Predictive analysis of student stress levels using ML

Shreelakshmi CM [1], Samana K [2], Samiksha V [3], Samyuktha S [4], Shilpa L [5]

Assistant professor, Department of CSE, GSSS Institute of Engineering and Technology for Women, Mysuru, India[1]

Student, Department of CSE, GSSS Institute of Engineering and Technology for Women, Mysuru, India[2-5]

Abstract: College students are suffering from many mental health problems including mental stress, somatization, obsession, interpersonal sensitivity, depression, anxiety, hostility, fear, paranoia, and psychosis, which can bring a lot of negative effects to them. Obviously, the mental health problems of college students not only directly affect their own growth, but also affect the stability of the campus. Most colleges also pay more and more attention to students' psychological crisis monitoring and prevention. All Colleges simply analyze whether students have mental health problems or what kind of problems they have. It cannot find hidden relationships in the psychological data. 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 workload, assignments, family issues, friends' issues, attendance, teaching, etc. Machine learning is a subject to predict the future based on past data. Using machine learning techniques, we predict student stress levels, and also the proposed system will give suggestions based on the stress levels of students.

Keywords: Student stress levels, psychological data, machine learning.

I. INTRODUCTION

Students are facing so many mental health problems such as depression, pressure, stress, interpersonal sensitivity, fear, nervousness, etc. Though many industries and corporate provide mental health-related schemes and try to ease the workplace atmosphere, the issue is far from control. Stress Prediction in college students is one of the major and challenging tasks in the current education sector. Stress is regarded as a major thing that is used to create an imbalance in the life of every character and it is additionally regarded as a major issue for psychological adjustments and trauma reduction.

Numerous studies work on stress management in school students. The students who are pursuing their secondary and tertiary education are widely facing ongoing stress level issues. It can be many times decided as day-to-day movements for a hassle-free mind to pay attention to lecturers. To decrease the individual stress rate, human societies have been in a position to boost a complete stage of progress in monitoring the stress stage of students and make them score well in academics.

Lack of stress administration can result in some drastic injury which can sometimes affect education completely and can even cause extreme injury to the fitness of the students at a variety of stages. Individual family background has been conceptualized as a major play that has been taking a path from our childhood. Children who are residing in rural or cities are consistently perusing an exclusive environment. The fulfillment of the faculty and students is majorly structured on both faculties instructing and learning at home. The current system is a manual process where it is difficult to identify the stress in college students. There is no automation for students' stress prediction.

Machine learning algorithms can be used to predict student stress levels by analyzing various data sources such as academic performance, social behavior, and self-reported stress levels. The algorithms can learn patterns and relationships from the data to generate a predictive model that can identify students at risk of high-stress levels.

Some potential data sources for predicting student stress levels using machine learning include:

- Academic performance data, such as grades, attendance, and course workload.
- Social behavior data, such as participation in extracurricular activities and interactions with peers.
- Self-reported stress level data, such as responses to surveys or wearable technology data.

Using machine learning to predict student stress levels can provide valuable insights for educators and mental health professionals to create personalized interventions for students who need them most. It can also help schools and universities to create a supportive environment that promotes academic success and student well-being.

The rest of this paper is organized as follows: Chapter 2 presents the related work. Chapter 3 presents the proposed work. Chapter 4 presents the module description. Chapter 5 presents the experimental results for Automatic Helmet Detection. Chapter 6 presents the conclusion.

II.   RELATED WORK

Projective Exploration on Individual Stress Levels using Machine Learning by Nagendra Prabhu S, Dr. Amarjeet Singh, Bhavya B, K. Sreeja, Bhargavi P: Recently, Stress Prediction in every individual based on their profile and behaviour is a challenging task in the current sector. Current system is a manual process where it is difficult to identify the stress in the college students or employees. There is no automation for any stress prediction. System uses machine learning algorithms or AI algorithms to find out the stress levels, these technologies are used for application development. The purpose of this project is to reduce stress in students and employees. Computer science has come a long way in the last few years. It is massive and multifaceted. It has been used in a variety of applications to meet the basic needs of human society. In the field of healthcare, machine learning has made significant progress. Stress is a deadly disease that kills a large number of people worldwide. How machine learning techniques can assist in reducing the risk of stress prediction, is examined in this study.

Stress Prediction in Indian School Students Using Machine Learning by Ena Jain, Aanchal Bisht, Shreya Vashisth, Muskan Gupta: 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 K- Nearest Neighbors giving the highest accuracy of 88%.

Prediction of Public Mental Health by using Machine Learning Algorithms by CH.M.H.Saibaba, Alekhya K V K, Yeshwanth K, Praveen Tumuluru: Ongoing investigations on mental health have shown a solid connection between the way of life factors and mental health. Additionally, late healthcare concentrated on discovering that way of life factors was emphatically connected with the results. The point of a review is to look at relationships in the middle of mental health and human lead factors as smoking, proactive tasks, drinking, and eating inclinations with occupants at the urban or state level utilizing accumulated information. More explicitly, the principal point of the review was to determine the connection between Frequent Mental Distress (FMD) and human being conduct designs among United States populace. This exploration needs utilized information from the most recent six years of AHR report; individual investigations have been done on female and male datasets, then, at that point, contrasted the conduct factors related and psychological health along with both female and male. An investigation discovered that places that get low extreme consumption rate high-level corpulence, and the high smoke rate has the most noteworthy successive of psychological trouble. Additionally, these guidelines have demonstrated great certainty limit as well as women rather than guys.

Predicting Mental health disorders using Machine Learning for employees in technical and non-technical companies by Rahul Katarya, Saurav Maan: After analyzing, they found that decision tree classifier has the best performance. As it has the best accuracy and precision with accuracy 84% and precision 83 followed by logistic regression with 84% accuracy and 82 precisions followed by Naïve Bayes with 79% accuracy and 78 precision, random forest with 77% accuracy and 81 precisions, SVM with 76% accuracy and 75 precision and KNN has the worst performance with 74% accuracy and 76 precisions. Also, Feature importance of the selected features showed that a history of mental health disorder contributes most during disorder prediction followed by family history. It was also found that rest of the features contributes bare minimum to the prediction with gender as their top rest of the features which includes mental health benefits or care provided by the employer, age and discussing mental health status with the employer barely makes any contribution to the prediction of mental health disorder. Thus, this paper contributes to proving that gender and company type that is the tech and non-tech even though have some influence on mental disorders. It is not prominent enough to come to a conclusion people of a gender or company type are more prone to mental disorders than the rest. But to further prove this, we need more data.

Detecting work stress in offices by combining unobtrusive sensors by Saskia Koldijk, Mark A.: Neerincx proposes a method for detecting work stress in offices by combining data from unobtrusive sensors. The authors argue that traditional methods for measuring work stress, such as self-reporting and physiological monitoring, can be intrusive and affect natural behavior, leading to unreliable results.The proposed method involves using unobtrusive sensors, such as motion sensors, light sensors, and computer usage logs, to capture data related to work behavior and environmental factors. The data is then analyzed using machine learning algorithms to detect patterns that indicate the presence of work stress.The authors conducted a pilot study with 15 participants, using both self-reporting and the proposed sensor-based method to measure work stress. The results showed that the sensor-based method was able to detect work stress with an accuracy of 78%, compared to 60% for self-reporting.The paper concludes that combining unobtrusive sensor data with

machine learning algorithms can provide a non-intrusive and reliable method for detecting work stress in offices, which can be used to inform interventions to improve employee well-being and productivity.

III. PROPOSED SYSTEM

This section presents the proposed approach for real time detection of stress level among students which has two phases. The first phase is to detect if the student is stressful or stress-free and the second phase is to provide the customized solution for the stressful students based on the factors affecting them. The advantages/features of the proposed system are:

- Proposed system is meant for education sector.
- System finds factors that strongly determine the stress levels.
- Stress is identified based on gender, family history and availability of health benefits in the education sector.
- By identifying the stress of students, we can come up with some approaches to reduce stress and create a much comfortable learning place for students.
- System uses many parameters such as gender, age, family history, provided health benefits, share about illness, tech company, tech role, acquiring leave, etc.
- System uses machine learning algorithms or AI algorithms to find stress of a student.
- System can be developed as a real time application which is useful for colleges. As Visual Studio and SQL Server is more supportive with real time application, we use these technologies for application development.
- Finding the risk factors which affects the student's mental health is the major objective of the system.
- System also gives suggestions for the students based on the student stress levels.
- System aims at identifying the factors and predicting the student stress levels and to come up with solutions to reduce the stress level so as to improvise the students' performances.
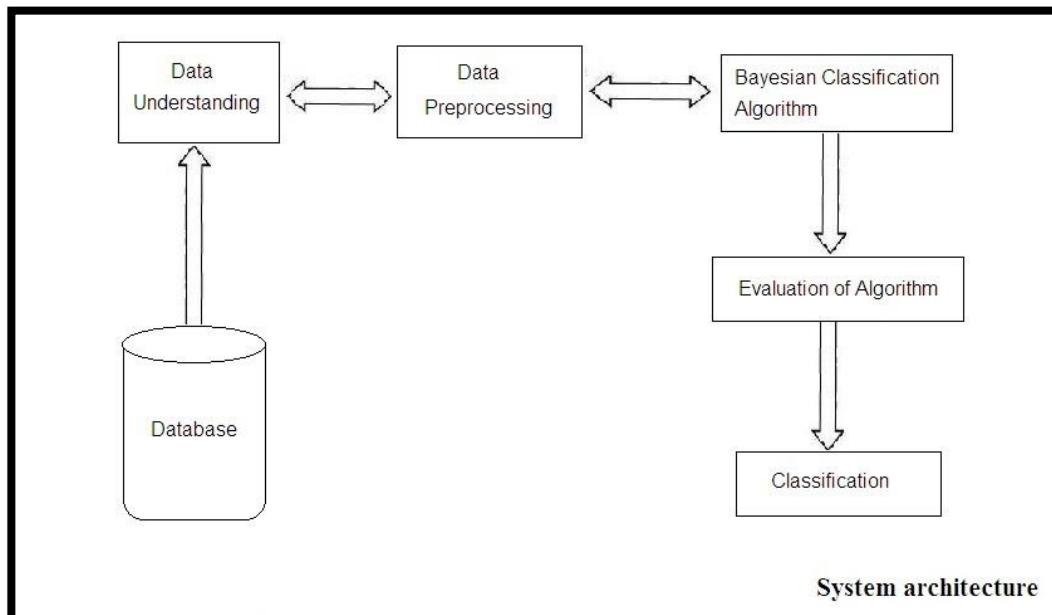

Figure 3.1: Stress prediction steps and flow

ML is concerned with the construction and study of systems that can learn from data. For example, ML can be used in E-mail messages to learn how to distinguish between spam and inbox messages.There are three types of Machine learning (ML) algorithms, they are

    i. Supervised Machine Learning: Here we have labels and the input is past examples.
    ii. Unsupervised Machine Learning: Extraction of patterns without labels.
    iii. Semi-Supervised Machine Learning: Mixture of both Supervised and Unsupervised Machine Learning

In this project, we use supervised learning techniques to process the training data-set. We use Naive Bayes Algorithm to predict student stress levels.

Reasons for selecting Naive Bayes:

- Most of the previous medical research papers use this algorithm.
- Survey says efficient algorithm to process medical data.

- Takes less time for data processing.
- Works fine for n number of parameters. Number of parameters need not be fixed.

IV. METHODOLOGY

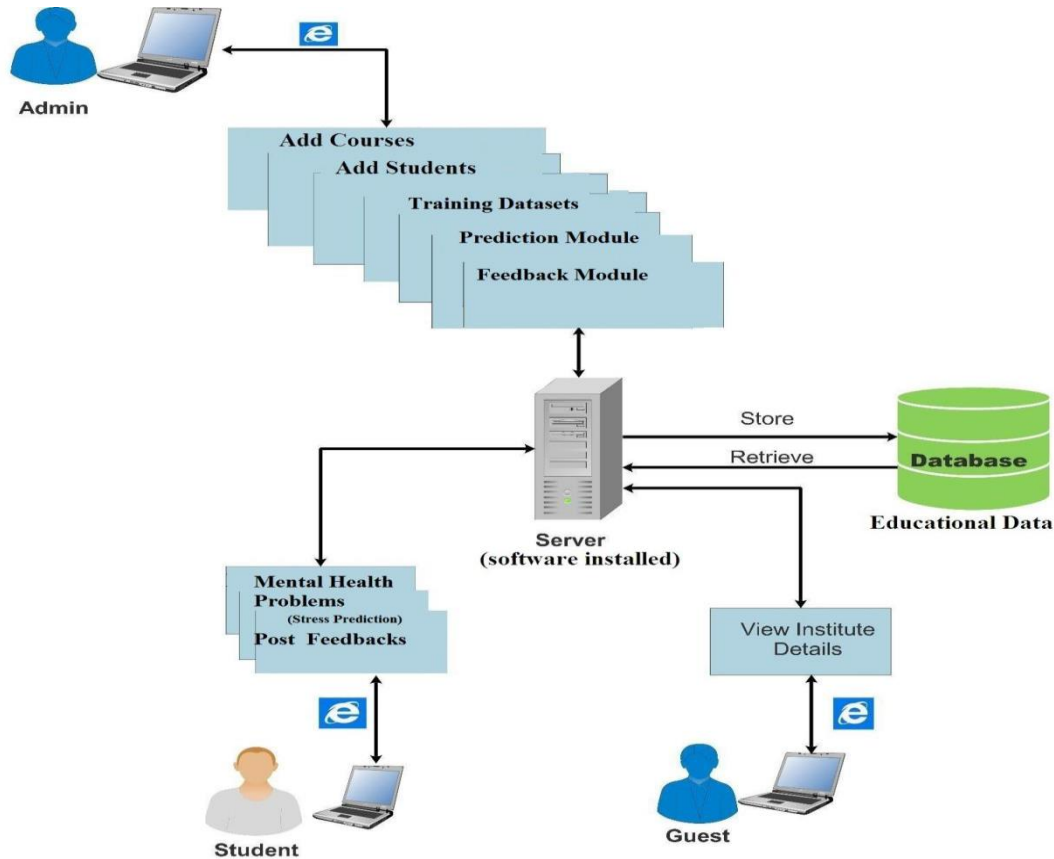The architecture diagram is shown below:

Figure 4.1: Architecture design

Data Collection (Stress Data):
This is the first step in the stress prediction process where we collect stress data. Data collected from many sources which contains parameters such as Gender, Age, Financial_Issues, Family_Issues, Health_Issues, Partiality_Fix, Pressure, Regular, Interaction, etc.

Data Preparation:
Here stress data was analyzed and only relevant data was extracted. The data required for processing is extracted and segmented according to the requirement. Required data extraction is done because entire data is not required for processing and if we input all data, it requires too much of time for processing, so data processing is done.

Bayesian classification algorithm:
Naive Bayes classification is a probabilistic algorithm used for classification tasks. It is based on Bayes' theorem, which describes the probability of an event occurring given the prior knowledge of conditions that might be related to the event. In Naive Bayes classification, the algorithm assumes that the input features are conditionally independent, which means that the probability of each feature occurring is independent of the other features. This assumption is called the "naive" assumption, as it is often not true in real-world data.

Stress Prediction:
The System predicts the stress based on the parameters using machine learning algorithm.

Results:
Here we find the accuracy of the algorithm by dividing the training datasets into training and testing datasets. 90% were considered as training datasets and 10% were considered as testing datasets.

Deploy model:

Once the model is trained and tuned, it can be deployed to predict the stress levels of new students. The new data can be fed into the model, and the model will output the predicted stress level of the student.
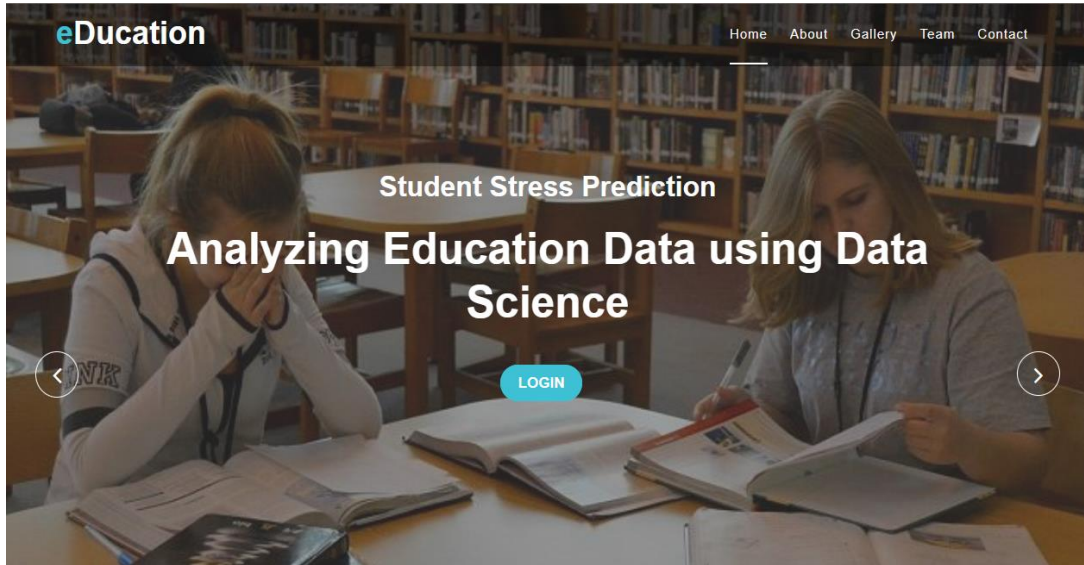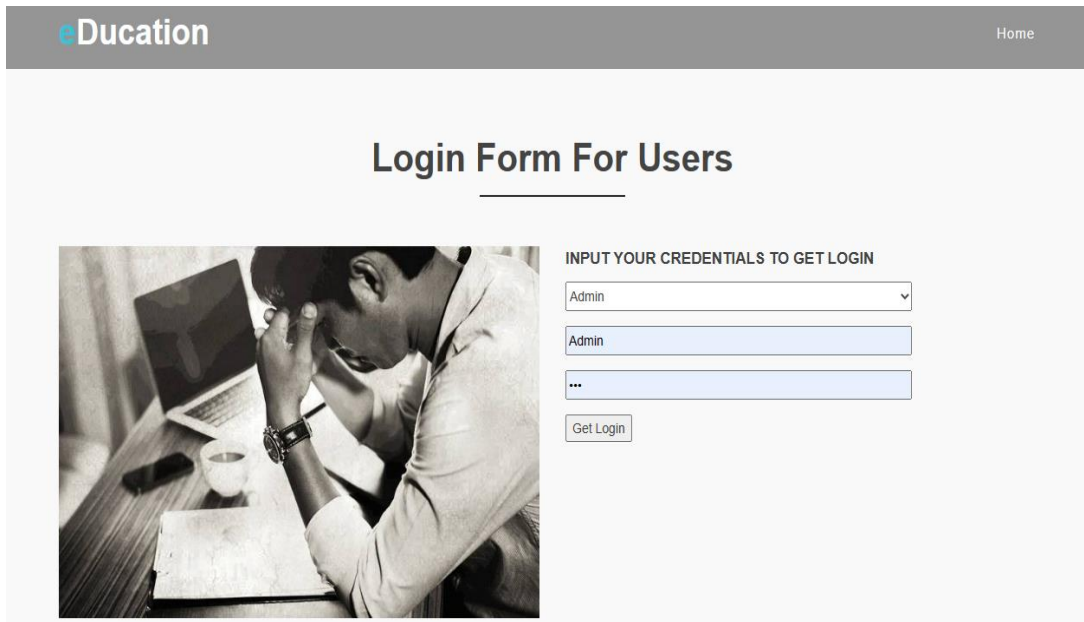
V. EXPERIMENTAL RESULTS

Figure 5.1: Home page

5.2:   Figure
page   Login
                                                                                                              for

both students and admin

The snapshot given above shows the login page of our project for both students and admin. Here admin can log in using login credentials such as username and password. If all login credential given by the admin is valid it directs to the main page or else it redirects to the login page. For students to log in, the admin should have already entered the students' details in the database, only then the student can access the page.
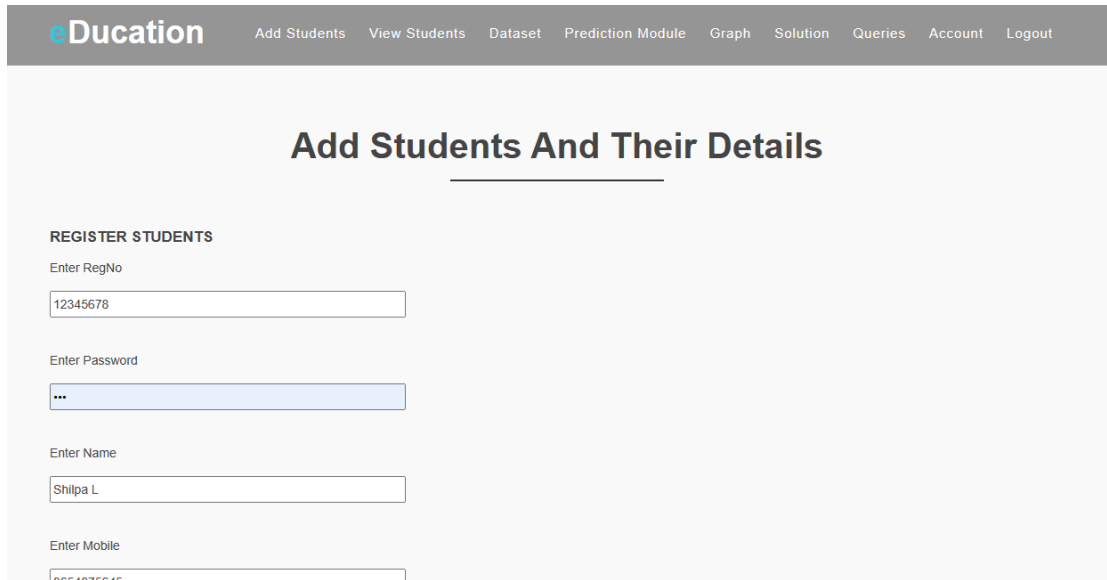
Figure 5.3: Admin Adding Students to the Database

The snapshot above shows the page where the admin can add the students' information to the database in order to predict the students' stress levels.
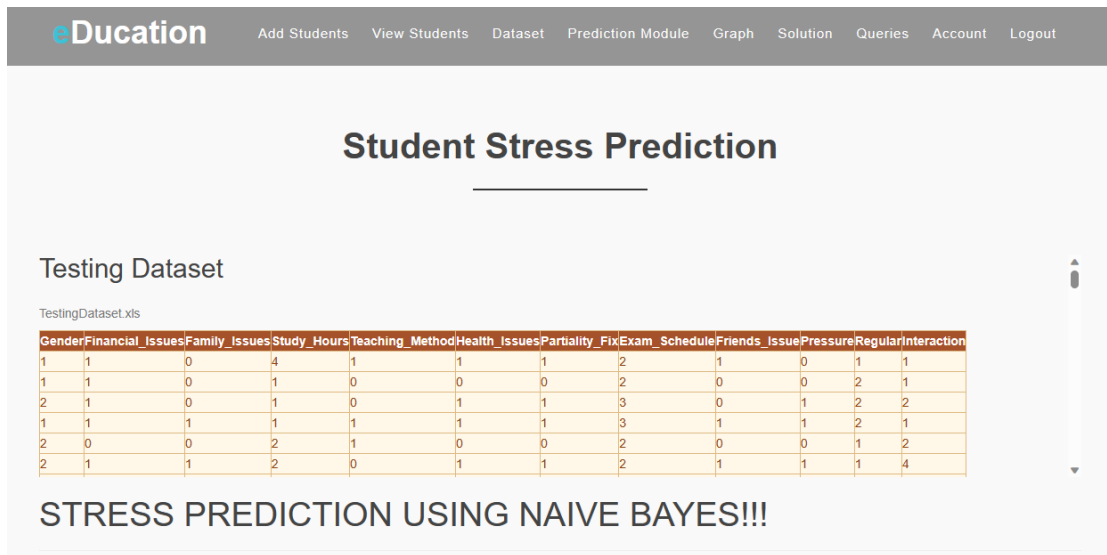
Figure 5.4: Student's Stress Prediction using the Testing dataset

The snapshot above predicts the stress of the students using the testing dataset provided by the algorithm which considers various factors as described above.
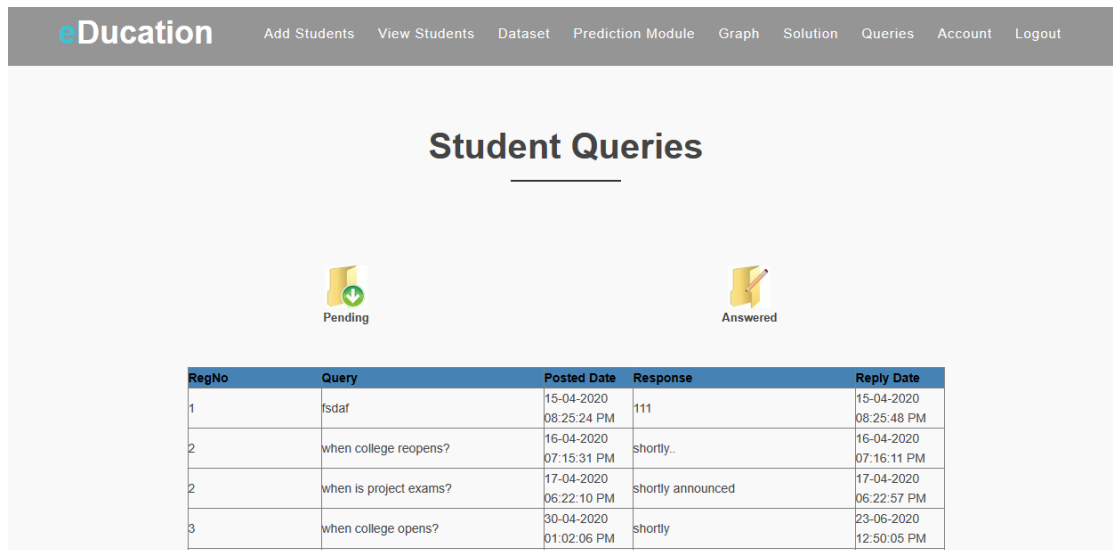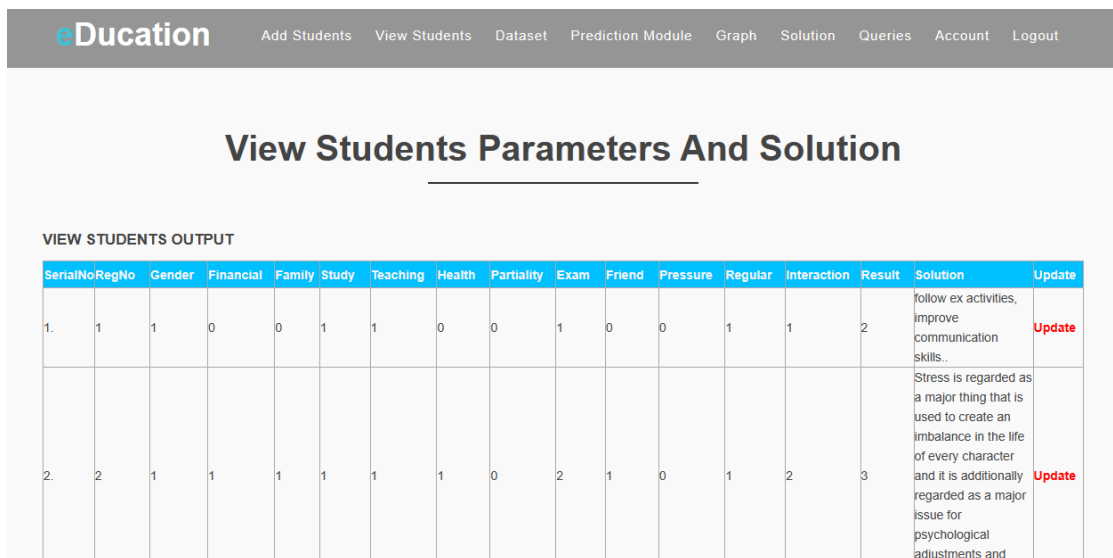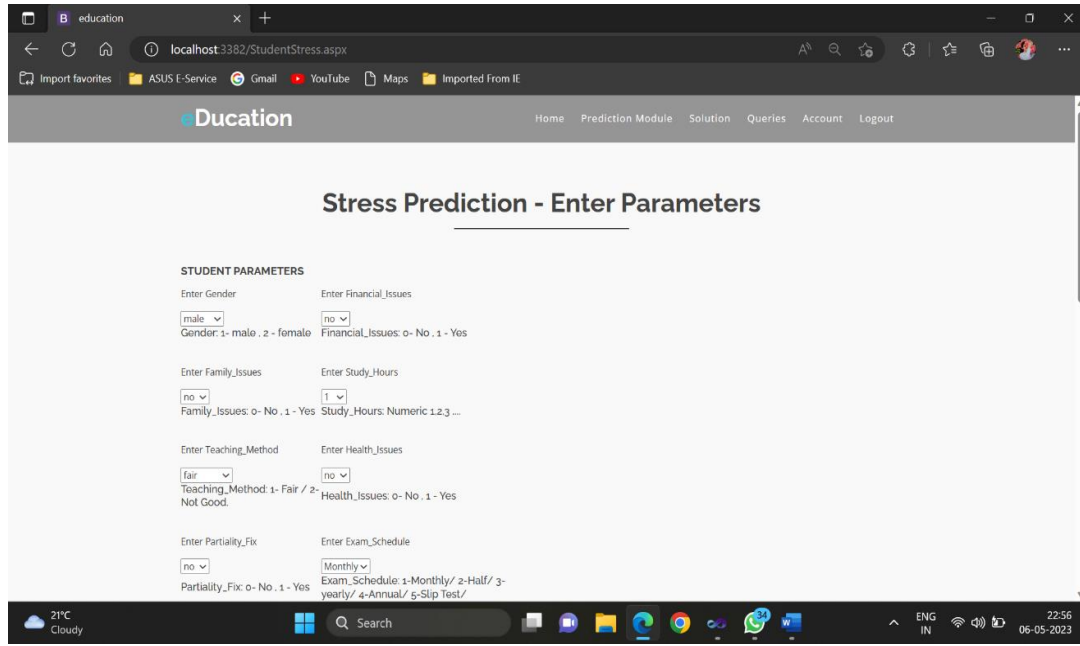
Figure 5.5: Admin Providing a Personalized Solution to the Student's stress.

Figure 5.6: Automatic Generation of Solution.

The above snapshots show how the solutions for the students' stress is generated with the help of admin and this solution is automatically provided for the students.
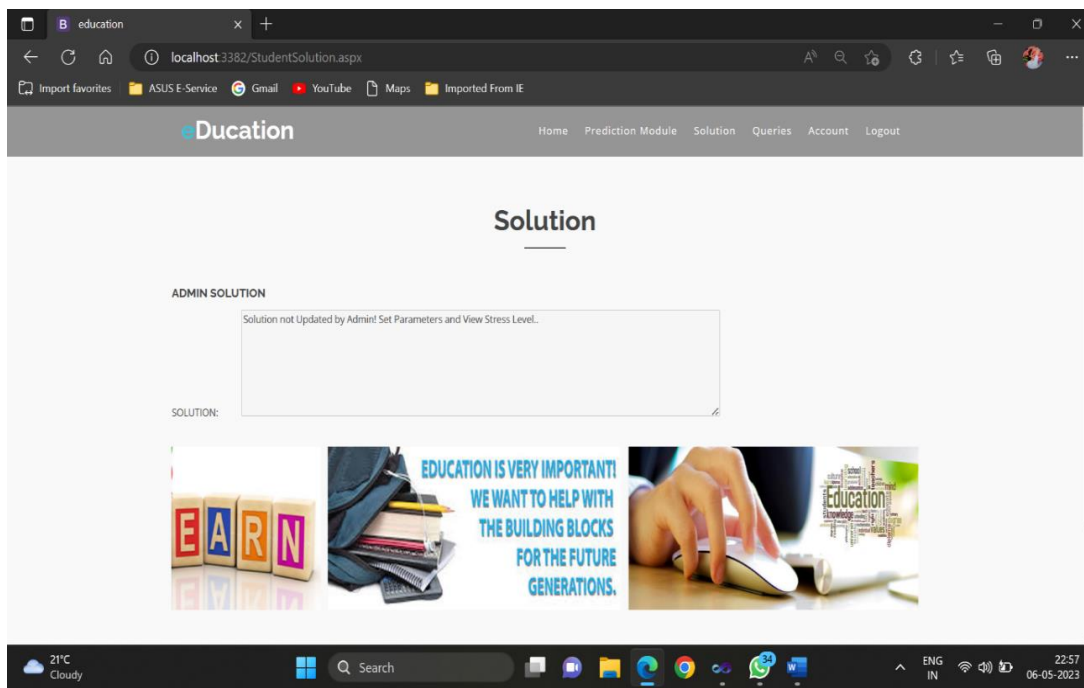
The next two snapshots show how a student can enter their problems which is considered as parameters for predicting their stress levels based on which the customized solution will be provided by the admin to overcome the stress among students.

Figure 5.7: Stress prediction parameters that should be entered by the students.

Figure 5.8: Solution provided by the admin.

VI. CONCLUSION

College students are suffering from many mental health problems including mental stress, somatization, obsessive, interpersonal sensitivity, depression, anxiety, hostility, fear, paranoia, and psychosis, which can bring a lot of negative effects to them. Machine learning is a subject to predict the future based on past data. Using machine learning techniques, we predict student stress levels, and also the proposed system will give suggestions based on the stress levels of students.

REFERENCES

[1]. Jena S, Tiwari H, "Stress and mental health problems in 1st year medical students: A survey of two medical colleges in Kanpur, India," International Journal of Research in Medical Sciences, vol. 3, pp. 1-5, Jan 2015.

[2]. Pedrelli P, Nyer M, Yeung A, Zulauf C and Wilens T, "College Students: Mental Health Problems and Treatment Considerations," Academic Psychiatry, vol. 39, pp. 503-511, Oct 2015.

[3]. Jenny K. Hyun, Brian C. Quinn, Temina Madon and Steve Lusting, "Graduate Student Mental Health: Needs Assessment and Utilization of Counseling Services," Journal of College Student Development, vol. 47, pp. 247-266, June 2006.

[4]. Gaihuan M A. "Construction of Fuzzy Evaluation Index System of College Student's Psychological Health Education," China Public Security, Feb 2012.

[5]. Agrawal R, Swami A, "Mining association rules between sets of items in large databases," Acm Sigmod International Conference on Management of Data, ACM, pp. 207-216, 1993.

[6]. Koh, H. C., and Tan, G., "Data mining applications in healthcare," Journal of Healthcare Information Management Jhim, vol. 19, pp. 64-72, 2005.

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

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

[9]. Jiang, H. B, and Yang, D L, "Application Research on Fast Discovery of Association Rules Based on Air Transportation," International Conference on Service Systems and Service Management, pp. 1-6, 2007. 723

[10]. Liu G, Jiang H, Geng R, and Li H, "Application of multidimensional association rules in personal financial services," International Conference on Computer Design and Applications, vol. 26, pp. 3877-3879, 2010.

[11]. Qi W, Yan J, Huang S, Guo L, and Lu R, "The application of association rule mining in college students′mental health assessment system. Journal of Hunan University of Technology," vol. 6, pp. 94-99, June 2013.

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

[13]. Huang S. C, Zhong J D, and Wen-Jua Q. I, "Statistical analysis and association rule mining of application in college students' mental health," Journal of Anyang Institute of Technology, pp. 108-111, 2014.

[14]. Long Z, Feng H, Xue D, and Xiangjun D, "Positive and Negative Association Rules Mining for Mental Health Analysis of College Students," Journal of Mathematics Science and Technology Education, vol. 13, pp. 5577-5587, 2017.

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

[16]. Han J, Pei J, and Yin Y, "Mining frequent patterns without candidate generation," ACM SIGMOD International Conference on Management of Data, pp. 1-12, 2000.

[17]. Wu X, Zhang C, and Zhang S, "Efficient mining of both positive and negative association rules," Acm Transactions on Information Systems, vol. 22, pp. 381-405, 2004.

[18]. Dong X J, Wang S J, Song H T and YC Lu, "Study on Negative Association Rules," Journal of Beijing Institute of Technology, pp. 978-981, 2004.

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

[20]. Dong X, Ma L, and Han X, "e-NFIS: Efficient negative frequent itemsets mining only based on positive ones," IEEE, International Conference on Communication Software and Networks,pp. 517-519, 2011.