

# Predictive Analysis of Student Stress Level using Machine Learning

Prakruthi Manjunath  
Department of Computer Science  
The National Institute of Engineering  
Mysuru, India

Pola Shreya  
Department of Computer Science  
The National Institute of Engineering  
Mysuru, India

Twinkle S  
Department of Computer Science  
The National Institute of Engineering  
Mysuru, India

Vismaya Ashok  
Department of Computer Science  
The National Institute of Engineering  
Mysuru, India

Dr. Shabana Sultana  
Professor & Controller of Examination  
The National Institute of Engineering  
Mysuru, India

**Abstract** - Students of today's era are invariably subjected to immense amount of stress, the contributing factors for this are in plenty. Many students are unable to cope up with the challenging and stressful environment and fail to receive help in the right way, thus leading to a persistent damage to their lifestyles. This is followed by performance degradation of the student and not being counted as an asset.

We propose a solution for the educational organization where the authorities can track the predicted stress percentages of each student enrolled. The student has the provision of taking up the survey, encompassing the parameters which is instrumental in bringing about mental distress and anxiety. The survey data is taken as the input for a pre-trained machine learning model which predicts the stress percentage of each student. A two-way classification of the stress level is brought about by the model as to whether the student is stress-free or stressful, and a further classification under stressful students about the range in which their stress percentages lie, as to low, medium or high is done. Based on the range of the stress level and the probabilistic parameters of stress, each stressful student is given a feedback and advisable solution from the educational institute. The student can adopt the solution and make way for his or her mental peace thereby reducing stress levels. Our work also enables the student to query his grief, and an apt answer would be received by the student from the authorities and the privacy of each student is maintained. The machine learning model is structured on the KNN-classification algorithm.

**Keywords** - *Machine Learning; KNN classification Algorithm; Stress Prediction*

## I. INTRODUCTION

The current educational system and the tough competition inclines the anxiety and stress amongst students. Other factors which contribute towards the mental disparity amongst students include parental pressure, peer pressure, health issues, financial conditions. The additive has been the pandemic of the corona virus, dysfunctioning the normalcy of student's lives and susceping them to more pressure thus leading to ill performance.

The automation for student stress prediction in institutes and educational organizations has been very minimal. Observing each student and his or her profile is a huge task. This responsibility lies under human interaction and that is why our work paves way for the automatic stress prediction of each student succumbing under various parameters and proposes the solution to each student rightly. This is done with the help of Machine learning and data science techniques. Keeping a check of each student's stress levels, and monitoring it closely, helps to heighten their performance in an organization.

Students are classified as to whether they are stress free or stress full and if they are stressful, their range of stress is highlighted. Based on this percentage, the authorities give each individual solutions and advice. This system reports the accurate predictions.

II. LITERATURE SURVEY

In Paper [2], Stress has been a havoc identified in software engineers. The authors have incorporated the OSMI (open sourcing Mental Illness) Survey dataset 2017 from tech industry. This dataset has been subjected to various machine learning strategies and the dataset hold the labels such as gender, age, family background etc. Their findings suggested that 75% of the people who worked in the IT industry are prone to pressure. Few of the techniques which they have worked on include boosting, packing, decision trees etc.

In Paper [3], authors have used decision tree algorithm which is applied on the data of students enrolled for an academic year and their stress levels are recorded at the beginning and at the end of the semester. The result is such that the model identifies students with more stress at the semester end, rather than at the beginning of the semester.

In Paper [4], authors have measured stress using different modes like EEG, GSR, EMG and SpO2. These were used to record or measure the automated stress detection. The measurement recorded from all the external and internal sensors were verified with the value of index which was initially set as a mark for stress prediction.

Paper [5] focuses on the occupational stress. A survey was conducted at different sectors to collect the data. It focused on three factors, i.e., psychosocial, environmental, and physical factors. Analysis is done using metric (Support Vector Machine and neural Network) and non-metric approaches. The metric approaches provide good accuracy.

III. METHODOLOGY

Machine Learning is a subset of Artificial Intelligence which trains the model without human Intervention. It has been categorized into Supervised, Unsupervised and semi-Supervised learning techniques.

The dataset has been collected and stored in an excel sheet. A method known as binning is implemented which removes the outliers and fills up the missing values with the highest count of the encoded numerical values of a respective parameter. This feature aids in maintaining the consistency of the data eliminating duplicate values. Further the data set is split into the training and testing data. The training data is fed into the model and the model is trained. The model can then predict the values for the testing dataset.

Our main objective is to train the model to give its best performance from the pre-processed data. To achieve this, a Supervised classification algorithm known as K-Nearest neighbor (KNN) is used. This algorithm is more suited as it is able to work at its best in a minimal period of time for a variable number of parameters. Firstly, a random value is assumed and assigned to a variable K. A Euclidean distance formula is applied to calculate the shortest distance between the new record and all the old records. All the shortest distances obtained are sorted in ascending order and the first K number of records are considered and amongst them, the highest count of the categorical variable is determined and assigned to the new record. The implementation of this KNN algorithm has been self-coded and not with help of inbuilt libraries.

Records	P1	P2	P3	Result
2	3	5	2	-
5	4	4	2	-
7	2	3	4	+
8	4	3	3	

Fig. 1 . KNN Classification

IV. IMPLEMENTATION

Our work offers a browser-based application where the students of an organization can take up the survey, input the parameters which they seem to be troubled about and obtain the percentage of stress as a result. Each student upon the correct verification of identity details can have access to his or her portal, where the survey can be taken up. Based on the result percentage and the probabilistic parameters the appropriate solution will be given by the admin to each student.

Firstly, the dataset has been collected, analyzed and pre- processed. The data which has been gathered contains various parameters such as Gender, Financial Issues, Family Issues, Health Issues, Partiality Fix, Pressure, Regular, Interaction. These parameters are encoded into numerical values which are then used for training of the model. The stress prediction is done with the help of a supervised learning machine learning approach known as “KNN classification Algorithm”. This algorithm is known for portraying accurate results.

The website has two major users namely Admin and Students where Students will login and input the parameters which causes them stress. The percentage of stress has been classified into four categories encoded as numbers which are ‘0’ indicating stress free, ‘1’ indicating between 1 to 25% of stress, ‘2’ indicating between 25 to 50% of stress, ‘3’ indicating between 50 to 100% of stress. Based on the percentage and the parameters which caused them stress the solution will be given by the admin.

The functionalities of the administrator include:

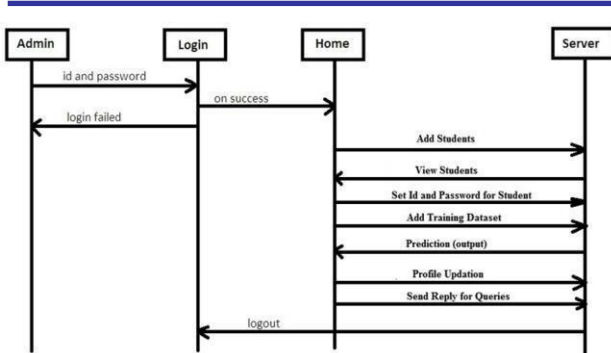


Fig. 2. Sequence Diagram of Administrator

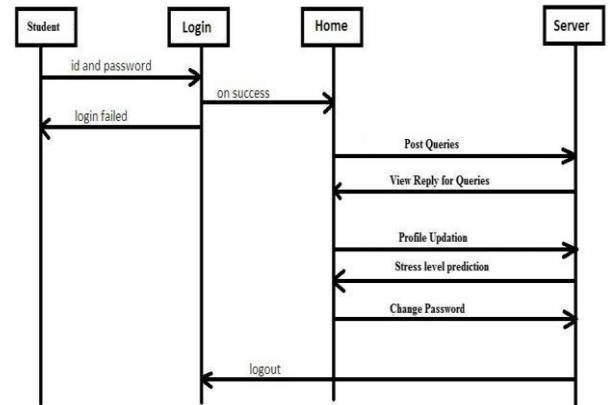


Fig. 3. Sequence Diagram of Student

- Login Module – The admin gets access to the admin portal by the input of his/her Login Credentials.
- Add Students – The Admin can record the details of every student of various departments. Upon addition of each student record, the automatic email will be sent to the respect individual with their login Credentials.
- View Students –The records of each student will be displayed and with the availability of edit or delete option, the admin can either update or delete the record of particular student.
- Prediction Module –This is the core module where the percentage of stress is detected in bulk of the testing dataset with the help of a Machine Learning technique called as K- Nearest Neighbor’s (KNN). The testing dataset has been imported to this module from the excel sheet. The various stress percentage of each testing dataset record is Classified as Stress-free and Stressful and further classified into the percentage of stress and is displayed with denominations as mentioned above. The Percentage of stress is also graphically Visualized.
- Profile Updation – Admin can change his/her password for security reasons.
- Queries – The queries posted by the particular student is stacked under the pending queries section in the admin's portal. The admin replies to each query. The queries replied by him/her will be displayed under answered queries section.
- Solution- The admin delivers each student the appropriate solution after analyzing the reasons of stress for the student and his or her stress category.

The functionalities of the student include:

- Login module - Students can login to the website using their credentials given by the admin.
- Stress prediction Module - Various parameters are enlisted and is visible to the students. Students can input these parameters to predict their stress level prediction. The obtained result is categorized into stressful and stress free and the stressful result is further classified into the percentage of stress and also represented graphically. Each parameter can be described as:

Sl.No	Factors	Sub Factors	Encoded Values
1	Gender		• 1-Male • 2-Female
2	Financial Issues	•Repay loan issues •Deadline of fee payment •Payment For Paying Guest(PG) or Hostel	• 0-No • 1-Yes
3	Family Issues	•Parental Expectations •Being bullied by siblings •Divorce of Parents •Negligence of Children •Poor Communication and misunderstandings	• 0-No • 1-Yes
4	Study Hours		• 1 • 2 • 3.....etc
5	Teaching Method		• 0-Fair • 1-Not Good
6	Health Issues	•Malnutrition •Sinus or Migraine •Insomnia(Sleep deprivation)	• 0-No • 1-Yes
7	Partiality Fix		• 0-No • 1-Yes
8	Exam Schedule		• 1-Monthly • 2-Half • 3-yearly • 4-Annual • 5-Slip Test
9	Friends Issues	•Conflicts •Comparison between them •Jealousy •Mistrust •Betrayal	• 0-No • 1-Yes
10	Pressure		• 0-No • 1-Yes
11	Regular	1.Hostler •Home Sick •Lack of nutritious food •Lack of Resources 2.Localite •Time management •Transportation Problem	• 0-No • 1-Yes
12	Interaction		• 1-Excellent • 2-good • 3-Average • 4-Poor

Fig. 4. Parameters List

- Solution module- Students find the solution proposed by the admin and can adopt this for reduction of mental stress.
- Post queries- Students can post their queries and grief. The queries which are yet to be replied by the admin are stored in Pending Section, and the queries which are answered are stored in the Answered section.
- Update Profile-Student can change his/her password.

The below figure shows that the student is classified as stress free.

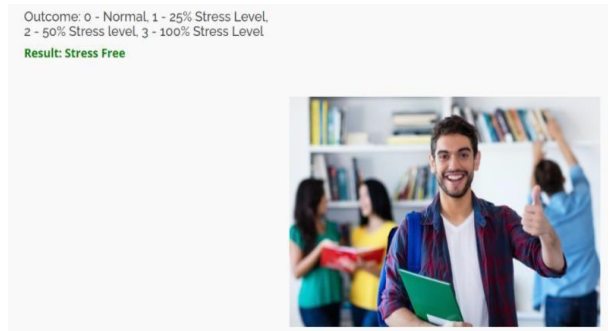


Fig. 5. Stress Free

The below figure shows that the student is classified as stress full and range of stress level is up to 100%.

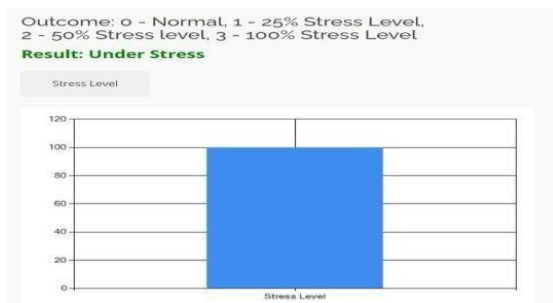


Fig. 6. Stress Full

V. RESULT AND ANALYSIS

The dataset consisting of different student records has been collected. The data has been encoded as integers under each of the twelve labels as shown in the below figure. The last column portrays the result label.

	Gender	Financial_Issues	Family_Issues	Study_Hours	Teaching_Method	Health_Issues	Partiality	FixExam_Schedule	Friends_Issue	Pressure	RegularInteraction	Result
1	0	0	5	2	1	0	1	0	0	1	4	0
1	1	0	4	1	1	1	2	1	0	1	3	2
2	0	0	3	1	1	0	2	0	0	2	2	1
2	1	0	5	2	2	0	2	0	1	1	4	1
1	0	0	2	2	1	0	3	0	1	2	1	3
2	1	0	4	1	1	0	2	0	1	1	2	0

Fig. 7. Training dataset

MODEL ACCURACY

Constraint	Algorithm
Accuracy	94.5054945054945%
Time (milli secs)	1789
Correctly Classified	94.5054945054945%
InCorrectly Classified	5.4945054945055%

Fig. 8. KNN Efficiency

The above table shows the performance of KNN algorithm. The testing and the training data is constituted as 10 percent and 90 percent of the original dataset. KNN finds the accuracy by matching the results of the predicted outputs of the testing dataset with the original dataset. The more the number of records correctly predicted upon the matching, the better accuracy of the model.

Formula:

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

The below figure is the graphical visualization of the stress percentages tabulated for the testing dataset.

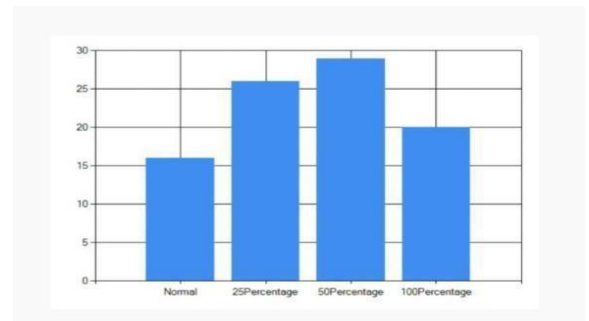


Fig. 9. Graphical Analysis of Tested Data

The tabulations are such that 15 students out of 91 test records have normal or nearly no stress. 26 students are identified with the range of 1% to 25% of stress, 29 students are found having the range from 25 % to 50% of stress, 20 students have 50% and above stress levels.

VI. CONCLUSION AND FUTURE SCOPE

Students are the victim of massive amounts of stress and this is increasing at an alarming rate with the ever growing competition on the educational grounds. This surfaces other problems such as peer pressure, parental expectations, health issues and many more. The tender minds of the students would

be subjected to these parameters and they won't be able to cope up well without the right guidance. Our work is an aid which helps in predicting the stress levels due to various parameters considered through a survey and it also lists as to whether the students are stress free or stressful and the range of their stress levels. This is accurately done with the KNN classification algorithm. This Machine Learning model holds an accuracy of 94%. Not just identifying the parameters and measuring the stress levels, our work tops this up by giving each student the appropriate solution for his or her grief. The student can incorporate the solution and work towards maintaining his or her mental equilibrium.

The future work would be to replicate this project in various universities for the beneficiary of the students and their mental well-being. The solution module could be further integrated with a Chatbot. A much more vast and detailed dataset can be collected and worked with other Machine learning techniques. The parameters for stress can further be classified into numerous sub parameters.

## VII. REFERENCES

- [1] American College Health Association (2018). ACHA /NCHA II Fall 2019 Reference Group DataReport. Hanover, MD: ACHA.
- [2] U Srinivasulu Reddy, Aditya vivek Thota, Adharun," Machine learning Techniques for Stress prediction in working employee" 2018 IEEE International Conference on computational intelligence and computing research.
- [3] Norizam, Sulaiman. Determination and classification of human stress index using nonparametric analysis of EEG signals. Diss. Universiti Teknologi MARA, 2015
- [4] Xu, Q., Nwe, T.L., Guan, C.. Cluster-based analysis for personalized stress evaluation using physiological signals.
- [5] IEEE Journal of biomedical and health informatics 2015; 19(1):275-281.
- [6] S.K. Yadav, Arshad Hashmi," An Investigation of Occupational stress Classification by using Machine Learning Techniques", Vol.-6, Issue-6, 2018.
- [7] D.Filip & C. Jesus. (2015). A Neural Network Based Model for Predicting Psychological Conditions International Conference on Brain Informatics and Health 252-261.
- [8] S. G. Alonso, I. Torre-Díez, S. Hamrioui, M.I López-Coronado, D. C. Barreno, L. M. Nozaleda, and M. Franco. Data Mining Algorithms and Techniques in Mental Health: A Systematic Review. J. Med. Syst. Vol. 42,9(September2018),1-15