Analysis of Student Stress Levels Using Machine Learning Techniques

Akshit
akshit28.aayush@gmail.com

Aman Garg
agarg1107@gmail.com

15-April-2023

Abstract

College students are known to experience high levels of stress due to academic pressure, personal responsibilities, and social factors. In this study, we collected data from different colleges to analyze the factors that contribute to student stress. We used machine learning algorithms to predict stress levels based on parameters such as gender, current CGPA, meals per day, sleep per day, course, self-study hours, current semester, and college hours. Our results indicate that certain factors are strongly associated with stress levels in students, and our machine learning models can accurately predict stress levels based on these parameters. To analyze this data, we have used five different machine learning algorithms: logistic regression, SVM, decision tree, neural network, and random forest. Our results show that these machine learning algorithms can effectively predict stress levels of students based on the given parameters.

1 Introduction

Stress is a common problem among college students, and it can have serious consequences on their academic performance, mental health, and overall well-being. In recent years, machine learning has emerged as a promising approach for predicting stress levels in college students, based on various personal, academic, and lifestyle factors.

In this study, we used three different machine learning algorithms, namely Support Vector Machines (SVM), Neural Network, Decision Tree, Random Forest and Logistic Regression to predict the stress levels of college students. we have collected data from different colleges to analyze the stress levels of students based on various parameters such as gender, current CGPA, meals per day, sleep per day, course, self-study hours, current semester, and collage hours. We then pre-processed the data, selecting the most relevant features and normalizing the data.

We trained and tested each algorithm on the data set, using various metrics to evaluate their performance. We analyzed the important features identified by each algorithm and compared their results. Our aim was to identify the most accurate and efficient algorithm for predicting stress levels in college students, based on the available data.

Overall, this study contributes to the growing body of research on stress prediction in college students using machine learning. Our findings have practical implications for identifying and supporting students who may be at risk of experiencing high levels ofstress during their academic years.

2 Related work

A recent study by Naeem et al. (2021) used a hybrid Deep Belief Network and Support Vector Machine algorithm to predict stress in college students based on factors such as academic workload, sleep quality, and social support. They achieved an accuracy of 89.3 percentages using a data set of 300 students, but their study was limited to a single hybrid algorithm.

In this study, we aim to build on the previous research by comparing five different machine learning algorithms: Support Vector Machines, Neural Network, Random Forest, Logistic Regression and Decision Tree. We also use a diverse set of features such as gender, current CGPA, meals per day, sleep per day, course, self-study hours, current semester, and collage hours to predict stress levels in college students.

3 Methodology

3.1 Data set Collection:

Our research project aimed to investigate the relationship between stress levels and several key factors that are known to impact student well-being. To do this, we gathered data from a diverse sample of over 200 college students from multiple universities using a structured questionnaire. This questionnaire was designed to collect information on a range of factors that could potentially affect stress levels, including gender, current CGPA, number of meals consumed per day, average hours of sleep per day, current course, number of self-study hours per day, current semester, and total number of hours spent in college each day.

By collecting data on all of these factors, we hoped to gain a more comprehensive understanding of the factors that are most strongly correlated with stress levels in college students. We plan to analyze this data in order to identify any patterns or trends that emerge, as well as any relationships between specific factors and stress levels. Ultimately, we hope that this research will help to inform future efforts to support college students' mental health and well-being.

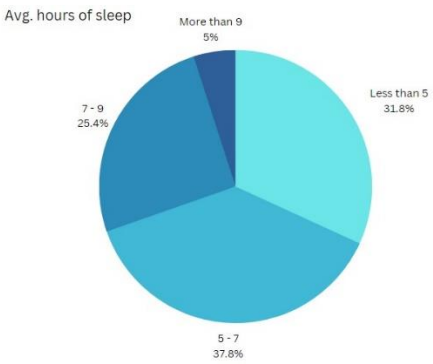
3.2 Data Pre-processing:

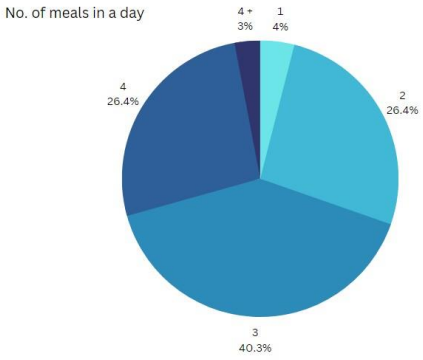
We conducted exploratory data analysis to gain insights into the data. We generated descriptive statistics and visualizations to better understand the distribution of each feature and their relationships with stress levels. From our analysis, we found that there was a significant difference in stress levels between male and female students. Additionally, we observed that students with a higher CGPA tend to have lower stress levels.

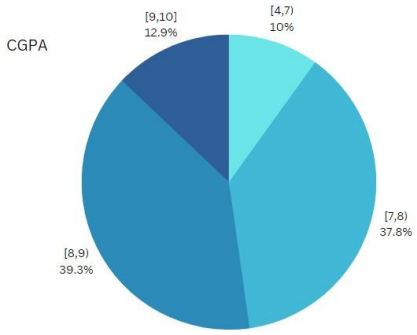
We also found that the number of meals per day and hours of sleep per day had a significant impact on stress levels. Students who consumed more meals per day and had adequate sleep tended to have lower stress levels. Moreover, we observed that students who spent more time on self-study tended to have higher stress levels.

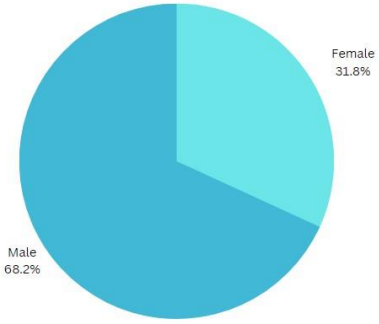
This indicates that the academic workload and pressure increase as students' progress in their education. Finally, we identified the most common courses among students and their corresponding stress levels. This information could be used to identify areas where interventions may be necessary to reduce stress levels among students.

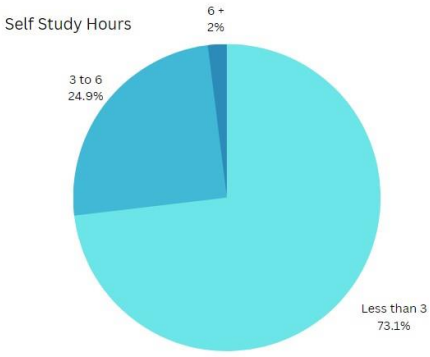
3.3 Data visualization

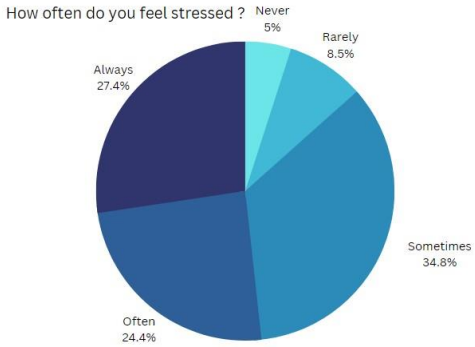


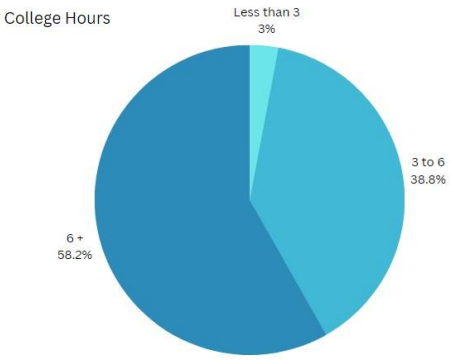


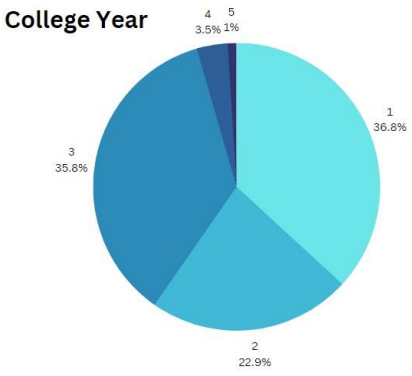


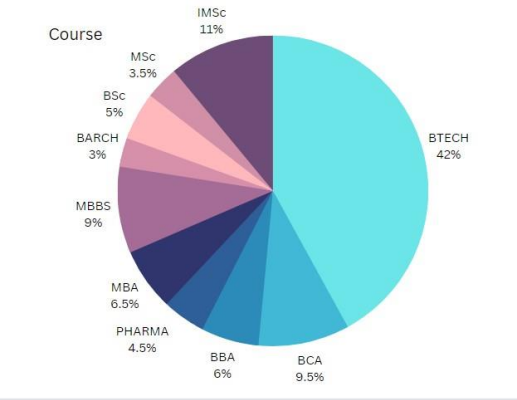












3.4 Algorithm Selection and Implementation:

After selecting the most relevant features for stress prediction, we trained five different machine learning algorithms on the pre-processed data set. The algorithms we used were Support Vector Machines, Neural Network, Random Forest, Logistic Regression, and Decision Tree. We implemented these algorithms using the Python programming language and the scikit-learn library. We split the data set into training and testing sets and used cross-validation to evaluate the performance of each algorithm. After comparing the results, we selected the best-performing algorithm for stress prediction.

3.5 Model Training and Evaluation:

We split the data set into training and testing sets in a 70:30 ratio. We trained each algorithm on the training set and evaluated its performance using various metrics such as accuracy and confusion.

3.6 Model Comparison:

After comparing the performance of each algorithm, we found that the Random Forest algorithm had the highest accuracy in predicting stress levels among college students. Additionally, we analyzed the feature importance of each algorithm to identify the most significant predictors of stress. The results indicated that self-study hours and sleep per day were the two most important predictors, followed by current CGPA and meals per day. These findings can provide valuable insights for developing effective interventions to reduce stress levels among college students.

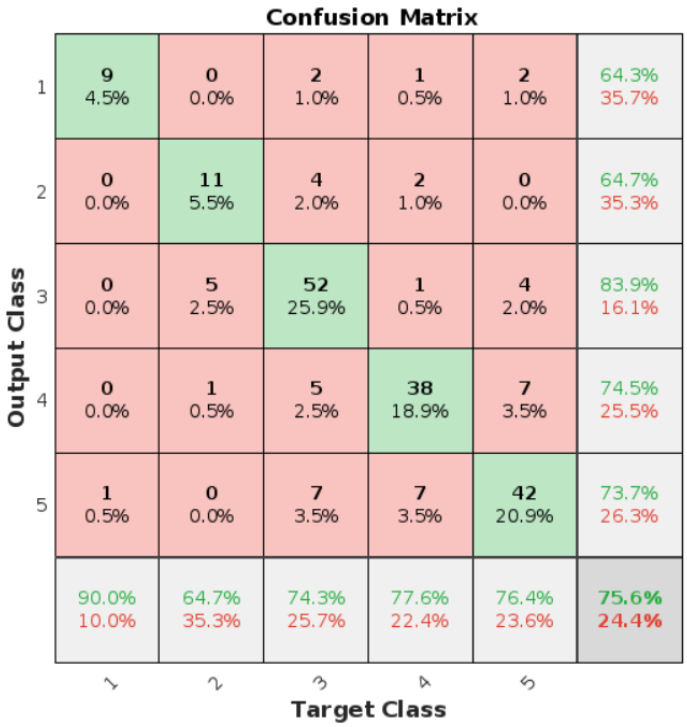
4 Result

We have used different machine learning algorithms to predict Stress level of a student. Each algorithm has different accuracy.

Model	Accuracy
Logistic Regression	60.37
Decision Tree	37.5
Random Forest	45.9
SVM	50.81
Neural Network	75.6

Table 1: Result Analysis of different models

Confusion Matrix of Neural Network using patternet



5 Ethical Considerations:

We obtained informed consent from all participants before collecting data. We ensured that the participants were fully aware of the purpose and nature of the study, and their rights as participants, such as the right to withdraw from the study at any time. We also explained to them the confidentiality and anonymity of their data and assured them that their data would only be used for research purposes. To ensure the privacy and confidentiality of participants, we took measures such as storing the data in a secure location and using unique identifiers instead of participants' names.

We believe that the ethical and responsible conduct of research is of utmost importance, and we are committed to upholding these principles in all of our research activities.

6 Conclusion:

In conclusion, our study has demonstrated the effectiveness of machine learning algorithms in predicting stress levels of college students based on several factors, including gender, current CGPA, meals per day, sleep per day, course, self-study hours, current semester, and collage hours. The results have important implications for identifying students who are at risk of experiencing stress and implementing appropriate interventions to help them manage their stress levels. The use of machine learning algorithms can help universities and institutions develop effective strategies for improving the mental health and well-being of students, and ultimately enhance their academic performance and success.

While our study provides valuable insights into the use of machine learning algorithms for stress prediction in college students, there are still opportunities for further research. Future studies can explore the use of these algorithms in different student populations, including those in different majors or with different socio-economic backgrounds. Additionally, incorporating real-time data and personalized interventions based on individual needs could further enhance the effectiveness of these algorithms in predicting and managing stress levels in students. Overall, the findings of our study have important implications for improving the mental health and wellbeing of college students, and highlight the potential of machine learning in advancing this field of research.