Mental Stress Detection in Students using Machine Learning Algorithms

Binny S [1], Darsana K Lal[2], Cina Mathew[3], Jefna Nazar[4]

(Associate Professor, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India)
([2]MCA, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India)
([3]Assistant Professor, KristuJyoti College of Management and Technology, Kottayam, Kerala, India)
([4]MCA, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India)

Abstract

Stress has become part of students' academic life due to the various internal and external expectations placed upon their shoulders. The age that was considered once most carefree is now under an outsized amount of stress. Stress increase nowadays results in many problems like depression, suicide, attack, and stroke. In this paper, we are trying to calculating the mental stress of scholars one week before the semester exams and after the semester exams. Our objective is to research stress within the college students at different points in his life. The effect that exam pressure or recruitments stress has on the scholar which regularly goes unnoticed. Here we perform an analysis on how these factors affect the mind of a student and may also correlate this stress with the time spent on the online .The dataset was taken from college students and it consisted of 150 students 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 knowledge are further enhanced by applying 10-Fold Cross-Validation. The absolute best accuracy recorded was by Support Vector Machine (85.71%).

Keywords: Classification algorithms: Linear Regression, Naïve Bayes, Random Forest, SVM; Mental Stress; Perceived Stress Scale; Sensitivity; and Specificity, Accuracy.

1. Introduction

University of Melbourne, on an average every four children out of five are suffering from the mental disorders. Psychological health plays a very important role in molding the development characteristics and academic performance of the child. The expanding pace of life hurried and focused ways of life imply that stress is an integral a part of human life. A man during a condition of adjusting to pressure demonstrates conduct resistances. This makes to changes in one's psychological procedures and enthusiastic scene. Stresses are often a sort of mental ordeal. It additionally features a role in response to nature, and inspirations. Due to stress there may be other health issues like obesity, heart attack, diabetes, asthma etc. Every hour, a student commits suicide within the different a part of the country. To determine and manage the mental stress among the individuals, the traditional approach is to concern the psychiatrists. Individuals can go to psychiatrists who will help them to identify the causes of stress and manage the stress levels accordingly. But due to the expensive fees that cannot be affordable by everyone. In this paper, we proposed the technical approach for determining the mental stress level of the students by using the machine learning model approach. This technical approach of determining stress eliminates the need of individuals to visit the psychiatrists physically. This is an approach with the assistance of which we will analyze the strain at its very initiative. If we can find out the stress level in the students, in the short or long term we can help them in recovering. Our inspiration for this paper is that the expansion within the number of suicide in our nation. To incorporate this, we have collected the real time survey data of students belonging to different streams from various colleges based on questionnaire. There are efforts made during this field by many individuals but the most focus is on the PSS test. The first being PSS to research the mental state of a private to some extent and to form a choice on whether to go on further or not.

2. Related Work

In paper [7], the authors calculated stress using pulse, EMG, GSR hand and foot data, respiration and concluded that respiration may be a critical parameter in stress. In the paper [8], the authors used ECG (Electrocardiogram) signals to predict stress. In the paper [9], where the authors calculated stress using signals such as EEG, GSR, EMG, and SpO2. Various pattern recognition algorithms are getting used for automated stress detection. The data received from all sensors are checked against the index value which is employed for detecting the strain. In paper [10] used HRV features and EEG signal to predict the strain level. Various features like HRV, pulse, ECG are wont to predict the strain level.The researchers conducted before were mainly concerned about the way to provides a standard measurement for stress like by reading the brain signals or by conducting a survey about different ways of living of individuals in rural or urban areas but nothing provided for a preliminary detection. These research papers mention tons about what all are often done to form these brain measurements more accurate by conducting some tests and recording brain signals or by measuring these brain signals in different situations
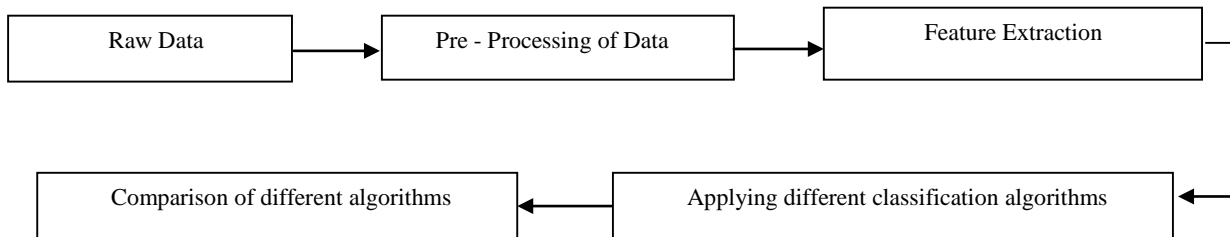
3. Dataset and Pre processing

The dataset was taken from the 150 students of Kristu Jyoti College of Management And Technology Changanacherry, Kottayam. Here we classify the information in two conditions one is before the exams and other is stress during the usage of the web. The set of information was collected for PSS test which includes 10 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 and the preference is given to every question. The students are divided into 3 categories based on highly stressed, stressed and normal. The set of data was collected from students in the college. They were asked basic questions on their feelings in situations that they could have encountered within the last month and their reactions thereto [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. Preprocess the collected dataset to analyze the weights and calculate final scores. We applied the K-fold cross-validation to improve the performance of our model.

Perceived Stress Scale (PSS): This was developed by a psychology professor Sheldon Cohen. It was basically developed to research what quite situations are how stressful for an individual and his ability to deal with such situations. The amounts are decided to research how uncontrollable and unpredictable people find their lives. It also made some queries about their recent experiences to research their mind situation and stress level when handling normal life scenarios [6].The questions are basically associated with certain common conditions that generally occur and ask the topic to reply in their concern. It basically focuses to understand their feeling and therefore the extent of their stress. PSS can reduce the threat of utmost stress in a private by diagnosing it at an initial level without investing much money at a preliminary stage.

4. Proposed Approach

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

Raw Data	→	Pre - Processing of Data	→	Feature Extraction

| Comparison of different algorithms | ← | Applying different classification algorithms | ← | |

5. Experimental Setup

Python language is used for implementation.

5.1 Performance Parameters

1.　　*Sensitivity:* This is often also referred to as truth positive rate. It is the ratio of true positive (TP) to truth positive and false negative (FN). This specifies the power of the model to properly identify with the diseases as shown in equation 1.

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

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

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

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 what proportion percentage of cases is correctly classified.

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

5.2 K-Fold Cross Validation

It is a way to generalize the behavior of knowledge and increase the information k-fold times on the idea of that analysis. It is basically helps to extend the performance of a model by increasing the set of data on which analysis is being performed. The method is to divide the information set into k folds then generalizes the behavior and then the data entries are increased thus increase the efficiency of our model [11]. In the case here, 10-fold cross validation is applied because of the small dataset.

6. Classification Algorithms

Classification Algorithm could also be a singular technique in processing through which one breaks down given information and takes each case of it. It classifies the instance to a specific class with a very less chance of error. It's utilized to urge obviate models that characterize imperative information classes inside the given informational index. Here we used some classification algorithms to detect the stress level in individuals. First of all trained our data then tested our model on the rest of our data. The train to check 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 the decision tree algorithms . This can be used for the classification also regression. These algorithms try to seek out best feature randomly among all the features. In our experiment, We've used 100 decision trees and Ginny for impurity index.

6.2 *Naive Bayes*

In the term of machine learning, naive Bayes classifiers consist of 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.

6.3 *Support Vector Machine*

This classifier generally works upon the hyper plane. This algorithm works upon the perfect hyper plane which is more useful in sorting new illustrations. During a 2-Dimensional plane, it's a line isolating a plane in two sections where each class lies in either of the edges.

6.4 *K- Nearest Neighbor*

This algorithm works upon whether the k-NN nearest, the classifier is employed for arrangement or relapse. It's a category membership type during which either a student belongs to A or b there's no between. If there are three groups then the info are going to be divided into three groups only. There could also be some compromise by the neighbors, with the question being allocated to the category , which one is its k closest neighbors (k may be a positive number and a little number).IF k = 1, then the protest are going to be allotted to the category of that solitary closest neighbor.

7. Results and Discussion

In this paper, we applied four machine learning algorithms such as (Random Forest, Naïve Bayes, K-Nearest Neighbor and Support Vector Machine) 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 85.81%, specificity 100%, and sensitivity of 77%, Random forest is performing next to support vector machine which giving an accuracy of 84%, specificity 68%, and sensitivity of 100%. Thus I 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%	68%	84%
2.	Naïve Bayes	65%	77%	70%
3.	Support Vector Machine	77%	100%	85.81%
4.	K-Nearest Neighbors	69%	40%	55%
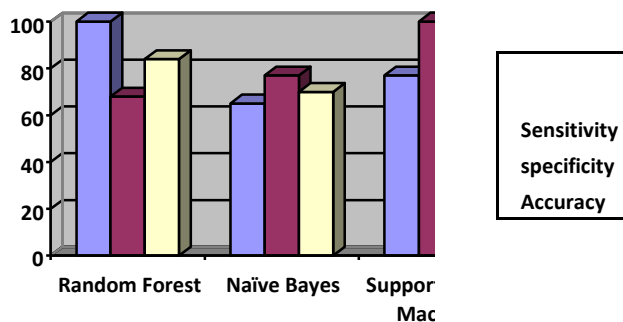
8. Conclusion and Future Works

We can find the strain level by using the Perceived Stress Scale (PSS) test. As with its help, one can perform an initial analysis to assist the person in his/her initial stages of stress if the person is in high mental stress state. In this paper, we've applied four classification algorithms as (Random Forest, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor) on the dataset of 150 students from different colleges, using the parameters sensitivity, specificity, and accuracy. Due to small set of data, we've applied the 10-fold cross validation also. We found out that of these four Algorithms, SVM is performing better as its geometric way of classification and amount of knowledge is additionally less. Analyzing and finding methods like PSS with more accurate results and fewer costs can help improve the psychological state 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):1130511317

[2] Norizam, Suleiman. Determination and classification of human stress index using nonparametric analysis of EEG signals. Diss. University Technology MARA, 2015 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.

[3] Binny S , A survey concept on Deep Learning, International Journal of Scientific & Engineering Research Volume10,Issue 6, June2019 1570

[4] International Journal of Engineering and Advanced Technology(IJEAT) ISSN: 22498558,Volume-9 Issue-1,October 2 019 335 Retrieval Number : F873088619/2019©BEIESP DOI : 10.33590/ijeat.F8573.109119 Published by : Blue Eyes Intelligence Engineering & Sciences Publication Detection and Analysis of Stress using Machine Learning Techniques by Reshma Radheshamjee Baheti, Supriya Kinariwala

[5] Digital Object Identifier 10.1109/ACCESS.2017.272362 Machine learning Framework for the Detection of Mental stress at multiple levels AHMAD RAUF SUBHANI, WAJD MUMTAZ, MOHAMED NAUFAL BIN MOHAMED SAAD, NIDAL KAMEL, AND AAMIR SAEED MALIK Centre for Intelligence Signal and Imaging Research, Department of Electrical and Electronic Engineering University Technology PETRONAS, Bandar Seri Iskandar 32610, Malaysia Corresponding author: Aamir Saeed Malik

[6] 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.

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

[8] 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.

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

[10]Susheel George Joseph, "The Usage of Machine Learning Evolutionary Algorithms in Medical Images Formed by Computer Tomography (CT) or X-Rays to Detect the Infections due to COVID 19", PENSEE (penseeresearch.com) ISSN: 0031-4773. Volume 51, Issue 4, Page No:1512-1518, April 2021. Available at: https://app.box.com/s/n6nsv8myosb0wb16psvtb8conekpy ohj

[12]. Susheel George Joseph, "A Machine Learning (ML) Modelling Approach in Monitoring and Controlling the Viral Pandemic- COVID 19", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.1709-1717, June 2020: http://www.jetir.org/papers/ JETIR2006575.pdf

[13]. S.Binny, "A survey concept on Deep Learning", International Journal of Scientific & Engineering Research (www.ijser.org), ISSN 2229-5518 , Volume 10, Issue 6, page 1570-1575, June-2019.

[14]. Susheel George Joseph, "A Machine Learning (ML) Modelling Approach in Monitoring and Controlling the Viral Pandemic- COVID 19", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.1709-1717, June 2020: http://www.jetir.org/papers/ JETIR2006575.pdf

[15]. Susheel George Joseph, Dr. Vijay Pal Singh, "Denoising of Images using Deep Convolutional Neural Networks (DCNN)", International Journal of Engineering Development and Research (IJEDR), ISSN:2321-9939, Volume.7, Issue 3, pp.826-832, September 2019, http://www.ijedr.org/papers/IJEDR1903143.pdf.

[16]. S. Binny, Dr. P. Sardar Maran, "Classification of Lung disease using deep learning neural networks, " PENSEE (penseeresearch.com) ISSN: 0031-4773. Volume 51, Issue 6, Page No: 466-470, June 2021.