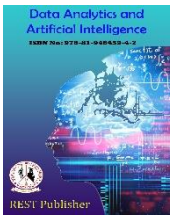
Data Analytics and Artificial Intelligence

Vol: 3(3), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/



Predictive Analysis of Student Stress Level Using Machine Learning

Kalai Vani. V, *Savitha. K, Barkavi. R, Eshaswetha. E

Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India.

*Corresponding Author Email: **asvitha.k2001@gmail.com**

Abstract: *Students are suffering from many mental health problems including mental stress, somatization, obsessive, interpersonal sensitivity, depression and fear which can bring a lot of negative effects to them. We need a system to handle the student mental health problems, here we are mainly concentrating on student stress prediction. There are so many factors related to stress, such as work load, assignments, family issues, attendance, teaching, etc. Anxiety, frustration, anger, or uneasiness can all cause stress. We have presented a methodology that will predict the level of stress a student is experiencing. Computer science has advanced significantly as the digital world has progressed, particularly in the field of health care with the application of machine learning. Machine learning is a subject to predict future based on the past data. We developed a real-time database in which we surveyed over 2000 students with a variety of 48 questions to analyze their stress and cure them early using machine learning algorithms. The survey data is taken as the input to predict stress of each student. Various parameters were considered to evaluate the accuracy of our trained model. Different types of machine learning algorithms are used, including Decision Tree, Random Forest, Logistic Regression, and K Nearest Neighbors, with a maximum accuracy of 96 percent using Random Forest.*

Keywords: *Machine Learning, Real-time data, Random Forest, Stress prediction*

1. INTRODUCTION

Stress is one of the most essential issues that needs to be discussed. It is a feeling of anxiety, loneliness, and emotion. Students and teachers worldwide have suffered mental stress since the outbreak of COVID-19, as they have had to strike a balance between home and online learning, and the post-COVID-19 wave also showed an increase in stress and a feeling of anxiety among students due to the intense amount of pressure after the re-opening of institutions worldwide. According to reports, 74% of students are suffering from stress right now. Stress in students can be caused by a lot of factors, including financial problems, health issues, problems with the teacher, and an unpleasant environment at home, to name a few. According to reports, students are pressured by their parents to score well in examinations, which leads to a great deal of stress. It can create dysfunctionality in a student's life, and stress should be predicted at an early stage so that it can be treated well, and the student can get help. We have applied different machine learning algorithms to predict stress and also compared them using various parameters such as accuracy score, ROC curve, precision, recall, and specificity to check which algorithm gives the best result. The algorithms were trained and tested on a real-time dataset of school students.

2. RELATED WORK

Garima Verma, Hemraj Verma, et al. [1] proposed a model to examine the mental stress among the students studying in higher education at the college level, especially those who are taking technical education in India. Also, the paper will identify the factors that affect the mental condition of students who have just taken admission from traditional school education to technical education in college or university. An ensemble prediction model to know the stress level of technical students has been proposed using the machine learning technique. Aanchal Bisht, Shreya Vashisth, et al. [2] proposed a methodology that will assist schools, parents, and students in predicting the level of stress a student is experiencing. They developed a real-time database in which they surveyed over 190 school children aged 14 to 18 years old with a variety of 26 questions to analyze their stress and cure them early using Machine Learning Algorithms. Different types of Machine Learning algorithms are used, including Decision Tree, Logistic Regression, K Nearest Neighbors, and Random Forest, with a maximum accuracy of 88 per cent using K Nearest Neighbors. K. Parthiban, Digvijay Pandey, et al. [3] a technique has been proposed to provide learners with a superior online classroom teaching experience, allowing the online classroom to be as good as, if not better than, a single online classroom. This study focused on daily teaching methods that employ online learning supported by a machine teaching approach to provide an individual with a relevant stress-free solution. Ravinder Ahuja, Alisha Banga, [4] proposed a system for calculating the

mental stress of students one week before the exam and during the usage of the internet. Main objective is to analyze stress in the college students at different points in his life. They performed 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 performance parameter. Aditya Vivek Thota, et al [5] proposed a model to analyze stress patterns in working adults and to narrow down the factors that strongly determine the stress levels using machine learning techniques. Towards this, data from the OSMI mental health survey 2017 responses of working professionals within the tech-industry was considered. Various Machine Learning techniques were applied to train the model after due data cleaning and preprocessing. The accuracy of the above models was obtained and studied comparatively.

3. MACHINE LEARNING ALGORITHMS

Random forest algorithm: Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

K-NN algorithm: K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It is used to identify the nearest neighbors of a given query point, so that we can assign a class label to that point. A class label is assigned on the basis of a majority vote.

Logistic regression: Logistic regression is basically a supervised classification algorithm. The target variable or output (y), can take only discrete values for given set of features or inputs (X). The logistic function or the sigmoid function is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$$1 / (1 + e^{-value})$$

- 'e' is the base of natural logarithms
- 'value' is the actual numerical value that you want to transform

Decision tree: Decision Tree is a Supervised learning technique that can be used for both classification and Regression internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

4. PROPOSED SYSTEM

We collected the real-time data from students. The dataset has structured data in CSV format. 48 attributes are present in the dataset which were the questions asked to students about what they felt, which defined their stress. The timestamp contains the exact time at which a student entered his/her details. Full Name, Gender, and Age contain their details. Next, training and testing datasets were generated using the refined real-time dataset with the percentage of 75% and 25% respectively. Finally, we applied the four machine learning algorithms to our collected dataset. Algorithms used here takes less time for data processing when compared to other existing systems. Produces accurate results by comparing various machine learning algorithms. All the responses from the students are monitored continuously to predict stress.

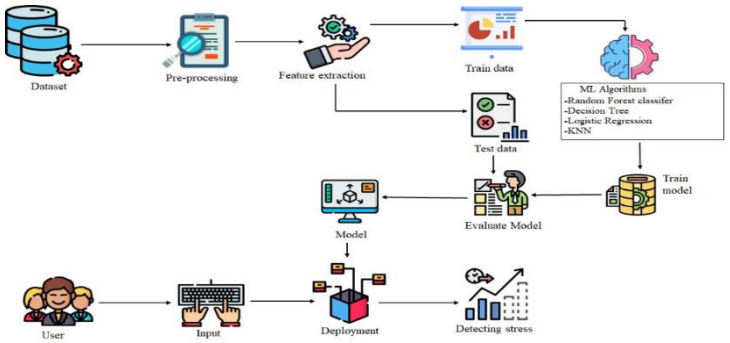


FIGURE 1.

5. MODEL DESIGN

Data collection: Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained. In our system we collected data from the students of different ages and stored in CSV format.

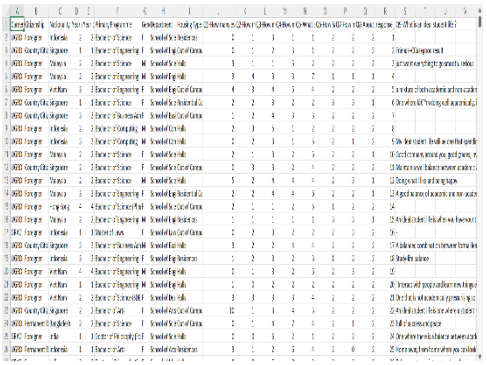


FIGURE 2. Questions

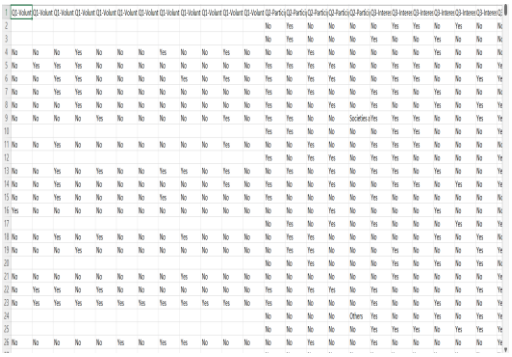


FIGURE 3. Responses

Data pre-processing: Real-world raw data are often incomplete, inconsistent and lacking in certain behaviors or trends. They are also likely to contain many errors. So, once collected, they are pre-processed into a format the machine learning algorithm can use for the model. Pre-processing includes a number of techniques and actions:

- **Data cleaning**: The process of adding missing data and correcting, repairing, or removing incorrect or irrelevant data from a data set.
- **Data integration**: Combining multiple datasets to get a large corpus can overcome incompleteness in a single dataset.
- **Data transformation:** Data transformation will begin the process of turning the data into the proper formats for analysis.

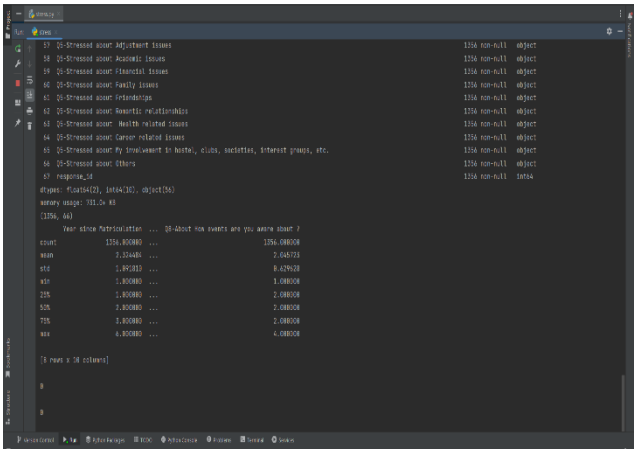


FIGURE 4. Data Pre-processing

Accuracy comparison: Accuracy is a proportional measure of the number of correct predictions over all predictions. Correct predictions are composed of true positives (TP) and true negatives (TN). All predictions are composed of the entirety of positive (P) and negative (N) examples. P is composed of TP and false positives (FP), and N is composed of TN and false negatives (FN). It is defined as the fraction of predictions that the model correctly predicted

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

```
DECISION TREE
the accuracy on testing data 0.916
the accuracy on training data 0.992

RANDOM FOREST
the accuracy on testing data 0.964
the accuracy on training data 1.0

KNN
the accuracy on testing data 0.776
the accuracy on training data 0.834667

LOGISTIC REGRESSION
the accuracy on testing data 0.896
the accuracy on training data 0.9655
```

FIGURE 5. Accuracy

6. CONCLUSION

The majority of the world's population is suffering from stress, and it is increasing at an alarming rate. It affects the individual's physical and mental health. The prediction or detection of stress among people has become an important aspect because every day billions of people are suffering from stress-related health problems like depression, insomnia, eating disorders, circulatory problems, heart diseases, anxiety and panic attacks, and the list goes on and these people are left unnoticed and untreated. From our study, we were able to compare the working of various algorithms on our real-time dataset with Random Forest giving the highest accuracy of 96%.

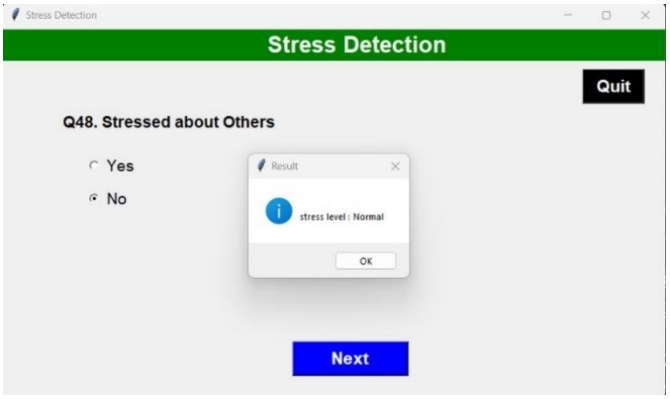


FIGURE 6. Stress Detection

In this conclusion, the students will be able to answer the questions and the system will predict the stress level of students with the help of responses.

REFERENCES

[1]. K. Parthiban, Digvijay Pandey, Binay Kumar Pandey, "Impact of SARS-CoV-2 in Online Education, Predicting and Contrasting Mental Stress of Young Students: A Machine Learning Approach," Augmented Human Research 08 June 2021.
[2]. Shivangi Dhawan "Online Learning: A Panacea in the Time of COVID-19 Crisis," Journal of Educational Technology Systems, June 2020.
[3]. Ravinder Ahuja, and Alisha Banga, "Mental Stress Detection in University Students using Machine Learning Algorithms," Ravinder Ahuja et al. / Procedia Computer Science 152 (2019).
[4]. U Srinivasulu Reddy, Aditya Vivek Thota, and A Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.
[5]. Garima Verma, Hemraj Verma "Model for predicting academic stress among students of technical education in India" International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 04, 2020.
[6]. Rumana Rois, Manik Ray, Atikur Rahman, and Swapan K. Roy, "Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms," Journal of Health, Population and Nutrition 40, 2021.
[7]. Reshma Radheshamjee Baheti, and Supriya Kinariwala, "Detection and Analysis of Stress using Machine Learning Techniques," International Journal of Engineering and Advanced Technology (IJEAT), Volume-9, October 2019.

[8]. Kavita Pabreja, Anubhuti Singh, Rishabh Singh, Rishita Agnihotri, Shriam Kaushik, and Tanvi Malhotra, "Stress Prediction Model Using Machine Learning," Proceedings of International Conference on Artificial Intelligence and Applications. Advances in Intelligent Systems and Computing, 2021.

[9]. Shruti, M. Harshini, Jeethu Philip, and I. Haritha, "Stress Level Detection of IT Professionals Using Machine Learning,"13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022.

[10]. S. Elzeiny, and M. Qarage, "Machine Learning Approaches to Automatic Stress Detection: Review," IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), October 2018.