ANALYSIS THE STRESS FACTORS USING MACHINE LEARNING

SUMATHI T[1*] KARTHIK S[2] BARANIDHARAN T[3]
[1*]Assistant Professor, Department of Information Technology, Government College of Engineering, Erode-638316, TamilNadu, India
E-Mail: sumathiirtt@gmail.com

[2]Professor and Dean, Department of Computer Science & Engineering, SNS College of Technology, Coimbatore-641035, TamilNadu, India
E-Mail: profskarthik@gmail.com

[3]Associate Professor, Department of Electronics and Communication Engineering, K S Rangasamy College of Technology, Tiruchengode-637215, TamilNadu, India
E-Mail: baraniinst@gmail.com

Abstract

Young people frequently struggle with stress disorder as a result of changing cultures, lifestyles, and competition. Stress is one of the mental illnesses that can have varying degrees of negative effects on both physical and mental health, according to medical science. The number of those affected by this ailment is increasing daily, and the proportion of teens affected is skyrocketing. As they transition from adolescence to adulthood, teenagers face a variety of challenges, many of which are made worse by their exposure to social media technology. Therefore, it's crucial to comprehend the myriad factors that contribute to stress and to pinpoint the traits that are the main causes so that appropriate measures may be taken to effectively manage it. The evaluation of student stress at several different educational institutions begins with this study. In order to identify the most significant and significant elements, we will utilise machine learning techniques to evaluate stress patterns in this work. The information we are utilising for this was gathered through a survey given to graduate and postgraduate students at several campuses and institutions. After cleaning and preparing our data, a variety of machine learning algorithms have been employed to train and evaluate our data. Academic, economical, and social aspects were shown to be major characteristics that impact stress utilising Decision Tree and Support Vector Machine. Additionally, Neural network is recognised as the Best approach for the presented data. With the aid of these findings, one may recommend the ideal model for worry data in addition to perhaps aiding in the recommendation of stress-reduction techniques like yoga to the pupils.

Keywords: Stress, Machine Learning, Random Forest, support vector machine

DOI Number:10.14704/nq.2022.20.8.NQ44426 **NeuroQuantology 2022; 20(8): 3953-3959**

1.Introduction

Stress is defined as a state of emotional or physical illness. It can be triggered by any incident or idea that causes you to feel dissatisfied, furious, or anxious. Your body's reaction to a challenge or demand is called stress. Stress may be beneficial in brief spurts, such as when it helps you escape danger or make a deadline. However, prolonged stress can be harmful to your health [1].

Stress is one of the key issues facing many nations in a globalised society. The performance parole can survive even a little level of stress. In some circumstances, prolonged, excessive pressure can be persistent and lead to a variety of health risks. Numerous different studies have been conducted in the past, raising alarm about the same problem. A healthy level of stress is regarded as beneficial since it generates the minimal amount of pressure necessary to inspire someone to complete a task and survive in this cutthroat society.

However, beyond a certain point and under sustained strain, mental disease can develop with mild to severe symptoms. Continuous stress or pressure can lead to a variety of health problems, including hypertension, insomnia, jitters, and aching muscles, among others. Depression or thoughts of suicide might even be the most serious effects. If we examine medical and research data, startling information is revealed that demonstrates that college students experience the same levels of stress as experts. Although there may be differences in the two factors, carrying stress is still a worry. Numerous studies document differences in stress levels by gender and age, but these are just two characteristics among many other crucial ones.

The 19–25 age group is now more likely to experience stress, which may be connected to studying, social instability, peer pressure, and many other factors. Similarly, if we consider gender differences, research may show that women are more likely to experience stress owing to mental and physical strain at home and at work. Stress is divided into two categories based on its intensity: Acute Stress: This type of short-term stress is extremely prevalent. Typically, a person reaches this level when they are under the pressure of recent or impending obstacles. Mild health effects from this stress include headaches, stomach trouble, and others. Chronic Stress: Chronic stress can develop from long-term stress, such a difficult financial condition. Additionally, chronic stress develops from acute stress that is repeated or persists over time. This kind of stress has a lot more negative effects, such as heart disease, suicidal thoughts, etc.

In Common Machine Learning Methodologies and logistic regression, using a dichotomous (binary) dependent variable, logistic regression is a statistical technique for creating machine learning models. Logistic regression is used to describe data and the connection between a dependent variable and one or more independent variables. Independent variables can be made up of nominal, ordinal, or interval variables. With respect to Random Forest, in order to handle classification and regression issues, people frequently employ the supervised learning approach known as random forest.

Using the average for regression and the majority vote for classification, it builds decision trees from a variety of samples. The Random Forest Algorithm's ability to handle data sets with both continuous and categorical variables, as in regression and classification, is one of its key features. It offers better outcomes for classification challenges. SVM Support vector machines, sometimes known as SVMs, are supervised learning techniques that may be applied to both classification and regression problems. The majority of the time, it is used in Machine Learning to solve classification problems. The goal of the SVM method is to determine the best decision boundary or line for classifying n-dimensional space into groups so that future additions of data points may be quickly assigned to the appropriate group.

The border of the ideal choice is called a hyperplane. SVM is used to choose the extreme

points and vectors that help build the hyperplane. The most extreme examples are termed support vectors, and the technique is known as a support vector machine. Take a look at the illustration below, which displays two unique groups that are divided by a decision boundary. KNN Classifier K-Nearest Neighbors (KNN) is a method for tackling classification and regression issues. KNN algorithms categorise new data points using data and similarity measures (e.g. distance function).

The classification of it is based on a majority vote of its neighbours. The class with the greatest number of neighbours is given the information. The value of k increases as the number of nearest neighbours does, and accuracy may rise as a result. Artificial intelligence is a subset of machine learning, while deep learning is a subset of machine learning. Deep learning algorithms analyse data according to a preset logical framework to arrive at judgments that are comparable to those of humans. Deep learning does this by using neural networks, which are multi-layered structures of algorithms.

2. Related Work

A test of human wellbeing and the nature of existence is being put to the test by the sharp increase in mental health problems or stress. A strange test of human well-being and the nature of life has been created by the fast escalation of mental crises or stress. It should be noted that the words pressure and unease are frequently used in opposition. Stress and anxiety frequently cause similar physical symptoms, such as an elevated heart rate, sweaty hands, and a racing heart, which are mostly covered by neural circuits when the brain fails to distinguish between a perceived and real threat. These comparisons touch on how each way of doing things seems on the surface. On stress, several research have previously been conducted.

In the works [2,3], the author used deep learning techniques to anticipate the patient's level of anxiety or sadness. A patient health questionnaire is given to participants, and the results are used to assess depression. A multimodel deep learning model is created to identify depression. The confusion matrix is employed to assess the outcomes. Precision: 68.615, NPV: 67.955, Sensitivity: 68.595, Specificity: 67.464, and F Score: 67.653 are the results of the paper. [2] found that Nave Bayes, convoluted neural networks may be employed and used Logistic regression, KNN classier, Decision tree, Random Forest, boosting for stress prediction with an accuracy of 75.13%. [3] The author described using logistic regression and SVM for stress prediction. Employing accuracy values of 66% and 68%, respectively, it can be shown that using SVM improved accuracy.

Occupational stress research has produced a few basic themes on the causes of stress during the last two decades, but it has not produced a comprehensive conceptualization of the problem. A clinical perspective of stress as a psycho-physiological condition that originates from an individual's perception of the balance between environmental demands and response capacities has dominated the discipline. Stress causes physiological, psychological, and behavioural changes, all of which are mediated through perceptions. According to this paradigm, the majority of stress research has concentrated on individual perceptions and vulnerability, and the majority of therapies have centred on individual coping techniques [9].

Class prediction, SVM, and Naive-Bayes classifiers have all been utilised in [4][7][8]. The main classification measures, such as the F1-score, accuracy, and confusion matrix, have been used to display the findings. reported seventy nine percent accuracy for SVM and eighty three percent accuracy for Naive-Bayes. Using emotive analysis on tweets [5] [6], the author employed the Multinomial Nave Bayes method and SVM to identify sadness. They discovered The Multinomial Nave Bayes method outperformed SVM in terms of accuracy and F1 score. Preprocessing of Data Set was done with 315 students from various graduate and postgraduate institutions made

up the dataset. Data has been concentrated on elements that might be the main cause of stress. The majority of data comprises yes answers. In order to prevent future suicides, our question set also focuses on suicidal thoughts. Label encoding and one hot encoding are employed for pre-processing, and feature significance is determined by using rrandom forest.

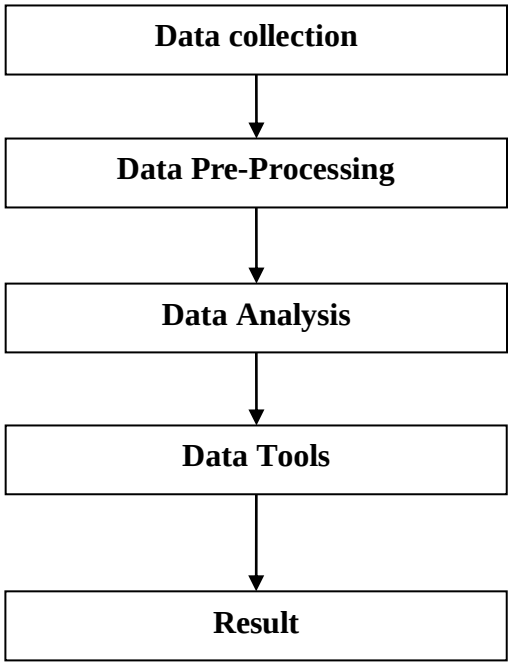

Fig 1: Proposed Approach

The first step in doing any analysis is gathering data. Both primary and secondary data sources may be used for analysis. In this study, primary data were employed. The researcher developed a mechanism to gather data from graduate and postgraduate students. After being tested by psychologists, the tool was used to gather data. Data preparation and cleaning: Data collection involves cleaning and pre-processing because it is primary data. Average values were used to fill in the blanks, and unnecessary cells were removed. The same Label Encoder and OneHot encoder technologies were used to further prepare the data for reading by machine learning algorithms.

After the data collection and applying the pre-processing, the data analysis can be done. Data

3. Proposed Method

In this poposed work, we used machine learning (ML) to uncover key approaches for detecting stressors and to identify stress among college students. A machine learning algorithm is used in the recommended method to gather datasets, pre-process them, extract features, classify them, and make predictions.

on a number of variables, including age, gender, and variables thought to be stressful, were analysed. A review of the literature was also done to take into account the main aspects of our study. Equipment and Methods: In order to use traditional group-based approaches for individual-level prediction, statistical models are usually validated in external samples, and this has contributed to the variety of tools offered by ML. In the years that followed, the machine learning paradigm went in and out of fashion, but it is now a part of our daily lives because advances in computing power and algorithm development produced performance gains that were greater than those of using conventional statistical methods or rule-based, deterministic programming across a wide range of fields.

SVM (Support Vector Machine) and Random Forest are the two most used methods for building prediction models. The same job has been accomplished using other strategies.

After doing all of the research, the results were created in the form of a graph, a confusion matrix, a comparison to show the best approach or tool, as well as the most important causes of stress among college students.

4. Performance Analysis and Results

Various machine learning models, including random forest, gradient boosting, SVM, neural networks, and others, were used in the experiment. The dataset was pre-processed by adding missing values by mean values and removing unnecessary replies before training the model. Train data and test data are separated from the entire set of data. Important elements that cause stress in students when they are in college have been discovered. This also aids in reducing the target area. The following phase involved testing with all of the machine learning models using the shared parameters.

Feature Importance: Important Feature get used to determine which Column in the given dataset is the most noticeable. It is useful to know which variable in our dataset is more significant. Although there were numerous assumed elements that might have been studied, a literature review was conducted in order to focus on only six, which were then confirmed by psychologists.

The goal of this study was to predict stress in college students and to determine which machine learning model performed the best. This investigation was carried out in Python utilising a Google-powered Colab simulation. The model's data was cleaned and pre-processed using a variety of methods prior to training. Data was split into a training and testing set with a ratio of 75% and 25%, respectively. Several models were used to the presented data set to predict student stress once key characteristics were identified.

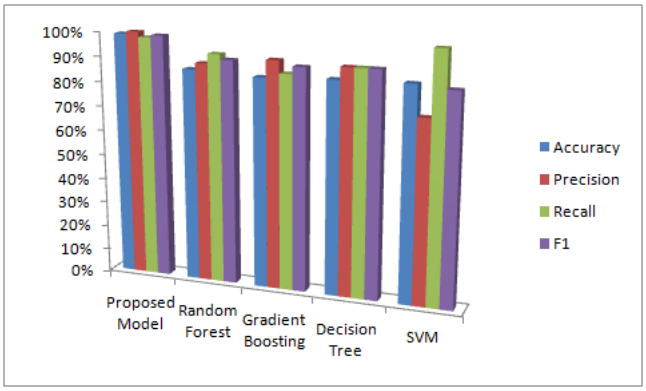

Figure 2. Comparision of Models

All of the machine learning models were tested using the same set of parameters. For additional comparison investigations, accuracy, precision, F1 score, recall value, and confusion matrix were employed. The first experiment used a decision tree and had an accuracy rate of about 86 percent and a precision rate of about.91. In the series of experiments, the following experiment was carried out using SVM, with an accuracy of 86.4 percent and a precision of 0.74; the next experiment was carried out using Random Forest, with an accuracy of 86.4 percent and a precision of 0.84. Gradient boosting was then employed, with an accuracy of 85% and a precision of 0.92. Finally, a neural network was used, which produced the greatest results in terms of accuracy (99%) and precision (1). Table 1 has further information.

Highest precision and accuracy were offered by the neural-based model. The suggested model comes close to being accurate for predicting student stress. A bar graph is generated to create thorough comparisons for better

visualisation and comparison. Performance comparison is represented by the bar graph in the picture.

VII Conclusion and Future Scope

The above-mentioned instruments will be used more often, ensuring more precise results and accurate detection of student stress so that quick attention may be called. This study was conducted to evaluate student attitudes and forecast student stress. Primary data were first collected for the study using a tool made by the researcher. Additional significant characteristics that indicate whether a student is experiencing stress or not and how it is impacting them were also collected. Researchers also attempted to pinpoint key sources of stress. In the study, it was discovered that financial stress, relationship stress, and academic pressure are the three main stressors that students experience and that have an impact on their physical health.

This study's secondary goal was to examine several machine learning models across a range of parameters. In this study, a neural network with six layers and four neurons produced the best results out of the five models tested. 99 percent of predictions are correct. Although alternative models are not claimed to be insufficiently accurate in this study, we may assert that neural networks do the best in studies of a comparable nature. To sum up this study, we can state that Neural Networks provide the best case precision for the presented data set. Here, we may draw the conclusion that the train model and test model are providing accuracy of over 99 percent and are capable of accurately predicting stressed pupils. More elements might be added in the future to forecast stress and scale it on low, high, and moderate to require rapid treatment. An app that has been validated and approved by a psychologist can be used to gather primary data in the future without the need for additional parameter verification.

References

[1]. https://medlineplus.gov/ency/article/003211.htm.

[2]. Shakir Khan, V. Saravanan, Gnanaprakasam C. N, T. Jaya Lakshmi, Nabamita Deb, Nashwan Adnan Othman, "Privacy Protection of Healthcare Data over Social Networks Using Machine Learning Algorithms", Computational Intelligence and Neuroscience, vol. 2022, 8 pages, 2022. https://doi.org/10.1155/2022/9985933

[3]. K. D. Mistry and B. J. Talati, "Integrated approach for bone tumor detection from MRI scan imagery," 2016 International Conference on Signal and Information Processing (IConSIP), 2016, pp. 1-5, doi: 10.1109/ICONSIP.2016.7857471.

[4]. Guimond, A., Subsol, G., & Thirion, J. (1997). Automatic MRI Database Exploration and Applications. *Int. J. Pattern Recognit. Artif. Intell., 11*, 1345-1365.

[5]. S. Handore and D. Kokare, "Performance analysis of various methods of tumour detection," 2015 International Conference on Pervasive Computing (ICPC), 2015, pp. 1-4, doi: 10.1109/PERVASIVE.2015.7087002.

[6]. Nilesh Bhaskarrao Bahadure, Arun Kumar Ray, Har Pal Thethi, "Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM", International Journal of Biomedical Imaging, vol. 2017, Article ID 9749108, 12 pages, 2017. https://doi.org/10.1155/2017/9749108

[7]. Vinit Kumar Gunjan, Y. Vijayalata, Susmitha Valli, Sumit Kumar, M. O. Mohamed, V. Saravanan, "Machine Learning and Cloud-Based Knowledge Graphs to Recognize Suicidal Mental Tendencies", Computational Intelligence and Neuroscience, vol. 2022. DOI: https://doi.org/10.1155/2022/3604113

[8]. M. Nasor and W. Obaid, "MRI Tumor Detection and Localization by Multiple Threshold Object Counting Technique," *2018 International Conference on Computer and Applications (ICCA)*, 2018,

pp. 158-161, doi: 10.1109/COMAPP.2018.8460322.

[9]. Y. Rongge, L. Huaiwen, B. Tong and Z. Luna, "Research on stress of magnetically controlled saturable reactor under practical working conditions," *2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, 2017, pp. 1-5, doi: 10.1109/ICEMS.2017.8056338.