International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, June 2022
Impact Factor: 6.252

AI Based Stress Detection using Machine Learning Techniques

Amrutha R J[1], Bindu M S[2], Saritha K[3]
PG Student, Computer Science, School of Technology and Applied Sciences, Kottayam, India [1]
Principal, Computer Science, School of Technology and Applied Sciences, Kottayam, India [2]
Assistant Professor, Computer Science, School of Technology and Applied Sciences, Kottayam, India [3]

Abstract:*The main objective of this study is to develop an AI based system that will predict the stress level of each student. First step is to collect the data needed to train the system. A total of 60 questions were prepared with the help of an expert. Each student will have to answer all the questions. Each question has four options never, somewhat frequent, frequent, always. Questions were provided to the students using google form. Then based on the information, we manually determine the stress level. Every answer has a score associated with it. Then total score is calculated and on the basis of the calculated score stress is determined as mild stress, moderate stress and severe stress. If the total score is greater than 180, then stress level is Severe. If the total score is greater than or equal to 90 and less than or equal to 180, then stress level is moderate. If the total score is less than 90, then stress level is mild. We append these stress level into the dataset. Then we will split the dataset into training and testing dataset. Then three classifiers were trained using the training data and the models are evaluated and best classifier is chosen for implementation.*

Keywords: Stress, Random Forest, Support Vector Machine, Naïve Bayes, etc.

I. INTRODUCTION

Data mining is the process of extracting information from a large amount of data. This process can be applied in various fields like medical field, real estate etc. Now a days, need of data mining to conduct research on educational field is on high demand [1]. Similarly scope of machine learning in educational research is also unlimited. [2]

Machine learning plays an important role in detecting many mental conditions. One such mental condition is stress. School and college students are suffering from stress due to many factors like academic problems, family problem, health issues and other socio-economic factors. Early detection of this stress is very important in this era. Many machine learning techniques like supervised and unsupervised algorithms can be used for this.

This paper is organised as follows. Section 2 deals with literature review. Section 3 describes the methodologies used. Section 4 deals explain details about the dataset used. Section 5 performs a detailed analysis of the methodologies used and section 6 explains conclusion and future enhancement.

II. LITERATURE REVIEW

Persulessy, G. B. V et. al [3] proposed a web-based model for predicting stress level using PHP. This model has 5 stages, knowledge acquisition, representation, inference engine, design and implementation. They represented the knowledge using if-then rules and used forward chaining inference method to predict the result. In this model a total of 31 questions were used and stress level is calculated based on the score obtained. If score is greater than 300, stress level is high. If score is in between 150 and 300 or equal to 150 and 300, stress level is medium and if score is less than 150, stress level is low.

Padmaja B et.al [4] proposed a model to detect the stress using wireless tracker that will track physical activities. This tracker was developed by FITBIT. In this model four conditions, sleep patterns, number of working hours, physical activity and change in heart rate are evaluated using logistic regression and then a combined model is built. A total of 10 IT workers were participated in this study for a duration of 10 months. Their age is in between 28 and 45.

Archana V R et.al [5] proposed a model to detect and analyse the stress using six attributes, electrocardiogram, electromyogram, galvanic skin response hand and foot, heart rate and respiration. Three algorithms were used. They are decision tree, Naïve bayes, K-nearest neighbour.

Ahuja R et.al [6] proposed a system to detect mental stress among university students. The study was conducted among 206 students of Jaypee Institute of Information Technology. They have used Random Forest, Naïve bayes, Support vector machine and K-nearest neighbour classification algorithms. They have used sensitivity, specificity and accuracy to measure the performance of each classification algorithms. Among these algorithms, SVM algorithm has an accuracy of 85.71, Random Forest has an accuracy of 83.33, Naïve bayes has an accuracy of 71.42 and K-nearest algorithm has an accuracy of 55.55.

III. METHODOLOGY

The main objective of this study is to predict the stress level among students using machine learning. Three methodologies were used. The algorithm, Algorithm 1 used in this study is shown in Figure 1.

```
Algorithm 1:
Input:
    Input the data as DF
Output:
    Proposed Classification Model
Method:
    1. Preprocess the data set for handling missing values
    2. Encode the categorical data
    3. For each row do:
        a) Calculate the sum
        b) If sum > 90:
              Append stress as Mild
           else if (sum>=90 and sum<=180):
              Append stress as Moderate
           else(sum>180):
              Append stress as Severe
        c) Encode the stress level
    4. Split the dataset into training & testing data
    5. a) Build the Random Forest classifier
       b) Build the Support vector machine classifier
       c) Build the Naïve bayes classifier
       d) Evaluate the models and choose the best one
```

Figure1: Algorithm

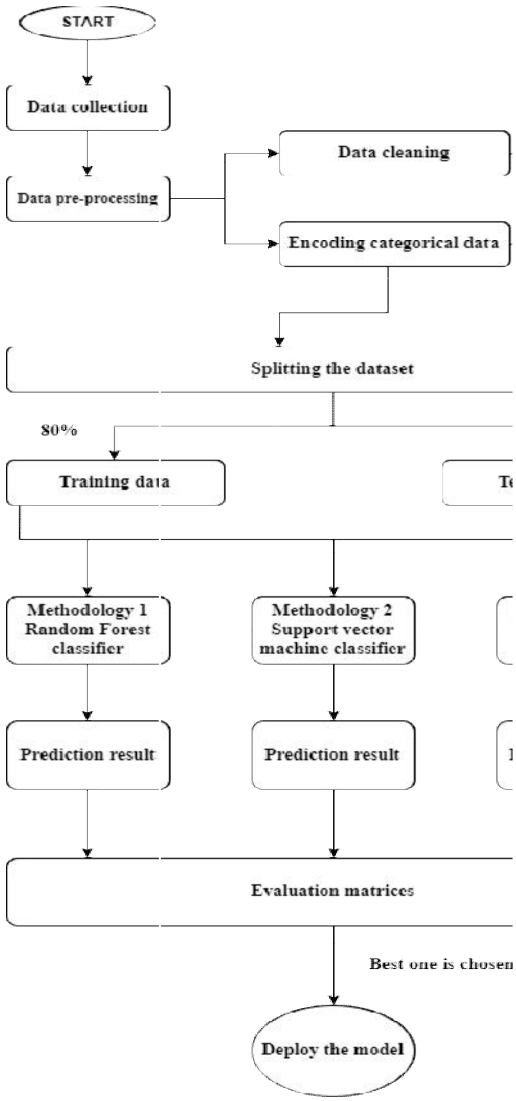
The process flow diagram is shown in Figure 2.

A. Data Collection and Data Pre-processing

The data needed for the study is collected from the students using a google form and the data is cleaned to remove the missing values if any. Categorical variables are also encoded into numbers.

1. *Handling Missing Values:*
   Several methods are available to handle missing values like deleting the rows, replacing the missing values with either mean, median, or mode, replacing the missing values with a global constant or we can fill the missing values with the most probable value.
2. *Encoding Categorical Data:*
   In this work we used 1,2,3,4 to encode the categorical data never, somewhat frequent, frequent, always.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, June 2022
Impact Factor: 6.252



B. Splitting the Dataset

Original dataset is divided into a training and testing dataset such that training data can be used to train the model and testing data can be used to evaluate the model. Dataset is divided in the ratio 80:20. The ratio is shown in Table 1.

Table 1: Train-test split ratio

DATASET	
Training Data	Testing Data
80%	20%

C. Building Classifier and Evaluation

Classifiers are built using Random Forest, Support vector machine and Naïve Bayes algorithm. Then accuracy, precision, recall, and F1 score are evaluated for these three classifiers and the best one is chosen.

1. **Confusion Matrix:**
   It is used to analyse the classifier. Confusion matrix for multi(three) class classification is shown in Table 2

Table 2: Confusion matrix for multiclass classification

	Predicted value			
		Class 1	**Class 2**	**Class 3**
Actual Value	Class 1	+ve$_{(cell\ 1)}$	-ve $_{(cell\ 2)}$	-ve $_{(cell\ 3)}$
	Class 2	-ve $_{(cell\ 4)}$	+ve $_{(cell\ 5)}$	-ve $_{(cell\ 6)}$
	Class 3	-ve $_{(cell\ 7)}$	-ve $_{(cell\ 8)}$	+ve $_{(cell\ 9)}$

Equation to calculate TP, FP, TN, FN is shown in Table 3

Table 3: Equation for calculating tp, fp, tn, fn

	Class 1	**Class 2**	**Class 3**
TP	cell 1	cell 2	cell 3
FP	cell 4 + cell 7	cell 2 + cell 8	cell 3 + cell 6
TN	cell 5 + cell 6 + cell 8 + cell 9	cell 1 + cell 3 + cell 7 + cell 9	cell 1 + cell 2 + cell 4 + cell 5
FN	cell 2 + cell 3	cell 4 + cell 6	cell 7 + cell 8

2. **Accuracy:**
   Accuracy is calculated using equation 1

$$accuracy = \frac{TP + TN}{Total\ number\ of\ classification} \quad (1)$$

3. **Precision:**
   Precision is calculated using equation 2

$$precision = \frac{TP}{TP + FP} \quad (2)$$

4. **Recall:**
   Recall is calculated using equation 3

$$recall = \frac{TP}{TP + FN} \quad (3)$$

5. **F measure:**
   F1 score is calculated using equation 4

$$F1\ score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (4)$$

IV. EMPIRICAL ANALYSIS

The dataset used in this study is created by using a google form containing 60 questions. Each question has four options as answer. Student has to choose one option for each question. Questions are shown in Fig. 3. and Fig. 4.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 3, June 2022

Impact Factor: 6.252

1. Have headache?
2. Have fatigue/constant tiredness?
3. Have difficulty in sleeping?
4. Have difficulty in concentrating?
5. Have upset stomach/nausea?
6. Have irritability?
7. Have chest pain/rapid heartbeat?
8. Have backpain?
9. Have difficulty in breathing?
10. Have Excessive worry?
11. Have sweating/sweaty hands?
12. Have frequent cold/flu/fever?
13. Have drastic weight loss?
14. Have memory problem?
15. Have nightmares?
16. Do you find it difficult to meet your high parent's expectation?
17. Do you think that your parents treat you as a helpless person?
18. Do you feel guilty if you fail to fulfil your parents hope?
19. Do you think that your parents are putting pressure on you to meet your future goals/for your success?
20. Do you find it difficult to get along with groupmates in doing academic task?
21. Do you think that your friends did not care about you?
22. Do you feel disturbed when having problem with your boyfriend/girlfriend?
23. Do you think that your family is not supportive?
24. Do you think that your lectures/teachers are not supportive?
25. Do you feel frustrated by the lack of faculty management?
26. Do you feel hurt or upset when negative feedback is given to you?

Figure 3: Question 1-26

27. Do you keep your worries a secret from other's life?
28. Do you feel frustrated of inadequate campus facilities?
29. Do you feel like withdrawing from family, friends and isolating yourself?
30. Do you have a financial problem because of the expenses of the university?
31. Do you find it difficult to juggle time between study and social activity?
32. Do you feel nervous delivering the class presentation?
33. Do you feel stressed to sit for examination?
34. Do you find it difficult to juggle time between study and society involvement?
35. Do you loss interest towards courses?
36. Do you feel burden of academic workloads?
37. Do you feel stressed dealing with difficult subject?
38. Do you feel difficult in handling your academic problem?
39. Do you experience acute fear while you speak or do something in front of others?
40. Do you get always fear when you express your opinions in front of others?
41. Do you feel that you are constantly observed and judged by other people?
42. Have you ever experienced the fear of failure and embarrassment/humiliation drawing you back from the things you desired to do?
43. Does your distressing fear prevent you from speaking to other people?
44. Do you experience acute fear when you meet new people or strangers?
45. Have you been experiencing physical problems like vomiting, shyness, shivering, sweating etc. during or before facing social situations?
46. Before or during social situations has your social anxiety brought in a panic attack?
47. Do you try to avoid the social situations? that cause your anxiety, if at all possible?
48. Do you sometimes use drugs or alcohol to make feared social situations easier to handle?
49. Has your anxiety about the social situations interfered with your normal routine?
50. Has your anxiety about the social situations interfered with your functioning at work or school?
51. Has your anxiety about the social situation kept you from having friendship, relationship or any kind of social life?
52. Do you feel nervous or worried about stressful situations?
53. Do you get angry easily?
54. Does stress make it hard for you to think or act?
55. Have you often being upset because of something that happened unexpectedly?
56. Do you feel stressed with bad living condition of hostel?
57. Does surrounding noise distract you?
58. Does messy living conditions distract you?
59. Does crowding makes you uneasy?
60. Do you feel scared being at the insecure place?

Figure 4: Question 27-60

These questions are prepared with the help of an expert and answers were collected from students of various schools and colleges. The options are never, somewhat frequent, frequent, always. Each option is assigned a score as given in Table 4.

Table 4: Score of each option

Options	Score
Never	1
Somewhat frequent	2
Frequent	3
Always	4

The dataset is pre-processed and we find the stress level by performing a summation of the score of answer in each observation. If score is greater than 180, the stress level is severe. If score is in between 90 and 180, stress level is moderate and if score is below 90, stress level is mild. Then we append the stress level in to the dataset. Thus, the pre-processed dataset used to implement the above methodologies contains 61 attributes. First 60 attributes are answers to the question which is of the type nominal and the last feature indicate the stress level mild, moderate and severe. It is shown in Table 5.

Table 5: Attributes of dataset

Attribute	Value	Type
Question 1 . . Question 60	Never - 1 Somewhat frequent - 2 Frequent - 3 Always - 4	Nominal
Stress Level	Mild Moderate Severe	Nominal

V. RESULT AND DISCUSSION

Initially dataset had 600 observations of 60 variables and after pre-processing and stress encoding, the dataset had 600 observations of 61 variables. The dataset had 201 observations for stress level mild, 367 observations for stress level moderate and 32 observations for stress level severe. Training dataset had 480 observations of 61 variables and testing dataset had 120 observations of 61 variables. Number of observations in the original as well as training and testing dataset for each stress level is shown in Table 6.

Table 6: Number of observations

Stress Level	Training Data	Test Data	Original Dataset
Mild	166	35	201
Moderate	288	79	367
Severe	26	6	32
Total	480	120	600

Confusion matrix for Random Forest classifier on training data is shown in Table 7.

Table 7: Confusion matrix for random forest classifier on training data

		Predicted Value		
		Mild	Moderate	Severe
Actual Value	Mild	166	0	0
	Moderate	0	288	0
	Severe	0	0	26

Confusion matrix for Random Forest classifier on testing data is shown in Table 8.

Table 8: Confusion matrix for random forest classifier on testing data

		Predicted value		
		Mild	**Moderate**	**Severe**
Actual Value	Mild	32	3	0
	Moderate	3	76	0
	Severe	0	0	6

Confusion matrix for Support vector machine classifier on training data is shown in Table 9.

Table 9: Confusion matrix for svm classifier on training data

		Predicted value		
Actual Value		Mild	Moderate	Severe
	Mild	288	0	0
	Moderate	0	166	0
	Severe	0	0	26

Confusion matrix for Support vector machine classifier on testing data is shown in Table 10.

Table 10: Confusion matrix for SVM classifier on testing data

	Predicted value			
Actual Value		Mild	Moderate	Severe
	Mild	34	1	0
	Moderate	1	78	0
	Severe	0	0	6

Confusion matrix for Naïve bayes classifier on training data is shown in Table 11.

TABLE 11: Confusion matrix for naïve bayes classifier on training data

	Predicted value			
		Mild	Moderate	Severe
Actual Value	Mild	159	7	0
	Moderate	18	270	0
	Severe	0	0	26

Confusion matrix for Naïve bayes classifier on training data is shown in Table 12.

TABLE 12: Confusion matrix for naïve bayes classifier on testing data

	Predicted value			
		Mild	Moderate	Severe
Actual Value	Mild	32	3	0
	Moderate	2	77	0
	Severe	0	0	6

Accuracy measure for different classifiers on training and testing data is shown in Table 13.

TABLE 13: Accuracy measure

Classifier	Accuracy	
	Training data	Testing data
Random Forest classifier	100%	95%
Support vector machine classifier	100%	98%
Naïve bayes classifier	95%	96%

Fig. 5. shows the pictorial representation of accuracy

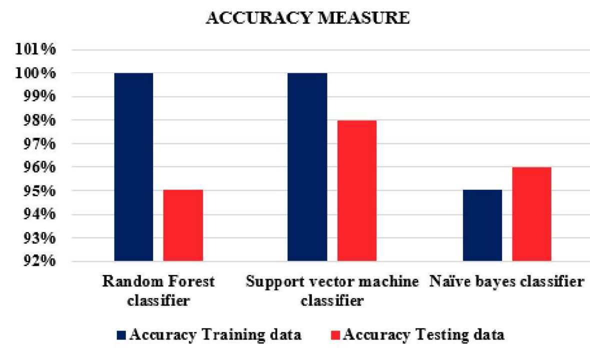


Fig. 5 Accuracy measure

Precision measure for different classifiers on training and testing data is shown in Table 14.

TABLE 14: precision measure on training and testing data

Classifier	Precision					
	Training Data			Testing Data		
	Mild	Moderate	Severe	Mild	Moderate	Severe
Random Forest classifier	1.00	1.00	1.00	0.91	0.96	1.00
Support vector machine classifier	1.00	1.00	1.00	0.97	0.99	1.00
Naïve bayes classifier	0.90	0.97	1.00	0.94	0.96	1.00

Recall measure for different classifiers on training and testing data is shown in Table 15.

TABLE 15: Recall measure on training and testing data

Classifier	Recall					
	Training Data			Testing Data		
	Mild	Moderate	Severe	Mild	Moderate	Severe
Random Forest classifier	1.00	1.00	1.00	0.91	0.96	1.00
Support vector machine classifier	1.00	1.00	1.00	0.97	0.99	1.00
Naïve bayes classifier	0.96	0.94	1.00	0.91	0.97	1.00

F1-score measure for different classifiers on training and testing data is shown in Table 16.

TABLE 16: F1 score measure on training and testing data

Classifier	F1-Score					
	Training Data			Testing Data		
	Mild	Moderate	Severe	Mild	Moderate	Severe
Random Forest classifier	1.00	1.00	1.00	0.91	0.96	1.00
Support vector machine classifier	1.00	1.00	1.00	0.97	0.99	1.00
Naïve bayes classifier	0.93	0.96	1.00	0.93	0.97	1.00

Thus, Support vector machine classifier is chosen as the best classifier and the model is implemented.

VI. CONCLUSION

The main objective of this study was to develop an AI based system for detecting stress among school and college students. A set of 60 questions were used to detect the stress level of student. Three methodologies were used to build the model and different evaluation matrices were used to evaluate the models and the best one is chosen. This system can be used by teachers to assess the stress level of students. The proposed work can be extended with deep learning techniques.

REFERENCES

[1] Baker, R. S. J. D. (2010). Data mining for education. International encyclopedia of education, 7(3), 112-118.

[2] Kučak, D., Juričić, V., &Đambić, G. (2018). MACHINE LEARNING IN EDUCATION-A SURVEY OF CURRENT RESEARCH TRENDS. Annals of DAAAM & Proceedings, 29.

[3] Persulessy, G. B. V., Pratama, N. S., Setiawan, N., &Sevani, N. (2019). Web-Based Expert System to Detect Stress on College Students. ComTech: Computer, Mathematics and Engineering Applications, 10(1), 9-14.

[4] Padmaja, B., Prasad, V. R., & Sunitha, K. V. N. (2018). A machine learning approach for stress detection using a wireless physical activity tracker. International Journal of Machine Learning and Computing, 8(1), 33-38.

[5] Archana, V. R., &Devaraju, B. M. (2020). Stress Detection Using Machine Learning Algorithms. International Journal of Research in Engineering, Science and Management, 3(8), 251-256.

[6] Ahuja, R., & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. Procedia Computer Science, 152, 349-353.