Journal of Information Systems Engineering and Management
2025, 10 (30s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Performance Analysis of Machine Learning Algorithms for Classification of Stress using PSS-10 scale of University Postgraduate Students

Pavitra Gautam[1], Dr. Ruchika Malhotra[2],
[1]*Mtech candidate, Department of Software Engineering, Delhi Technological University, Delhi 110042, India*
[2]*Professor, Department of Software Engineering, Delhi Technological University, Delhi 110042, India*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 20 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Stress is a person's normal reaction to difficulties or circumstances. In brief spurts, it might be useful in encouraging someone to fulfill a deadline, prepare for an event, or respond to danger. However, chronic or unmanaged stress can negatively affect the individual physical and mental health. Students are not untouched by this. Every year, thousands of students commit suicide due to stress. In this paper we examine the performance of machine learning algorithms which help in early prediction of stress using the best prediction model. The dataset was taken from postgraduate (Master of Technology) students using Google form , it consisted of 57 students data. We have applied 6 types of classification algorithms: Logistic (75.00%), KNN (83.33%), SVM (66.67%), the Decision Tree (92.00%), Random Forest (92.00%)and Naive Bayes' (91.67%) and also we calculated their accuracy with the help of confusion matrix. In this study, Decision Tree algorithm and Random Forest algorithm shared an equal and highest accuracy of 92.00% as compared to other algorithms.

Keywords: Stress Prediction, DTU, PSS-10, Machine learning algorithm, student, dataset, SVM, RF,, Decision Tress, Random Forest , Naïve Bayes, Logistic, Stress, Postgraduate Students, Classification Algorithm, Supervised learning .

INTRODUCTION

Student life is often regarded as a golden period in one's life, filled with opportunities for learning, personal growth, and preparation for the future. However, beneath the surface of excitement and academic challenges lies a less-discussed but critical issue that is stress. Student stress has become a global phenomenon, affecting individuals at various educational levels, from primary school to higher education. It arises from various sources, including academic pressures, social expectations, career expectations, financial constraints, and personal challenges and many more. In addition to depression, schizophrenia, attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and other conditions, an estimated 450 million people worldwide suffer from mental fitness issues. [1]. There are many consequences of unmanaged stress, it is highly affect academic achievement of students, physical health, mental health, social relationships etc. Understanding student stress and making strategies for management, is essential for educators, policymakers, and society as a whole to ensure the well-being of future generation. With the introduction of machine learning (ML) and big data analytics, the ability to conduct a more holistic analysis is increasing [2]. According to the World Health Organization (WHO), depression is the most common mental health issues, impacting more than 300 million individuals globally. [3]. Poor stress management can lead to severe injuries that can occasionally have a total impact on education and even seriously harm students' fitness at different phases [4].Stress specially in students is a multi-faceted phenomenon influenced by various internal and external factors, such as rigorous study schedules, competition, parental expectations, and the transition to new environments like college or university. Occasional stress can be beneficial for performance while chronic stress often leads to burnout, anxiety, depression, and other health issues. For students, these challenges can manifest in difficulty concentrating, poor academic performance, strained relationships, and even long-term mental health concerns.

Numerous studies have been undertaken regarding stress, particularly among college students. ***Cheng Ding, Yuhao Zhang, and Ting Ding [5***] in their research authors introduced a hybrid model that integrates two approaches: a random forest model and a gradient boosting machine (GBM), achieving an accuracy rate of 100%. They carried out a 10-fold cross validation and statistical T-test with suggested model to demonstrate the importance of the suggested method over other techniques, This paper showed that HB model performs better with a mean accuracy of 1 and a standard deviation of +/- 0 .

Disha Sharma, Sumit Chaudhary [6] , stress is the main cause that have high impact on mental and physical health of students. In this research they have applied number of classification techniques as Naive Baye's, Logistic Regression, Multilayer perceptron technologies for the prediction of stress in professional students. They have used Weka tool, the accuracy measures of various techniques are calculated and compared in this study. Authors investigate 220 undergraduate and postgraduate understudies and observed that Baye's Net classifier gives the longest accuracy of 88 % by using Kappa, statistic F-measure, MCC, mean absolute error, ROC area, false positive, true positive etc.

Garima Verma*, Sandhya Adhikari, [7] ,the objective of the paper was finding the variables that influence mental health conditions including anxiety, stress, and depression in college students—particularly in engineering school. The size of the dataset was 513 individuals enrolled in graduation-level engineering programs in northern India. Data was gathered using both online and offline surveys. The machine learning models used in this model are Logistic and second Support Vector Machine (SVM) resulting accuracy of 67% and 86.84% respectively.

Ishrak Jahan Ratul a, Mirza Muntasir Nishat [8] , This study was to develop a reliable machine learning-based prediction model aimed at forecasting perceived stress and to validate it against real-world actual data. To achieve feature reduction, the authors employed the chi-squared test along with Principal Component Analysis (PCA). Two approaches one genetic algorithms (GA) and grid search cross-validation (GSCV) were used for important phase that is hyper parameter optimization (HPO) to control the learning process. The results showed that approximately 11.26% of people had significant levels of social stress. 24.10 % of participants had extremely high psychological stress. Additionally, the ML models' prediction results showed the most impressive recall value (0.826), F1 score (0.890), accuracy (80.5%), and precision (1.000).

***Joey Man Yee KWOK & Douglas Kei Shing[9]*,** This study analyzed 337 undergraduate students at the Open University of Hong Kong . Perceived Stress Scale-10 (PSS-10) along with The Beck Anxiety Inventory (BAI) scale was used as the measure of perceived stress level and evaluating the convergent validity of PSS-10 respectively. The General Self-Efficacy (GSE) Scale and the Subjective Happiness Scale (SHS) were employed to assess the associated divergent validity. The primary results indicated that, the stress level of the participants in age group of 18-29 had an average score of 19.02 which was considered to be higher than the standard score (M = 14.2; SD =6.2), and thus undergraduate students who belonged to this age group were found to present a potential higher stress level among those participants.

Ms. Ancy Paul1, Ms. Resija P R2 et all[10], The objective of the study was to detect the different level of stress among students of Vimla College. Data from 954 students was gathered online. Nine algorithms were trained across three categories: a) chronic, b) episodic, and c) acute, achieving satisfactory accuracy levels. Among the nine algorithms, two models demonstrated superior accuracy: the RF classifier and logistic regression. Subsequently, the random forest classifier was used during the testing phase of the model development, effectively identifying the level of stress and categorizing, achieving an accuracy rate of 99% for both the classifers.

Prakruthi Manjunath, Twinkle S et all [11], A solution was suggested for the educational institution that enables authorities to monitor the anticipated stress levels of each enrolled student. The input for this process consists of survey data, which is utilized by a pre-trained machine learning model to estimate the stress percentage for every student. The model performs a two-tier classification of stress levels, determining whether a student is stress-free or experiencing stress. Additionally, among those identified as stressed, a further categorization is made regarding the severity of their stress, classifying it as low, medium, or high. The underlying framework of the ML model is based on the KNN classification algorithm with the accuracy of 94.50%.

Ravinder Ahuja, Alisha Banga [12] , In this study, the authors assessed the mental stress experienced by students one week prior to examinations and during their internet usage. The aim was to evaluate stress levels among college students at various stages of their academic journey. Data was collected from 206 students at the Jaypee Institute of Information Technology. Four machine learning classification algorithms were applied with resulting accuracy as Random forest (83.33 %), Naïve Bayes (71.42 %), SVM (85.71%) and KNN (55.55%). Utilizing 10-fold cross-validation led to a notable enhancement in the accuracy and performance of the data. The analysis revealed that the support vector machine exhibited the highest performance among the four algorithms, attaining an accuracy of 85.71%.

Reshma Radheshamjee Baheti, Supriya Kinariwala [13], In this research, the author introduced a system framework designed to identify users' psychological stress levels through the analysis of weekly social media data. A word dictionary was employed, assigning ratings from -5 to +5 for each term. To classify and predict the data, the study utilized Support Vector Machine (SVM) and Naive Bayes (NB) algorithms. Additionally, two models, n-gram and Skip-gram, were implemented to enhance the accuracy of the results, incorporating Word Sense Disambiguation. Support Vector Machine with WSD and Ngram gives 65% precision and 67% recall.

MATERIALS AND METHODS

Dataset and preprocessing

We have used statistician random sampling for data collection. The data collection process yielded information from 57 students total of 106 final year students pursuing M-Tech degrees in Computer Science, Information Technology, Software Engineering, and Data Science, conducted via a Google Form. The questioner designed for stress prediction is having 14 questions out of which 10 questions are of PSS-10 scale and four questions focuses on kindness for exploring the perceived kindness with stress in our study. The dataset consists of three targeted classes low, moderate and high in alignment with the prescribed standard defined by well-established and validated PSS-10 scale. We have used label encoding for converting the categorical data values to the numerical data . The dataset further divided into training set and testing set in the ration 80:20 ration respectively for Machine Learning model training and evaluation. Accuracy, precision, recall and confusion matrix are used to evaluate the performance of a classification model.

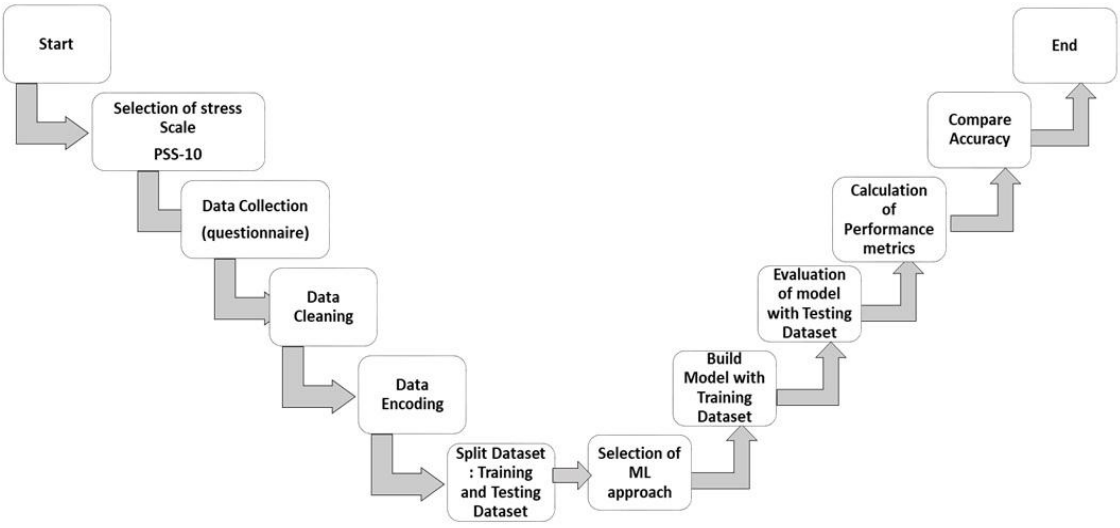

Figure 1: Methodology

The Perceived Stress Scale-10 (PSS-10)

The Perceived Stress Scale-10 (PSS-10) is a widely used psychological assessment tool designed to measure perceived stress in individuals. Formulated by Sheldon Cohen and colleagues. It consists of 10 items that are rated on a 5-point Likert scale, ranging from 0 ("never") to 4 ("very often"). The PSS-10 is valuable because it measures perceived stress over the past month, reflecting a subjective appraisal of life circumstances rather than specific stressors or their

intensity. This subjective approach makes it particularly useful for comparing stress levels across diverse populations and contexts. The stress level is detected based on the total of PSSS score as below:

1. If the PSS score is between 0 to 13, it denotes that the stress level of students is low.

2. If the PSS score is between 14 to 26, it implies that the stress level of students is moderate.

3. If the PSS score is between 27-40, it indicates that the stress level of students is high.

Machine Learning Techniques.

Categorical data often requires specialized machine learning algorithms or preprocessing techniques to handle its discrete nature. Here are brief descriptions of a few key algorithms used in the study.

1. **Decision Trees** (DT): Decision trees work well with categorical data as they partition the data based on feature values at each node. They handle non-linear relationships effectively, can process mixed types of data, and are intuitive to interpret. Overfitting can occur, but this can be mitigated with pruning techniques or ensemble methods.

2. **Naive Bayes:** Naive Bayes is a probabilistic algorithm that assumes independence between features. It is particularly suitable for categorical data because it calculates the likelihood of each class given the feature values. Despite its simplicity, it often performs well in text classification and spam detection tasks.

3. **K-Nearest Neighbors** (KNN): KNN can work with categorical data by using distance metrics such as the Hamming distance. It classifies new instances based on the majority class of their nearest neighbors. While simple to implement, it can be computationally expensive for large datasets.

4. **Logistic Regression:** Logistic regression is applicable for the classification of categorical data in both binary and multiclass scenarios. Through the implementation of the sigmoid function, it forecasts the probabilities for the various classes. Categorical variables often need to be encoded using techniques like one-hot encoding or label encoding before being used.

5. **Random Forests (RF):** Random forests are an ensemble method based on decision trees. They are robust to overfitting and provide strong performance on categorical data. By averaging or voting across multiple trees, random forests reduce variance and improve predictions.

6. **Support Vector Machines (SVM):** Although primarily designed for continuous data, SVMs can handle categorical data through kernel tricks or appropriate encoding. This technique is particularly effective particularly in scenarios where the data cannot be separated linearly.

Performance measures.

There are many performance measures we can use to compare the performance of machine learning models. We have used following prime performance measurers which are calculated using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) predictions of confusion matrix.

Accuracy: It is the ratio of correct prediction (all positive and negative prediction) to the total predictions.

Formula used to calculate the accuracy is:

ACCURACY = (TP + TN)/ (TP + TN + FP + FN)

Precision: It is the ratio of correctly positive to all positive predictions. Formula used to calculate the precision is:

PRECISION = TP/ (TP + FP)

Recall: It represents the ratio of accurately identified positive predictions to the overall count of actual positive occurrences. The formula used to determine recall is:

RECALL = TP/ (TP + FN)

F1-score : It is defined as the harmonic mean of precision and recall. Formula used to calculate the F1score is:

F1 score=2[(Precision*Recall)/(Precision +Recall)]

RESULTS

In this study , jupyter notebook and python libraries (pandas, numpy, matplotlib, sklearn) are used for metric calculation model selection, data preprocessing , descriptive statistics of machine learning algorithms and data. Details of Dataset used in this study are:

Total Students=106 (M.Tech(DS,IT,SWE and CS))

Data collected=57

Stress Scale used=PSS-10

Data Collection Technique=Random Sampling

Data collection method=questionnaire

After preprocessing, mean, median, SD, MIN, MAX ,and interquartile range for all parameter of the data are calculated as shown in Table 1.

Table1: Descriptive statistics of the dataset

	Gender	Course	Age	Branch	Staying	Q1	Q2	Q4	Q5	Q6	Q8	Q10	Q11	Q12	Q14	pss_total	stress_class
count	57.000000	57.0	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000
mean	0.210526	1.0	1.421053	1.298246	1.105263	2.298246	2.087719	2.368421	1.333333	2.140351	2.140351	1.614035	2.105263	2.157895	1.877193	20.122807	1.070175
std	0.411306	0.0	0.822613	1.148989	0.816880	1.322402	1.138305	1.062873	0.969782	1.076346	1.125017	1.064934	0.976221	1.177004	1.134998	7.599507	0.622772
min	0.000000	1.0	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
25%	0.000000	1.0	1.000000	0.000000	0.000000	2.000000	1.000000	2.000000	1.000000	2.000000	2.000000	1.000000	2.000000	2.000000	1.000000	15.000000	1.000000
50%	0.000000	1.0	1.000000	1.000000	1.000000	2.000000	2.000000	2.000000	1.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	20.000000	1.000000
75%	0.000000	1.0	2.000000	2.000000	2.000000	4.000000	3.000000	3.000000	2.000000	3.000000	3.000000	2.000000	3.000000	3.000000	3.000000	26.000000	1.000000
max	1.000000	1.0	4.000000	3.000000	2.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	34.000000	2.000000

Table 1: Descriptive Statistics

As per the analysis maximum students are found at moderate level of stress, which is pre-stage to the high level of perceived stress. In a group of 57 students, 9 are identified as having low stress levels, 35 are assessed to be at a moderate level of stress, while 13 students are classified as experiencing high perceived stress levels as shown in Fig 2.

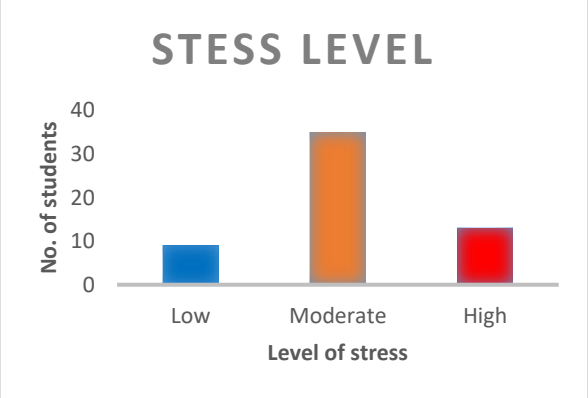

Figure 2:Stress Level

Table2: M Model Efficiency Comparison Table on student dataset

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.75	0.77	0.75	0.74
KNN Classifier	0.8333	0.87	0.83	0.80
Support Vector Machine Classifier	0.6667	0.68	0.67	0.65
Decision Tree Classifier	0.92	0.93	0.92	0.91
Random Forest	0.92	0.93	0.92	0.91
Naïve Bayes Classifier	0.9167	0.93	0.92	0.91

With the help of accuracy and model we have generated accuracy percentage graph shown in Fig 3.

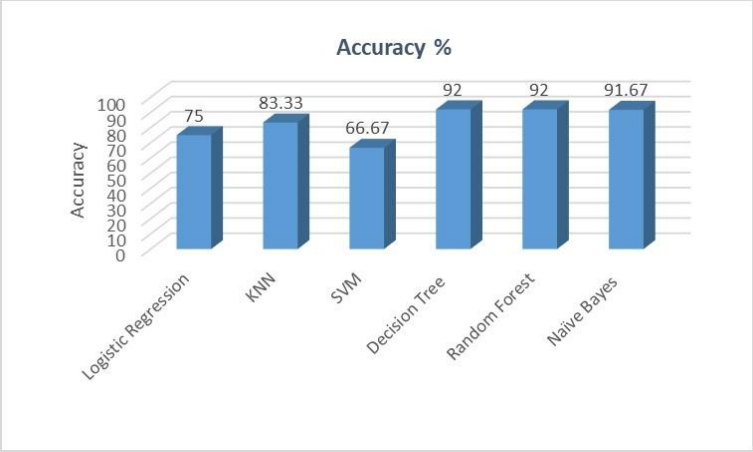

Figure 3: Accuarcy Percentage Bar Graph.

DISCUSSION

Stress is a term habitually used equivalently with negative valuable encounters or life events, this compromising individual wellbeing's [14]. Stress is identified as one of the leading health issues faced by individuals today. It is a primary contributor to numerous other health problems, highlighting the importance of its careful and effective management. In our research, we utilized six machine learning algorithms to assess their predictive accuracy: Logistic Regression (75.00%), K-Nearest Neighbors (83.33%), Support Vector Machine (66.67%), Decision Tree (92.00%), Random Forest (92.00%), and Gaussian Naive Bayes (91.67%). Among these, the Decision Tree and Random Forest algorithms demonstrated the highest levels of accuracy in predicting stress, outperforming the other classifiers. In future we can validate the PSS-10 stress model and conduct sequential modeling to gain deeper insights from data. Additionally, we have the option to integrate several additional variables while performing PCA factor analysis. There are various approaches that the university can implement to address stress management, ensuring that students remain free from stress, depression or anxiety and are able to deal with stress in a positive and efficient manner Periodic counselling and mindfulness practice sessions will be best to detect stress before it gets severe. The mentor assigned to the student should be in touch with the student and must have periodic sessions with them to understand their current state of mind to avoid unwanted consequences.

REFRENCES

[1] Salma S Shahapur and Praveen Chitti," Decoding Minds: Estimation of Stress Level in Students using Machine Learning", Indian Journal of Science and Technology ,vol-17,issue-19, May 2024

[2] Rocco de Filippis and Abdullah Al Foysal ," Comprehensive analysis of stress factors affecting students: a machine learning approach", Discover Artificial Intelligence,vol-4,September (2024)

[3] Anu Priyaa and Shruti Garga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms",Procedia Computer Science ,vol-167,April 2020

[4] Shreelakshmi CM and Samana K ,"Predictive analysis of student stress levels using ML." International Advanced Research Journal in Science, Engineering and Technology, Vol. 10, Issue 5, May 2023
[5] Ding C, Zhang Y, Ding T. A systematic hybrid machine learning approach for stress prediction. PeerJ Comput Sci. Feb 2023
[6] Disha Sharm and Sumit Chaudhary ," Stress Prediction of Professional Students using Machine Learning", International Journal of Engineering and Advanced Technology (IJEAT), Volume-9 Issue-5, June 2020
[7] Garima Verma and Sandhya Adhikari, "Machine Learning Model for Prediction of Stress Levels in Students of Technical Education", Conference Paper · August 2020
[8] Ishrak Jahan Ratul and Mirza Muntasir Nishat ," Analyzing Perceived Psychological and Social Stress of University Students: A Machine Learning Approach" ,Cellpress, Heliyon 9 (2023) .
[9] Joey Man Yee KWOK & Douglas Kei Shing," A Study of the Perceived Stress Level of University Students in Hong Kong", International Journal of Psychological Studies; Vol. 8, No. 4; 2016.
[10] Ms. Ancy Paul and Ms. P R Resija ," Stress Detection in College Students Using Machine Learning Algorithm", International Journal of Research Publication and Reviews, Vol 5, February 2024
[11] Prakruthi Manjunath, Twinkle S et all ," Predictive Analysis of Student Stress Level using Machine Learning", International Journal of Engineering Research & Technology (IJERT) , Volume 9, Issue 12, Special Issue – 2021.
[12] Ravinder Ahuja and Alisha Banga "Mental Stress Detection in University Students using Machine Learning Algorithms", Procedia Computer Science vol-152 ,pp-349–353 ,May 2019
[13] Reshma Radheshamjee Baheti and Supriya Kinariwala ," Detection and Analysis of Stress using Machine Learning Techniques", International Journal of Engineering and Advanced Technology (IJEAT) Volume-9 Issue-1, October 2019
[14] Dr M Anbarasi , Sethu Thakkilapati , Veeragandham Rajeev Vas , Sarabu Venkata Bharath Viswath, , "Predictive Analysis of Student Stress Level using Machine Learning", International Journal Of Engineering Research & Technology (IJERT) Volume 12, Issue 05 ,May 2023
[15] A. Kene and S. Thakare, "Mental Stress Level Prediction and Classification based on Machine Learning," Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, pp. 1-7,2021