pengson, C. G. V. Razon, K. C. G. Tungcul, and S. V. Prado, ―Classification of stress into emotional, mental, physical and no stress using electroencephalogram signal analysis,‖ in 2017 IEEE

[8]. International Conference on Signal and Image Processing Applications (ICSIPA), Sept 2017, pp. 11–14.

[9]. J. Ham, D. Cho, J. Oh, and B. Lee, ―Discrimination of multiple stress levels in virtual reality environments using heart rate variability,‖ in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), July 2017, pp. 3989–3992.

[10]. Y. Xue, Q. Li, L. Jin, L. Feng, D. A. Clifton, and G. D. Clifford, ―Detecting adolescent psychological pressures from micro-blog,‖ in [11] Health Information Science. Springer International Publishing, 2014, pp. 83–94.

[11]. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, ―Recognizing depression from twitter activity,‖ in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 3187–3196.

[12]. 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 HealthInformatics, 2016.

[13]. Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen ―Subconscious Crowd- sourcing: A Feasible Data Collection Mechanism for Mental Disorder Detection on Social Media‖ 2016.

[14]. J. Gill and G. King. What to do when you're Hessian is not invertible: Alternatives to model specification in nonlinear estimation. Sociological Methods and Research, 2004.

[15]. Kohei Hayashi † Takanori Maehara § Masashi Toyoda Ken-ichi Kawarabayashi ―Real-Time Top-R Topic Detection on Twitter with Topic Hijack Filtering‖,2015

[16]. B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. WOSN, 2009.

[17]. 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.

[18]. Liqiang Nie, Yi-Liang Zhao, Mohammad Akbari, Jialie Shen, and Tat-Seng Chua.‖ Bridging the vocabulary gap between health seekers and healthcare knowledge.‖ Knowledge and Data Engineering, IEEE Transactions on, 27(2):396–409, 2015.

[19]. Chi Wang, Jie Tang, Jimeng Sun, and Jiawei Han.‖ Dynamic social influence analysis through time-dependent factor graphs.‖ Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on, pages 239 – 246, 2011.

[20]. Lexing Xie and Xuming He. ―Picture tags and world knowledge: learning tag relations from visual semantic sources. ―In ACM Multimedia Conference, pages 967–976, 2013.

[21]. Yuan Zhang, Jie Tang, Jimeng Sun, Yiran Chen, and Jinghai Rao. ―Moodcast: Emotion prediction via dynamic continuous factor graph