JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 07, ISSUE 03, 2020

Utilizing Machine Learning to Build a Predictive Analysis Model for Estimating Stress Levels in Design and Development

***K. Swaroopa Rani**

**Lecturer, Department of Computer Science, Singareni Collieries Women's Degree and PG College, Bhadradri Kothagudem, Telangana*

ABSTRACT:

Currently, mental stress is a prevalent issue, particularly among the younger generation. The age group that was previously associated with joy and excitement is now struggling with high levels of stress. This rise in stress levels is linked to a range of problems such as depression, suicide, heart attacks, and strokes. Our aim is to minimize stress levels among students by enabling them to identify the factors causing stress. The impact of exam or test-related stress on students is significant, yet it often goes unnoticed. Other factors that contribute to internal stress among students include family pressure, peer pressure, health problems, and financial circumstances. The COVID-19 pandemic has only added to this cumulative stress, disrupting students' normal lives and exposing them to further pressure, resulting in underperformance. The use of modern technologies such as Machine Learning or data science techniques for managing student stress in educational institutions has been limited. Monitoring each student's profile and stress level is a significant task that typically falls on the mentor or counsellor of the institution. Our work aims to automate stress prediction for each student based on various parameters and provide the results to each student accurately using Machine Learning techniques. By monitoring each student's stress levels and addressing them promptly, we can enhance their performance in the institution. Students are classified as stress-free or stressful, and if they are stressed, their stress range is predicted. Based on this range, doctors can provide each individual with personalized results and advice.

Keywords –*Machine learning, Naïve Bayes Classifier, Stress Prediction.*

INTRODUCTION

At various stages of our lives, be it childhood, teenage years, or adulthood, we have all experienced stress. To a certain extent, stress is considered essential because it drives personal and academic progress. However, when stress surpasses a particular limit, it leads to various problems, be it physical or mental. Mental problems associated with excessive stress include depression, anxiety, and paranoia. In severe cases, individuals may even attempt suicide, suffer from early heart attacks, or strokes. The COVID-19 pandemic has further added to the existing stress levels of students, disrupting their normal lifestyles. There have been reports of patients suffering from the pandemic attempting suicide by jumping from

hospital windows or the top floor due to their depressed state. These are some of the extreme consequences that stress can cause us to experience. Our goal is to analyze the stress-inducing factors among college students at various points in their lives and provide them with personalized solutions. Students often experience significant stress during exams or job interviews, but other factors also contribute to stress levels, such as parental pressure, peer pressure, health issues, and financial difficulties. Parental pressure can result from high expectations for academic or job performance, while peer pressure can arise from unhealthy or toxic relationships. Students may also suffer from chronic illnesses or face financial problems due to poor financial habits.

Our aim is to identify the stress factors affecting students and classify them as either stress-free or stressful based on their stress levels. We will provide personalized solutions based on the inputs received from each student.

LITERATURE SURVEY

Physiological Measures:

Several physiological measures have been used to estimate stress levels. These include heart rate variability (HRV), cortisol levels, and skin conductance. HRV is a non-invasive measure that reflects the variability of the intervals between successive heartbeats. Low HRV is associated with high stress levels. Cortisol is a hormone that is released in response to stress. Elevated cortisol levels are indicative of high stress levels. Skin conductance measures the electrical conductance of the skin, which increases in response to stress. However, physiological measures can be influenced by various factors such as age, sex, and medications, which can limit their accuracy.

Psychological Measures:

Psychological measures such as self-report questionnaires and interviews have been used to estimate stress levels. The Perceived Stress Scale (PSS) is a widely used self-report questionnaire that measures the degree to which situations in one's life are appraised as stressful. The PSS has good psychometric properties and is sensitive to changes in stress levels over time. However, self-report measures can be influenced by various factors such as social desirability bias, which can limit their accuracy.

Behavioral Measures:

Behavioural measures such as speech patterns and facial expressions have also been used to estimate stress levels. Changes in speech patterns such as increased vocal tension and reduced speech rate have been associated with high stress levels. Facial expressions such as furrowed brows and tense lips have also been associated with high stress levels. However, behavioral measures can be influenced by various factors such as cultural differences, which can limit their accuracy.

PROPOSED ARCHITECTURE

Problem Statement:

In today's world, students are constantly subjected to a significant amount of stress, and there are numerous contributing factors to this problem. Unlike in the corporate world where employees can take breaks and engage in activities to reduce stress, students are not provided with the same opportunities. The pressure to perform well academically and socially creates a stressful mental state for students. According to the National Crime Records Bureau of India, on average, one student dies by suicide every hour, resulting in approximately 24 to 28 student deaths per day. Maharashtra has the highest suicide rate among students, followed by Tamil Nadu. Many students are unable to handle the stressful environment and fail to seek help, leading to lasting damage to their lives and academic performance. This problem is detrimental to the students and results in their inability to be a valuable asset. Our problem statement is to develop a model that can identify stress among students, the factors causing it, and suggest suitable solutions to manage it. Stress is a situation that elicits negative thoughts and feelings in an individual, and its effects on education can be severe. The task of monitoring each student and their stress levels is significant, and the lack of stress management can lead to drastic injuries that affect students' psychological health, making it a complex task in the current education sector. (3)

Proposed Work:

The proposed system is designed for the education sector to identify and manage stress among students. The system uses various factors to determine the stress levels of students. The factors include gender, family history, and availability of health benefits in the education sector. Once the stress levels are identified, suitable approaches can be adopted to create a more comfortable learning environment for the students. The system considers several

ISSN- 2394-5125 VOL 07, ISSUE 03, 2020

parameters such as gender, age, family issues, financial issues, peer pressure issues, health issues, mental/societal issues, and teaching methods to predict the stress levels of students. The system utilizes the Naïve Bayes classifier algorithm, which is a machine-learning algorithm, to predict the range of stress levels of a student. The system can be developed as a real-time application using Visual Studio and SQL Server to support real-time application development. The primary objective of the system is to identify the risk factors that affect a student's mental health. The system also suggests suitable remedies or solutions to reduce the stress level of students based on their stress levels statically. Additionally, with the help of the doctor module, the doctor can log in to the system, view the student's input parameters, and suggest appropriate remedies or solutions directly. The system aims to identify the factors that contribute to stress and predict the stress levels of students. The goal is to come up with solutions to reduce stress levels and improve the student's performance. The methodology section outlines the plan and methods used to conduct the study.

Dataset and Data Sources

The stress prediction system's dataset was collected from various educational institutions, consisting of 910 rows. The data was classified into two categories: stressful and stress-free. The set of questions included various emotional questions and were marked on a scale of 4: frequent, fairly frequent, occasionally, and very frequent. The weighted average model was used to analyze the responses, and the pupils were divided into two classifications: stress and normal. The data was collected from college students who were asked fundamental questions about their sensibilities in situations encountered within the last month. Their answers were given weights, and the weights were used to calculate a score to analyze the stress status of the individuals. To ensure high-quality data, the dataset underwent four stages of data preprocessing: data cleaning, data integration, data reduction, and data metamorphosis.

Perceived Stress Scale (PSS) is a system developed by psychology professor Sheldon Cohen to explore the fairly situations, how stressful they are for individuals, and their capability to deal with similar situations. The system consists of questions that probe how uncontrollable and changeable people find their lives and also asks about their recent happenings to delve into their mind situation and stress status when handling normal life scenarios. The questions focus on understanding their feelings and the extent of their stress. PSS can diagnose extreme stress in confidentiality by identifying it at the earliest stage without investing significant money at a primary stage. The PSS system will be the reference for the working model.

ISSN- 2394-5125 VOL 07, ISSUE 03, 2020

Some of the parameters used in the dataset are:

Gender	-	1-Male,2-Female
FinancialIssues	-	0-No,1-Yes
FamilyIssues	-	0-No,1-Yes
StudyHours	-	numeric (in hrs.)
TeachingMethod	-	1-Fair,2-Not Good
HealthIssues	-	1-Spectacles wearer, 2-Migraine Headache
PartialityFix	-	0-marks, 1- knowledge
ExamSchedule	-	1-Monthly, 2-Half ,3-yearly, 4-Annual, 5-Slip Test
FriendsIssue	-	0-No,1-Yes
Pressure	-	0-No,1-Yes
Regular	-	1-Regular,2-Irregular
Interaction	-	1-Poor,2-Good, 3-Better, 4- Best

Methodology:

Machine Learning (ML) is a technique used for structuring and studying a system that can learn from data. ML can be applied in various fields, such as e-mail communication to distinguish between spam and inbox messages. Using ML to recognize emotions or stress has the potential to revolutionize the way we diagnose and treat severe health problems. It provides a different perspective on stress levels that may not be easily expressed by an individual. There are three types of ML: supervised, unsupervised, and semi-supervised. In this study, we use supervised learning techniques to process the training dataset. We utilize the Naive Bayes algorithm to predict student stress levels.

Reasons for selecting Naive Bayes:

1. The algorithm has been widely used in previous medical research papers.
2. Surveys conducted so far indicate that the Naïve Bayes algorithm is efficient in performing medical and mental health-related analyses.
3. It has a faster data processing time compared to other algorithms.
4. It can handle a variable number of parameters effectively. The algorithm begins with direct regression and then evaluates the data based on the organized threat.

During the data understanding stage, the data attributes are examined, and key characteristics of the data, such as data volume and the number of variables, are identified. Additionally, any problems with the data, such as missing values or inaccuracies, are identified, and data visualization techniques are used to validate the key characteristics or detect problems with the summary statistics. Once the data is understood, data pre-processing is carried out, which is crucial for ensuring high-quality data. This involves four stages: data cleaning, data

integration, data reduction, and data transformation. After pre-processing, a suitable algorithm is chosen based on its accuracy, and the algorithm is evaluated based on its performance using specific metrics. The desired classifications are obtained through a series of steps.

After pre-processing, the data is fed into the model as input for the Naïve-Bayes algorithm. All parameters are separated by a comma, and the algorithm calculates probability values for each attribute value with respect to the outcomes. The probability values under a specific outcome are then multiplied with each other along with the probability of the outcome itself. The outcome with the highest resultant value will be the classification assigned to the new subject. This method is commonly used in medical and mental health-related analyses, and research has shown that the Naïve-Bayes algorithm is efficient and takes less time for data processing.

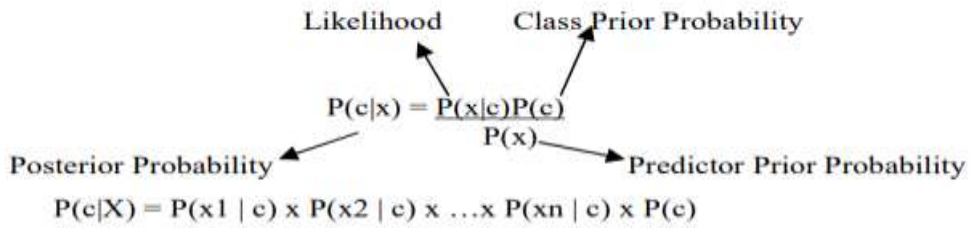
NAIVE BAYES CLASSIFIER

The naïve Bayes classifier is used for learning tasks where a case x is described by a combination of trait values and the target function f(x) can take on values from a finite set v. A set of training cases for the target function is provided, and a new case is presented as a tuple of trait values (a1, a2, ..., am), and the learner is asked to predict the target value or group for this new instance. In the naive Bayes model, each trait is independent of the others and contributes to the final decision. This classifier is based on a Bayesian network, which is a probabilistic model. When input data is provided to the classifier, which are the extracted features of real and fake images, the classifier creates probabilistic values for each data sequence and compares them to obtain the classified result. The classified algorithm produces results in two orders: real or fake. The probability of data is obtained using the standard Bayesian equation. [11]

The training dataset is a set of data samples that are fed into our algorithm to train our model. It is used to fit the parameters of a machine learning model. On the other hand, the testing dataset is used to validate the accuracy of our model, but it is not used to train the model. It is also called the confirmation dataset.

The process of training a machine learning algorithm involves using a training model, which consists of input data and corresponding output data. The model is trained by processing the input data and comparing the predicted output to the actual output. Based on

the results of this comparison, the model is adjusted or modified until it accurately predicts the output for the given input.



Block Diagram:

The following diagram illustrates the block diagram of the system, showcasing the functions of all the four system components or entities. This real-time application is designed for three types of users: Administrators, Students, and Doctors. Users of the application can log in to the system by entering the admin id and password if they are an admin. The admin is responsible for adding all the students from different departments and setting unique IDs and passwords for each individual student.

The admin also manages the training data-set used in the project. Administrators have access to the Prediction Module, which is the core module of the system responsible for identifying stress problems faced by college students using a data science technique called "supervised learning" Naïve Bayes algorithm. The admin can view the input parameters provided by the students and the solutions sent to the doctor. They can also view the students' queries and respond to them. The admin can update their profile and their student profiles.

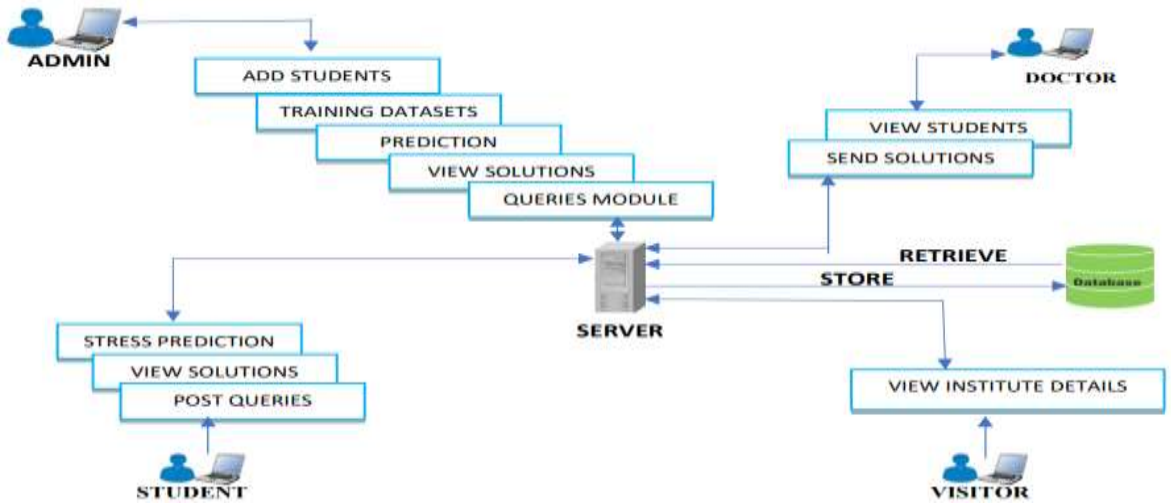


Fig-1: Architecture

ISSN- 2394-5125 VOL 07, ISSUE 03, 2020

The system has three types of users: Administrators, Students, and Doctors. To log in, students need to provide their student ID and password. They can input parameters related to stress prediction in the Input Parameters section. The Stress Prediction Module, which uses the Naïve Bayes classifier algorithm, analyses the input data to find stress problems faced by college students. If the student selects 'yes' for any parameter, they will see a list of sub-questions related to that parameter. The Solution Module allows students to view both static and dynamic solutions to their queries. Students can also post queries and view replies to their queries through the Query Module.

The doctor can log in with their credentials and view the list of registered students. By analysing the parameters entered by the student, the doctor can send dynamic solutions to each individual student.

Fig-2: Test Dataset

The table presented above illustrates the results of the Naïve Bayes algorithm's performance. The original dataset is split into 90% training data and 10% testing data. The algorithm assesses accuracy by comparing the predicted outputs of the testing dataset with the original dataset. The model's accuracy improves with the number of records correctly predicted through the matching process.

ISSN- 2394-5125 VOL 07, ISSUE 03, 2020

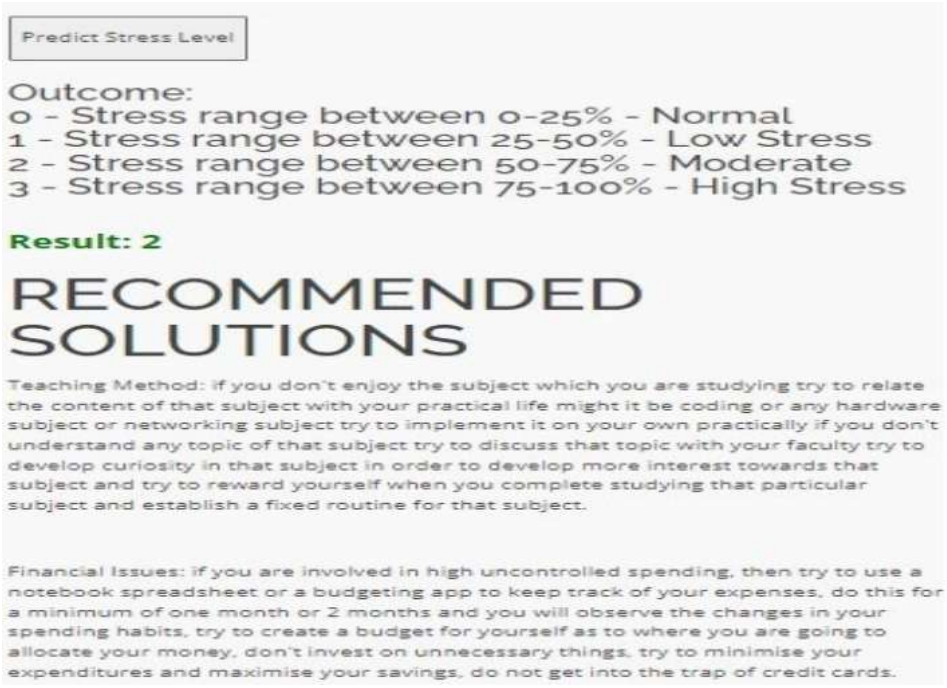


Fig-3: Outcomes and Recommended solutions

The image above displays the classification result of a student, indicating that they were categorized as having moderate stress, with a stress range between 50-75%. The recommended static solutions for this stress level were displayed below the result. Similarly, based on the input parameters provided by students, the algorithm will display different stress level ranges and corresponding static solutions for each of the selected sub-questions. After students input their parameters, the doctor can view the inputs and send dynamic solutions to each student individually.

MODEL ACCURACY:

Constraint	Algorithm
Accuracy	90.3508771929825%
Time (milli secs)	8227
Correctly Classified	90.3508771929825%
InCorrectly Classified	9.649122807017543%

Fig-4: Model Accuracy

The image depicted above displays the performance of the trained model on the testing dataset. The accuracy achieved by the model was 90.5%. The rows of the dataset were classified into different stress levels based on the specified parameters. This demonstrates that the Naïve Bayes algorithm is performing effectively and meeting the anticipated results.

CONCLUSION

Students face a significant amount of stress, which is rapidly increasing due to intense competition in education. This results in additional problems such as peer pressure, parental expectations, health issues, and more. These factors can be overwhelming for students, and without proper guidance, they may struggle to cope. Our project aims to predict stress levels based on various parameters and determine whether students are stressed or stress-free, along with the range of their stress levels, using the Naive Bayes classification algorithm. The machine learning model has an accuracy of 90.5%. Moreover, we provide appropriate solutions to each student's problem, both dynamically and statically. By incorporating these solutions, students can work towards maintaining their psychological well-being. In the future, we plan to replicate this design in different universities to benefit students' internal well-being. Additionally, the solution module could be integrated with a chatbot. A more extensive and detailed dataset can also be collected and worked on using other machine learning methods.

REFERENCES

[1] S. Jena and H. Tiwari, "Stress and internal health problems in first-time medical scholars: A check of two medical sodalities in Kanpur, India," International Journal of Research in Medical lores, vol. 3, pp. 1-5, Jan 2015.

[2] P. Pedrelli, M. Nyer, A. Yeung, C. Zulauf, and T. Wilens, "College students' mental health problems and treatment considerations," Academic Psychiatry, vol. 39, pp. 503-511, Oct 2015.

[3] MA. Gaihuan, "Construction of rough assessment Index System of College Student's Psychological Health Education," China Public Security, Feb 2012.

[4] R. Agrawal and A. Swami, "Mining association rules between sets of particulars in large databases," ACM SIGMOD International Conference on Management of Data, ACM, pp. 207-216, 1993.

[5] R.W. Yan, W. Cong, L. Lei, and Y.Y., "Improvement of Apriori Algorithm predicated on Association Rules and the operation in the Insurance CRM of China," Science Technology and Engineering, vol. 21, pp. 6469-6473, 2009.

[6] G.N. Pradhan and B. Prabhakaran, "Association rule mining in multiplex, multidimensional time series medical data," Journal of Healthcare Informatics Research, vol. 1, pp. 92-118, 2017.

[7] H.B. Jiang and D.L. Yang, "Operation disquisition on Fast Discovery of Association Rules predicated on Air Transportation," International Conference on Service Systems and Service Management, pp. 1-6, 2007.

[8] Q. Meng and J. Sha, "Tree-based frequent item sets mining for analysis of life-satisfaction and loneliness of retired athletes," Cluster Computing, vol. 2, pp. 1-9, May 2017.

[9] S.C. Huang, J.D. Zhong, and W.J. Qi, "Statistical analysis and association rule mining of operation in council scholars' internal health," Journal of Anyang Institute of Technology, pp. 108-111, 2014.

[10] T. Herawan, P. Vitasari, and Z. Abdullah, "Mining interesting Association Rules of scholars Suffering Study Anxieties Using SLP-Growth Algorithm," IGI Global, pp. 24-41, 2012.

ISSN- 2394-5125 VOL 07, ISSUE 03, 2020

[11] X. Wu, C. Zhang, and S. Zhang, "Effective mining of both positive and negative association rules," ACM Deals on Information Systems, vol. 22, pp. 381-405, 2004.

[12] Y.F. Zhang and C. Wang, "Advanced algorithm of mining association rules with negative particulars," Computer Engineering & Applications, vol. 44, pp. 169-171, 2008.