International Conference on Pervasive Computing Advances and Applications – PerCAA 2019

Mental Stress Detection in University Students using Machine Learning Algorithms

Ravinder Ahuja[a], Alisha Banga[b]

[a]Jaypee Institute of Information Technology, Noida 201304, India
[b]Satyug Darshan Institute of Engineering and Technology, Faridabad-121002, India

Abstract

Mental stress is a major issue nowadays, especially among youngsters. The age that was considered once most carefree is now under a large amount of stress. Stress increase nowadays leads to many problems like depression, suicide, heart attack, and stroke. In this paper, we are calculating the mental stress of students one week before the exam and during the usage of the internet. Our objective is to analyze stress in the college students at different points in his life. The effect that exam pressure or recruitments stress has on the student which often goes unnoticed. We will perform an analysis on how these factors affect the mind of a student and will also correlate this stress with the time spent on the internet.

The dataset was taken from Jaypee Institute of Information Technology and it consisted of 206 student's data. Four classification algorithms Linear Regression, Naïve Bayes, Random Forest, and SVM is applied and sensitivity, specificity, and accuracy are used as a performance parameter. The accuracy and performance of data are further enhanced by applying 10-Fold Cross-Validation. The highest accuracy recorded was by Support Vector Machine (85.71%).

* Corresponding author.
E-mail address:ahujaravinder022@gmail.com

1. Introduction

Stress is a term frequently utilized synonymously with negative life experiences or life occasions. Logical research on pressure and uneasiness offers different points of view on the issue. The expanding pace of life hurried and focused ways of life imply that stress is an integral part of human life. A man in a condition of adjusting to pressure demonstrates conduct resistances. This prompts changes in one's psychological procedures and enthusiastic scene. Stress can be a kind of mental ordeal [1]. It additionally has a role in response to nature, and inspirations. A large damage may be there due to measures of pressure [5]. Due to stress there may be other health issues like obesity,

heart attack, diabetes, asthma etc.

Every hour, a student commits suicide in the different part of the country. Our country has reported large suicide cases of the youngsters aging between the age group of fifteen to twenty nine, as per Lancet report(2012). This is an approach with the help of which we can analyze the stress at its very first step. If we can find out the stress level in the students, in the short or long term we can help them in recovering. In the year 2015, the number of the suicide of students was 8,934. From 2010-2015, 39,775 students were dead due to suicide.

Our inspiration for this paper is the expansion in the number of suicide in our nation. Consistently around 92 individuals commit suicide around the world, which makes it 800,000 for every year. Out of this 135000(17%) are an inhabitant of India. The scale is vast. As indicated by the investigation the fundamental reason for suicide is the psychological issue [3]. Suicides are an impulsive reaction due to stress, for example, there may be money related challenges, several issues with one's connection or partner, or maybe due to bullying. Once a person thinks about it several different methods came into his mind, for example, guns, medications, and toxin. Treatment of mental clutters is one of the ways to deal with diminish the rate of suicide in the coming year.

We will play out an investigation on how these elements influence the psyche of these individuals utilizing their mind wave flag which is gathered from the PSS dataset. There have been efforts made in this field by many individuals but the main focus is on the PSS test and it inflicts on the subject under consideration thus to avoid any such situation or potential economic loss we consider making an effort to conduct such studies in a sequence of steps. The first being PSS to analyze the psychological state of an individual to some extent and to make a decision on whether to go on further or not.

2. Related Work

In paper [12], the authors calculated stress using heart rate, EMG, GSR hand and foot data, respiration and concluded that respiration is a critical parameter in stress. In the paper [13], the authors used ECG (Electrocardiogram) signals to predict stress. In paper [14], the authors calculated stress using signals like EEG, GSR, EMG, and SpO2. Various pattern recognition algorithms are being used for automated stress detection. The data received from all sensors are checked against the index value which is used for detecting the stress. In paper [5], authors applied the J48 algorithm, SMO, Bayesian Network algorithm for predicting stress on the data collected from 16 peoples under four different stressful conditions. In paper [16] used HRV features and EEG signal to predict the stress level. Various features like HRV, heart rate, ECG are used to predict the stress level. In paper [17], authors used decision tree algorithm is applied on a dataset collected from two test completed that these test to be unsatisfactory. Students stress level is calculated in the starting of the semester and in the last of the semester. The study found that stress in the start was less and higher in the last [4]. The researchers conducted before were mainly concerned about how to give a standard measurement for stress like by reading the brain signals or by conducting a survey about different ways of living of people in rural or urban areas but nothing provided for a preliminary detection. These research papers mention a lot about what all can be done to make these brain measurements more accurate by conducting some tests and recording brain signals or by measuring these brain signals in different situations.

3. Dataset and Pre processing

The dataset was taken from the 206 students of Jaypee Institute of Information Technology Noida. We have classified the data in two conditions one is before the exams and other is stress due to the usage of the internet. The dataset was collected for PSS test which includes 14 questions overall including the entire emotional question. The marking for the questions was in 5 ways (a) Never (b) Almost Never (c) Sometimes (d) fairly Often (e) Very Often. Then the weighted average model is used, preference is given to every question. The students are divided into 3 categories highly stressed, stressed and normal.

The dataset was collected from students in the college. They were asked basic questions about their feelings in situations that they might have encountered in the last month and their reactions to it [9]. Their answers are given some amount of weights and the weights thus help to calculate a score to analyze the stress level of the individuals.

The dataset was pre processed to analyze the weights and calculate final scores. To improve the performance of our model we applied K-fold cross-validation.

Perceived Stress Scale (PSS): This was developed by a psychology professor Sheldon Cohen. It was basically developed to analyze what kind of situations are how stressful for a person and his ability to cope with such situations. The levels are decided to analyze how uncontrollable and unpredictable people find their lives. It also made some queries about their recent experiences to analyze their mind situation and stress level when dealing with normal life scenarios [6].

The questions are basically related to certain common conditions that generally occur and ask the subject to respond in their concern. It basically focuses to know their feeling and the extent of their stress.

PSS can reduce the threat of extreme stress in an individual by diagnosing it at an initial level without investing much money at a preliminary stage.

4. Proposed Approach

In this paper, we used machine learning (ML) to identify the increasing stress level in the students and to predict the stress beforehand and be able to stop the major damage to their life before happening. In the test, we evaluate students amongst different situations. The level of stress was approved by the undertaking execution. The proposed model includes PSS dataset collection, pre processing, feature extraction and applying machine learning algorithm (Random Forest, SVM, NB, KNN) and comparing them on three performance parameters as shown in figure 1.

Figure 1: Overall Methodology used

5. Experimental Setup

We have used python language for implementation.

5.1 Performance Parameters

1. Sensitivity: This is also known as the true positive rate. It is the ratio of true positive (TP) to the true positive and false negative (FN). This specifies the ability of the model to correctly identify with the diseases as shown in equation 1.

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

2. Specificity: This is also known as the true negative rate. It is the ratio of true negative (TN) to the true negative and false positive (FP) as shown in equation 2. This specifies the ability to correctly identify without the diseases.

$$Specificity = \frac{TN}{TN+FP} \tag{2}$$

3. Accuracy: This is the ratio of true positives plus true negative to the true positives plus true negatives plus false positive plus false negative as shown in equation 3. It calculates how much percentage of cases is correctly classified.

$$Accuracy = \frac{TP+TN}{TN+FP+TP+FN} \tag{3}$$

5.2 K-Fold Cross Validation:

It is a technique to generalize the behaviour of data and increase the data k-fold times on the basis of that analysis. It basically helps to increase the performance of a model by increasing the dataset on which analysis is being performed. The process is to divide the dataset into k folds and then generalize the behaviour and increases the data entries and thus increase the efficiency of our model [11]. In our case, we have applied 10-fold cross validation because of the small dataset.

6. Classification Algorithms

Classification Algorithm is a unique technique in data mining through which one breaks down given information and takes each case of it. It classifies the example to a specific class with a very less chance of error. It is utilized to remove models that characterize imperative information classes inside the given informational index. We herein used some classification algorithms to detect stress level in individuals. We first trained our data and then tested our model on the rest of our data. The train to test ration used was 1:3.

6.1 Random Forest

This algorithm considers numerous decision trees, thus forming a forest. It is also called an ensemble of decision tree algorithms [7-8]. This can be used for classification as well regression. This algorithms tries to find out best feature randomly among all the features. In our experiment, we have used 100 decision trees and Ginni for impurity index

6.2 Naive Bayes

In the term of machine learning, naive Bayes classifiers consists a group of straightforward "probabilistic classifiers". They work upon the probability, highly scalable. Naive Bayes classifiers are quite adaptable. They require various parameters that are straight in the number of factors (highlights/indicators) in learning issue [10].

6.3 Support Vector Machine

This classifier that generally works upon the hyper plane [10]. This algorithm works upon the ideal hyper plane which is more useful in sorting new illustrations. In a 2-Dimensional plane, it is a line isolating a plane in two sections where each class lies in either of the sides.

6.4 K- Nearest Neighbour

This algorithm works upon whether the k-NN nearest, the classifier is utilized for arrangement or relapse. It is a class membership type in which either a student belongs to group a or b there is no between. If there are three groups then the data will be divided into three groups only. There may be some compromise by the neighbours, with the question being allocated to the class, which one is its k closest neighbours (k is a positive number and a small number). IF k = 1, then the protest will be allotted to the class of that solitary closest neighbour.

7. Results and Discussion

In this paper, we have applied four machine learning algorithms (Random Forest, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbour) and calculated specificity, sensitivity, and accuracy of all these. We found that support vector machine is performing well out of all four algorithms giving an accuracy of 85.71%, specificity 100%, and sensitivity of 75%, Random forest is performing next to support vector machine giving an accuracy of 83.33%, specificity of 66.66%, and sensitivity of 100%. Thus we can say that SVM is performing well out of these four algorithms in this scenario.

Table 1: Comparison of Different Algorithms

Sr. No.	Algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	Random Forest	100	66.66	83.33
2	Naïve Bayes	66.66	75	71.42

| 3 | Support Vector Machine | 75 | 100 | 85.71 |
| 4 | K-Nearest Neighbors | 70 | 44 | 55.55 |

8. Conclusion and Future Work

We can find the stress level by using the Perceived Stress Scale (PSS) test. As with its help, one can perform an initial analysis to help the person in his/her initial stages of stress if the person is in high mental stress state. In this paper, we have applied four classification algorithms (Random Forest, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbour) on the dataset of 206 students of JIIT Noida using sensitivity, specificity, and accuracy parameters. Due to small datasets, we have applied 10-fold cross validation also. We found that out of these four algorithms SVM is performing better as its geometric way of classification and amount of data is also less. Analyzing and finding methods like PSS with more accurate results and less cost can help improve the mental health of individuals and make our people mentally sound.

References

[1] Jung, Yuchae, and Yong Ik Yoon. "Multi-level assessment model for wellness service based on human mental stress level." *Multimedia Tools and Applications* 76.9 (2017): 11305-11317.

[2] Norizam, Sulaiman. *Determination and classification of human stress index using nonparametric analysis of EEG signals*. Diss. Universiti Teknologi MARA, 2015.

[3] http://www.mindgarden.com/documents/PerceivedStressScale.pdf

[4] Adnan, Nadia, et al. "University students stress level and brainwave balancing index: Comparison between early and end of study semester." *Research and Development (SCOReD), 2012 IEEE Student Conference on*. IEEE, 2012.

[5]https://www.hindustantimes.com/health-and-fitness/every-hour-one-student-commits-suicide-in-india/story-7UFFhSs6h1HNgrNO60FZ2O.html

[6] Subahni, Ahmad Rauf, Likun Xia, and Aamir Saeed Malik. "Association of mental stress with video games." *Intelligent and Advanced Systems (ICIAS), 2012 4th International Conference on*. Vol. 1. IEEE, 2012.

[7] https://www.deccanchronicle.com/nation/in-other-news/190916/stress-depression-lead-to-suicides.html

[8] Khosrowabadi, Reza, Chai Quek, Kai Keng Ang, Sau Wai Tung, and Michel Heijnen. "A Brain-Computer Interface for classifying EEG correlates of chronic mental stress." In *IJCNN*, pp. 757-762. 2011.

[9] Subhani, Ahmad Rauf, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, and Aamir Saeed Malik. "Machine learning framework for the detection of mental stress at multiple levels." *IEEE Access* 5 (2017): 13545-13556.

[10] Lederbogen, Florian, Peter Kirsch, Leila Haddad, Fabian Streit, Heike Tost, Philipp Schuch, Stefan Wüst et al. "City living and urban upbringing affect neural social stress processing in humans." *Nature* 474, no. 7352 (2011): 498.

[11] Khosrowabadi, Reza, Chai Quek, Kai Keng Ang, Sau Wai Tung, and Michel Heijnen. "A Brain-Computer Interface for classifying EEG correlates of chronic mental stress." In *IJCNN*, pp. 757-762. 2011.

[12] Ghaderi, A., Frounchi, J., Farnam, A.. Machine learning-based signal processing using physiological signals for stress detection. In: 2015 22nd Iranian Conference on Biomedical Engineering (ICBME). 2015, p. 93–98.

[13] Liu, David, and Mark Ulrich. "Listen to Your Heart: Stress Prediction Using Consumer Heart Rate Sensors." *Online]. Retrieved from the Internet* (2014)

[14] Xu, Q., Nwe, T.L., Guan, C.. Cluster-based analysis for personalized stress evaluation using physiological signals. IEEE Journal of biomedical and health informatics 2015; 19(1):275–281.

[15] Villarejo, M.V., Zapirain, B.G., Zorrilla, A.M. A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. Sensors 2012; 12(5):6075–6101.

[16] Gjoreski, M., Gjoreski, H., Lustrek, M., Gams, M.. Continuous stress detection using a wrist device: in the laboratory and real life. In: ˇProceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM; 2016, p. 1185–1193.

[17] Selvaraj, N. Psychological acute stress measurement using a wireless adhesive biosensor. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015, p. 3137–3140.