Research Paper

The Role of Socioeconomic Factors in Machine Learning-Based Classification of Student Stress Levels

Tapu Biswas

University of Essex

Abstract

Stress among university students is a significant issue affecting their emotional, physical, and cognitive well-being. Stress can lead to anxiety, depression, and impaired academic performance. Early detection is important for mental health and academic success. This study used machine learning to predict stress levels in undergraduate students based on socioeconomic data from an online survey in Bangladesh and to identify if only the socioeconomic factor affects the student stress, using the PSS-10 stress scale. Results showed varying stress levels among students. Ten models were tested—including Naive Bayes, Random Forest, Decision Tree, CART, Logistic Regression, K-Nearest Neighbors, SVM, AdaBoost, and Bagging with the highest accuracy of 97.92% achieved by Decision Tree, CART, AdaBoost, and Bagging using all the factors. To examine the specific influence of socioeconomic factors, a secondary experiment was conducted using only socioeconomic attributes (family income, living status, fee provider, socioeconomic class). The SES-only model achieved lower accuracy (46%), indicating that socioeconomic conditions are not the only factors that influence stress among students.

Keywords: Student Stress, University, Machine Learning, Socioeconomic Factors, PSS-10, Stress Classification

Date of Submission: 26.08.2025 **Date of Acceptance:** 15.01.2026

AI Statement

I have used AI tools to review the text and enhance the clarity, grammar, and readability of the writing. All ideas, interpretations, and conclusions presented are my own.

Introduction

Stress is a reaction to pressure and can manifest in various ways, affecting both the physical and mental aspects of the human body (Yılmaz Koğar and Koğar, 2024). Stress is a regularly occurring emotion in all aspects of human life, including among students, that affects them academically,

DOI: https://doi.org/10.5526/esj.453

socially, physically, and emotionally (Di Mario et al., 2024). Stress is a normal feeling intended to assist a person in managing specific stressful circumstances. In small amounts, it can be productive; otherwise, excessive stress can lead to mental health problems such as depression and anxiety that affect a student's academic performance (Öztekin et al., 2025, Ribeiro et al., 2018). Socioeconomic status includes family income, parental education, parental occupation, and the social class of a family within society (Erola et al., 2016). Low socioeconomic status can affect students with cognitive ability, language, memory, and socioemotional processes, and hence, income and health in later years (Nja et al., 2022). Schools located in low socioeconomic status areas can be compared to lacking the necessary resources for students to progress academically, which can later affect students in their academic life (Meier-Faust and Watermann, 2024). Reasons like this can create stress among students regarding their studies at their undergraduate level. Female students are more likely to experience depression, anxiety, and stress compared to male students (Gao et al., 2020). A study found that undergraduate students have a higher level of stress compared to postgraduate students (MacKinnon et al., 2024).

Motivation

Student stress is increasing globally and can have significant impacts on both academic and psychological health. Socioeconomic factors have a direct impact on the daily lives of students and greatly contribute to the amount of stress they endure. Identifying stressed students at the outset is difficult because their stress levels are influenced by a range of sociocultural and economic circumstances. Machine learning (ML) allows for the extraction of intricate relationships within complex datasets that might otherwise evade conventional analysis. The study uses ML to analyse socioeconomic data to create a reliable method for identifying levels of stress among students. Many ML models have been introduced to assess the stress of the student based on many factors, but they don't focus on socioeconomic status. For countries that are developing or underdeveloped, this factor affects a student's stress significantly. Also, we have used data collected from all social statuses of students so that a fair identification can be performed of the stress because of this factor.

An ML approach is proposed in this paper to identify student stress levels using socioeconomic status (SES) variables. Various ML models were evaluated based on how accurately they predicted student stress based on the analysed data. Ensemble models outperformed other classifiers and revealed their ability to identify subtle patterns and relationships within data on economic information. Certain key SES factors are shown to affect student stress, and the study suggests how predictive modelling can contribute to early intervention efforts. The results support initiatives that use data analysis to promote positive student mental health and enhance educational programs.

To collect and preprocess socioeconomic and stress-related data from students to create a comprehensive dataset for analysis.

DOI: https://doi.org/10.5526/esj.453

To evaluate and compare the performance of multiple machine learning models in the accurate classification of student stress levels based on socioeconomic indicators, and other factors.

To identify key socioeconomic factors or other factors influencing student stress and develop a predictive model that can assist educators in early stress detection and intervention planning.

Section 2 outlines the research methodology, including dataset description, sample size, data collection techniques, data preprocessing steps, and ML models used for stress detection and their analysis. Section 3 provides an in-depth analysis of the study's findings. Section 4 discusses the results obtained from the models and compares their accuracy. Section 5 summarizes the recent trends of this sector of the study. Section 6 concludes the paper by summarizing the key findings and proposing future work directions.

Research Method

This paper identifies the stress level among students through their socioeconomic status. It begins by describing the method used to select participants, followed by the techniques used for data collection and the factors of the dataset. Additionally, this section explains the measurement framework using the PSS-10 (Perceived Stress Scale), details the data preprocessing steps, and presents the ML models used for stress classification and analysis. This section provides a framework for understanding how the study was designed, executed, and evaluated.

Population of this Study

The data for this study were collected from students at three types of universities in Bangladesh: public, private, and national. Students of all genders participated in this survey.

Sample Size & Data Collection

Convenience sampling techniques were used in this study. The survey data was collected via Google Forms, and 150 students from different universities filled out this Google form. Among them, 51.9% were female students, and 49.1% were male students. The questionnaire was divided into two parts. The first part of the questionnaire collects demographic information and comprises a total of 14 questions. The second part of the questionnaire is designed to measure the perceived stress level of a student, and for this purpose, we utilize the PSS-10 (Perceived Stress Scale) scale in this study. This survey was completely anonymous, and the responses were kept confidential.

Data Collection Technique

We carried out a survey among undergraduate students using Google Forms. It had 14 questions, and people were allowed to respond between January and February 2025. The identities of the students were not revealed during the process of collecting the data. Table 1 shows the Demographic Questions in the survey, discusses their impact on student stress, and how it is important to identify student stress.

DOI: https://doi.org/10.5526/esj.453

Dataset

The Student_Stress.csv dataset contains raw survey details from university students about their demographics, financial matters, academic situation, and stress levels. Data includes descriptions, classifications, entries with time stamps, and detailed data gathered from questionnaires. The table of the dataset and the description of the factors are shown in Appendix A.

A table of responses of the students is shown in Appendix B from the Student_Stress.csv raw file, and the first 2 rows of the FinalStudent_Stress.csv file is shown in Appendix C, which has been cleaned and organized from the raw CSV file. For this version, textual data is translated into numbers or categories useful for analysis in machine learning. Unused and redundant entries have been eliminated, and important columns have either been renamed or encoded for gender (0 for male, 1 for female). The abbreviations used in Table 1 are as follows: Gen refers to Gender, Rel to Relationship Status, SES to Socioeconomic Status, Fee to Education Fee Provider, Sch to Scholarships, Univ to University Type, Yr to Current Year of Study, CGPA to Cumulative Grade Point Average, Liv to Living Status, Inc to Monthly Family Income, Mem to Family Members, Sleep to Average Sleeping Hours, Tot to Total of the 10 question point (a pseudocode is given in Appendix Section D for the calculation of the stress level), and Stress to feeling nervous and stressed. We've included another feature called Stress Level to sort stress scores into different groups.

PSS-10 (Perceived Stress Scale)

The 10-item Perceived Stress Scale (PSS-10) was created by Cohen et al. (CORC, no date) and is frequently used to measure stress levels in adults and children 12 years of age and older. It assesses how much a person has felt that life is unexpected, out of control, and overwhelming throughout the past month. According to preliminary data, the PSS-10 might make it possible to make insightful comparisons between various racial, ethnic, or linguistic groups. The questions focus on thoughts and feelings throughout the past month. On a five-point scale ranging from "Never" to "Very Often," respondents are asked how frequently they felt a particular way in each instance. Following that, responses are graded as follows:

Table 4: Rating Scale for stress level for the 10 questions in the dataset.

Never	Almost never	Sometimes	Fairly often	Very often
0	1	2	3	4

First, the responses to the four positively stated items score (items 4, 5, 7, and 8) 0 =a > b 4; 1 =a > b 3; 2 =a > b 2; 3 =a > b 1; 4 =a > b0, like this. Then add all 10 item scores. Low stress would be defined

DOI: https://doi.org/10.5526/esj.453

as scores between 0 and 13, while moderate stress would be defined as scores between 14 and 26. High felt stress would be defined as scores between 27 and 40.

Data Preprocessing

Steps were taken to clean the data so it could be adequately analysed and used for modelling. The column that displayed time information was excluded, since it provided no data useful for predicting stress levels. At that point, the column names were changed to Gender, Age, Relationship, Socioeconomic Status, and so on for better clarity and understandability. The survey's ten questions (q1 to q10) were combined into a single score called Total. Stress Level was formed by using a personal Python function on the basis of this score. If the score was 13 or lower, it was 'Low', ranging from 14 to 26 was 'Medium', and scores of 27 or higher were 'High'. The Stress Level was created, after which the individual question columns were dropped. The dataset was examined for missing values using functions such as isnull(), and no null entries were present. Similarly, the dataset was checked for duplicate rows using drop_duplicates(), and no repeated records were detected, ensuring that each observation is unique. Categorical fields contained inconsistent text variations and trailing spaces, so these values were standardized manually. Numerical columns were also inspected for abnormal ranges to detect possible outliers. Finally, the categorical variables were encoded into numbers using label encoding for use in the machine learning algorithms. For instance, Gender has 0 for Male and 1 for Female, and Relationship has 0 for Single and 1 for Married. Furthermore, the same transformations were applied to Socioeconomic Status, Edu_Fee Provider, Scholarships, and other variables. The dataset was randomly split using train_test_split() into 60% training and 40% testing data (test_size=0.40, random_state=1), with no separate validation set.

Machine Learning Model Analysis

The Naive Bayes (NB) classifier is a probabilistic model based on Bayes' theorem, assuming independence between features (Wickramasinghe and Kalutarage, 2021). The Equation 1 predicts the class Ck for a feature vector $\boldsymbol{x} = (x1, x2, \ldots, xn)$ by maximizing the posterior probability given by:

$$P(C_k \mid x) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(x)} \qquad 1$$

The Random Forest algorithm is an ensemble method that constructs multiple decision trees *T1, T2, . . . ,* Tmusing bootstrap samples and random subsets of features (Fawagreh, Gaber, and Elyan, 2014). The final prediction is obtained by majority voting over the individual trees demonstrate in the Equation 2:

$$y = \text{majority vote}\{T_1(x), T_2(x), \ldots, T_m(x)\} \qquad 2$$

DOI: https://doi.org/10.5526/esj.453

A Decision Tree (DT) splits the feature space recursively to create homogeneous nodes (Costa and Pedreira, 2023). A common splitting criterion is the Gini impurity, defined in Equation 3:

$$Gini = 1 - \sum_{k=1}^{K} p_k^2 \qquad 3$$

where p_k is the proportion of samples belonging to class k in a node.

The CART (Classification and Regression Trees) algorithm is a popular type (Cleophas and Zwinderman, 2021, pp. 383-391) of decision tree that builds binary splits by minimizing the weighted Gini impurity in child nodes:

$$\text{Gini}_{\text{split}} = \frac{N_{\text{left}}}{N}\,\text{Gini}_{\text{left}} + \frac{N_{\text{right}}}{N}\,\text{Gini}_{\text{right}} \qquad 4$$

where N_{left} and N_{right} are the number of samples in the left and right nodes, respectively.

Logistic Regression (LR) models the probability of a binary outcome using the logistic sigmoid function (Irimia-Dieguez et al., 2015). Given a linear combination $\boldsymbol{w}^T\boldsymbol{x} + b$, Equation 5 represents the predicted probability:

$$P(y = 1 \mid x) = \frac{1}{1+e^{-(w^\top x+b)}} \qquad 5$$

The K-Nearest Neighbors (KNN) algorithm classifies (Halder et al., 2024) a data point by majority voting among its k closest neighbors, typically using Euclidean distance, shown in Equation 6:

$$d(x, x') = \sqrt{\sum_{i=1}^{n} \left(x_i - x_i'\right)^2} \qquad 6$$

Support Vector Machine (SVM) finds the (Shawe-Taylor and Sun, 2011) hyperplane that maximizes the margin between two classes by solving Equation 7:

$$min\ {}^{1}/_{2}\ \|w\|^2 \text{ subject to } y_i(w^\top x_i + b) \geq 1, \quad \forall i \qquad 7$$

where w is the normal vector to the hyperplane and b is the bias term.

AdaBoost is an ensemble method that combines weak classifiers sequentially (Wang, 2012). At each iteration t, it updates the sample weights $w_i^{(t)}$ to emphasize misclassified examples like in Equation 8:

$$w_i^{(t+1)} = w_i^{(t)} \exp\backslash big\left(\alpha_t\, I\left(y_i, h_t(x_i)\right)\backslash big\right) \qquad 8$$

where αt is the weight of the $t\text{-}th$ classifier and I is the indicator function.

DOI: https://doi.org/10.5526/esj.453

Bagging (Bootstrap Aggregating) trains multiple models on different bootstrap samples and aggregates their predictions, reducing variance and overfitting. It is similar to RF but typically without feature randomization (Kadiyala and Kumar, 2018).

Finally, Gradient Boosting builds models sequentially by fitting the residual errors of previous models (Bentéjac, Csörgő, and Martínez-Muñoz, 2021). At iteration *t*, the model is updated as in Equation 9:

$$F_t(x) = F_{t-1}(x) + \nu \cdot h_t(x) \tag{9}$$

where $ht(x)$ approximates the negative gradient of the loss function with respect to F_{t-1}, and ν is the learning rate controlling the step size.

Results and Discussion

Comparisons between ML models were conducted based on their accuracy, precision, recall, and F1-score, as shown in Table 5. Models including Decision Tree, CART, AdaBoost, Bagging, and Gradient Boosting achieved the highest performance with an accuracy of 0.9792. Random Forest also shows an accuracy of 0.9375. In contrast, simpler models such as Naive Bayes and K-Nearest Neighbors performed poorly, achieving accuracies of 0.3958 and 0.5417. Logistic Regression (0.6667) and Support Vector Machines (0.5625) achieved moderate accuracy, suggesting limited suitability for modelling stress when relationships are highly nonlinear. To further investigate the specific influence of socioeconomic factors, a separate experiment was conducted using only socioeconomic features. As shown in Table 6, model performance decreased significantly, with the highest-performing model achieving only 46% accuracy. This indicates that stress levels cannot fully depend on socioeconomic factors alone, and that behavioural, academic, and psychological factors play a significant role too.

Table 5: Metrics of all Factor Models

Model	Accuracy	Precision	Recall	F1-Score
NB (Naive Bayes)	0.3958	0.3481	0.3395	0.2243
Random Forest	0.9375	0.7982	0.7593	0.7730
Decision Tree (DT)	0.9792	0.9167	0.9815	0.9429
CART	0.9792	0.9167	0.9815	0.9429
Logistic Regression (LR)	0.8125	0.5344	0.5741	0.5527
KNN	0.5417	0.3817	0.3580	0.3632
SVM	0.5625	0.1875	0.3333	0.2400

DOI: https://doi.org/10.5526/esj.453

AdaBoost	0.9792	0.9167	0.9815	0.9429
Bagging	0.9792	0.9167	0.9815	0.9429
Gradient Boosting	0.9792	0.9167	0.9815	0.9429

Table 6: Random Forest Model Metrics for SES

Metric	Score
Accuracy	**0.46**
Recall	**0.46**
F1-Score	**0.36**
Precision	**0.36**

Analysis of All Factors vs Socioeconomic Factors

Models trained using the complete feature set achieved high classification accuracy, with tree-based models such as Decision Tree, CART, AdaBoost, and Bagging achieving 97.92% accuracy. The feature importance plot (Figure 2) highlights that the Total stress score contributed the most to classification, followed by academic and lifestyle attributes such as CGPA, Sleeping Hours, and Family Income. This outcome is expected because the Total score is derived from stress-related questionnaire responses and therefore strongly shows students' perceived stress levels. These findings indicate that when stress-related behavioural factors are all included, ML models can classify stress accurately. Evaluation of the socioeconomic factors, a secondary experiment was conducted using only socioeconomic features. The SES-only Random Forest model achieved an accuracy of 46%, significantly lower than the full-feature model. The feature importance results (Figure 1) indicate that Family Income was the strongest socioeconomic predictor, followed by Living Status and Socioeconomic Class, while the fee provider had comparatively less influence. This demonstrates that socioeconomic conditions contribute to stress, but stress doesn't depend only on the SES factors.

Figure 1: Feature importance for SES-only model

DOI: https://doi.org/10.5526/esj.453

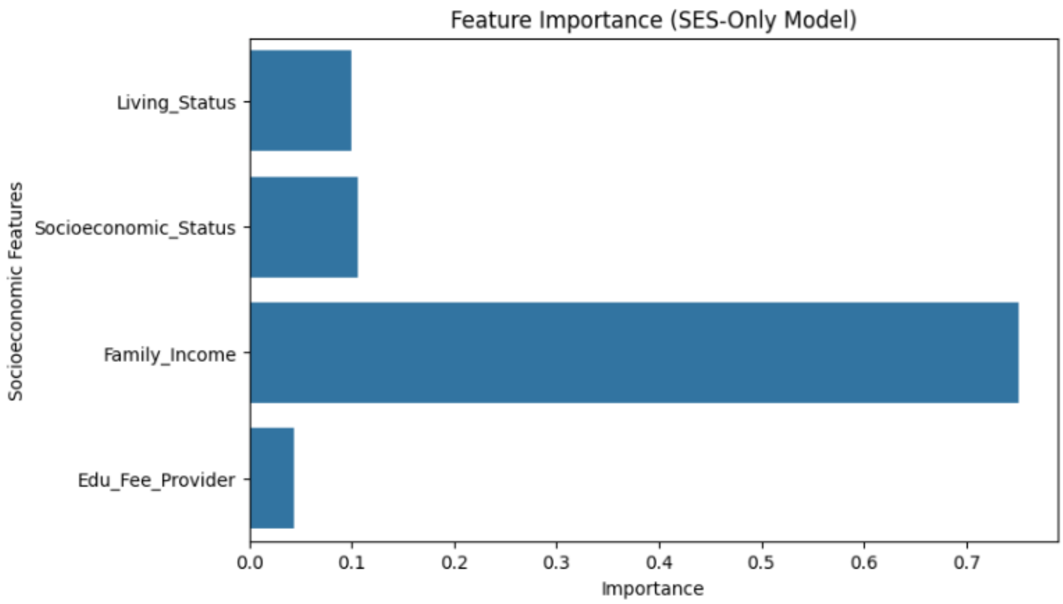
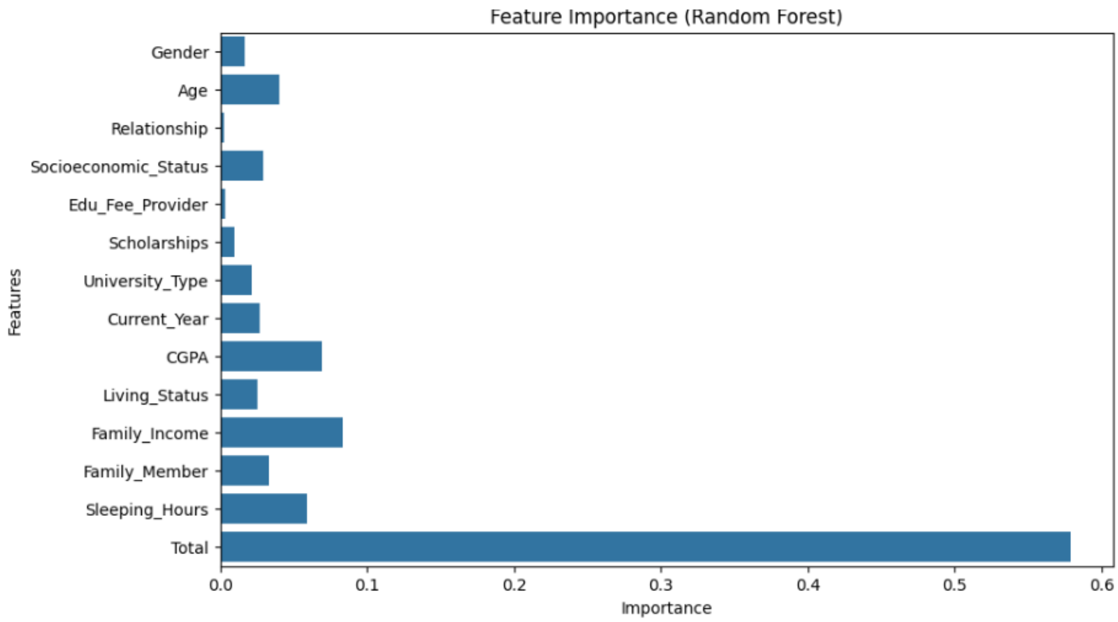


Figure 2: Feature importance for full-feature model



Recent Research Trends

Table 7 shows a comparison of several 2024 and 2025 experiments that applied ML and DL models to diverse data. The great majority of the datasets were gathered by participants, except for one that was taken from a public source. There is a big range in the dataset sizes seen, from just 1,602 entries to over 61,000. When measured by accuracy, performance metrics do change, but the best results are reported as 91.1% when using both ML and DL models on a public data set. Alternatively, an ML model alone on a dataset gathered by us achieved 66.2% accuracy, but a DL model on another set collected by us achieved 88%. They show how the kind of data, the amount, and the way it's modelled matter for predictive performance.

DOI: https://doi.org/10.5526/esj.453

Table 7: Summary of datasets and model performance

Ref	Year	Dataset	ML/DL Model	Data Size	Accuracy (%)
(Kong et al., 2025)	2025	Own collected	ML	3,281	66.2
(Ma, 2025)	2025	Own collected	DL	–	88
(Al Masud et al., 2025)	2025	Public Dataset	ML and DL	1,602	91.1

Conclusion & Future Work

The study examined different ML models to determine if SES could be used to predict stress levels among students. Ensemble-based models, specifically Decision Trees, CART, AdaBoost, Bagging, and Gradient Boosting, had the most accurate results (97.92%). Their effectiveness suggests that they are well-suited for discovering intricate relationships and patterns present in the data. However, Naive Bayes and K-Nearest Neighbors models were less effective because they require specific data characteristics and are more heavily influenced by variable distributions. These results highlight the potential for ML to aid in gaining psychological and educational understanding, as well as help identify groups of students at risk due to their socioeconomic factors to some level. Early warning signs of distress from these models enable teachers and counsellors to take necessary actions to protect students' emotional health. Future research could incorporate cross-validation, tuning hyperparameters, and computing metrics like precision, recall, and F1 score to improve performance assessment.

Appendix

A: Table 1: Demographic Questions and Description

#	Demographic Question	Description
1	Gender	Gender roles, support networks, and cultural expectations affect stress.
2	Age	Different age groups face unique hurdles: older students worry about graduation and employment, younger students about independence.
3	Relationship Status	Relationship status influences stress: partnerships can bring emotional support or relational stress.
4	Socioeconomic Status	Financial security impacts stress, especially worries about resources, living costs, and tuition.
5	Education Fee Provider	Paying education fees can cause additional stress.

DOI: https://doi.org/10.5526/esj.453

6	Scholarships	Lack of scholarships causes financial distress: scholarships may increase performance pressure.
7	University Type	Resources, atmosphere, and financial aid differ between public and private colleges.
8	Current Year of Study	Stress varies by year: adjustment stress in first year, graduation or career stress in final year.
9	CGPA	Academic pressure to maintain grades corresponds with stress.
10	Living Status	Living alone may cause loneliness: sharing space may reduce privacy or stability.
11	Family Income	Lower income limits access to resources, increasing stress.
12	Family Members	Family size affects financial obligations and support networks.
13	Chronic Illness	Managing health and academics is stressful with chronic conditions.
14	Sleeping Hours	Poor sleep worsens anxiety and coping.

B: Transposed View of First Two Responses (Raw Dataset)

Question	**Respondent 1**	**Respondent 2**
Timestamp	2024/06/19 12:01:02 AM GMT+6	2024/06/19 7:40:02 PM GMT+6
Gender	Male	Female
Age	24	24
Relationship Status	Single	Single
Socioeconomic Status	Upper Middle Class	Upper Middle Class
Education Fee Provider	Parents	Parents
Scholarships	No	No
University Type	Private	Private
Current Year of Study	4	3
CGPA	3.76	2.97
Living Status	With Family	With Family
Monthly Family Income	220000	50000

DOI: https://doi.org/10.5526/esj.453

Family Members	4	6
Average Sleeping Hours	8	8
Unable to control important things	1	5
Felt nervous and stressed	1	5
Angered by things outside your control	1	5
Could not cope with responsibilities	1	5
Felt difficulties piling up	1	5
Felt things going your way	1	5
Able to control irritations	1	4
Felt confident handling problems	1	5
Felt on top of things	1	5

C: First Two Rows of Cleaned Dataset.

Gen	Age	Rel	SES	Fee	Sch	Univ	Yr	CGPA	Liv	Inc	Mem	Sleep	Tot	Stress
0	24	0	1	0	1	1	4	3.76	0	220000	4	8	10	0
1	24	0	1	0	1	1	3	2.97	0	50000	6	8	49	2

D: Pseudocode for Stress Level Detection

Stress_Level(Total):

If Total ≥ 27 → High

Else if Total ≤ 13 → Low

Else → Medium

DOI: https://doi.org/10.5526/esj.453

References

Al Masud, G.H., Shanto, R.I., Sakin, I. and Kabir, M.R. (2025) 'Effective depression detection and interpretation: Integrating machine learning, deep learning, language models, and explainable AI', *Array*, 25, p. 100375. Available at: https://doi.org/10.1016/j.array.2025.100375

Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G. (2021) 'A comparative analysis of gradient boosting algorithms', *Artificial Intelligence Review*, 54, pp. 1937–1967. Available at: https://doi.org/10.1007/s10462-020-09896-5

Cleophas, T.J. and Zwinderman, A.H. (2021) *Regression Analysis in Medical Research: for Starters and 2nd Levelers,* 2nd edn. Springer.

CORC (no date) *Perceived Stress Scale (PSS-10).* Available at: https://www.corc.uk.net/outcome-experience-measures/perceived-stress-scale-pss-10/ (Accessed: 17 July 2025).

Costa, V.G. and Pedreira, C.E. (2023) 'Recent advances in decision trees: An updated survey', *Artificial Intelligence Review*, 56, pp. 4765–4800. Available at: https://doi.org/10.1007/s10462-022-10275-5

Di Mario, S., Rollo, E., Gabellini, S. and Filomeno, L. (2024) 'How stress and burnout impact the quality of life amongst healthcare students: an integrative review of the literature', *Teaching and Learning in Nursing*, 19(4), pp. 315-323. Available at: https://doi.org/10.1016/j.teln.2024.04.009

Erola, J., Jalonen, S. and Lehti, H. (2016) 'Parental education, class and income over early life course and children's achievement', *Research in Social Stratification and Mobility*, 44, pp. 33–43. Available at: https://doi.org/10.1016/j.rssm.2016.01.003

Fawagreh, K., Gaber, M.M. and Elyan, E. (2014) 'Random forests: from early developments to recent advancements', *Systems Science & Control Engineering: An Open Access Journal*, 2(1), pp. 602–609. Available at: https://doi.org/10.1080/21642583.2014.956265

Gao, W., Ping, S. and Liu, X. (2020) 'Gender differences in depression, anxiety, and stress among college students: a longitudinal study from China', *Journal of Affective Disorders*, 263, pp. 292–300. Available at: https://doi.org/10.1016/j.jad.2019.11.121

Halder, R.K., Uddin, M.N., Uddin, M.A., Aryal, S. and Khraisat, A. (2024) 'Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications', *Journal of Big Data*, 11, article 113. Available at: https://doi.org/10.1186/s40537-024-00973-y

DOI: https://doi.org/10.5526/esj.453

Irimia-Dieguez, A.I., Blanco-Oliver, A. and Vazquez-Cueto, M.J. (2015) 'A comparison of classification/regression trees and logistic regression in failure models', *Procedia Economics and Finance*, 23, pp. 9–14. Available at: https://doi.org/10.1016/S2212-5671(15)00493-1

Kadiyala, A. and Kumar, A. (2018) 'Applications of Python to evaluate the performance of bagging methods', *Environmental Progress & Sustainable Energy*, 37(5), pp. 1555–1559. Available at: https://doi.org/10.1002/ep.13018

Kong, W., Pei, Z., Guo, Z., Xu, R. and Zhao, J. (2025) 'Relationship matters: Using machine learning methods to predict the mental health severity of Chinese college freshmen during the pandemic period', *Journal of Affective Disorders*, 369, pp. 392–403. Available at: https://doi.org/10.1016/j.jad.2024.09.168

Ma, X. (2025) 'Research on virtual entertainment robots based on machine learning algorithms providing psychological health services for college students', *Entertainment Computing*, 52, p. 100819. Available at: https://doi.org/10.1016/j.entcom.2024.100819

MacKinnon, M., Mitchell, S. and Linden, B. (2024) 'A cross-sectional comparison of the association between self-reported sources of stress and psychological distress among Canadian undergraduate and graduate students', *Journal of American College Health*, 73(9), pp. 3625-3635. Available at: https://doi.org/10.1080/07448481.2024.2431714

Meier-Faust, E. and Watermann, R. (2024) 'Perceiver Effects and Socioeconomic Background: Contrasting Parent-Reports against Teacher-Reports of Elementary School Students' Personality', *Journal of Personality Assessment*, 106(4), pp. 482–495. Available at: https://doi.org/10.1080/00223891.2023.2286449

Nja, C.O., Neji, H.A., Orim, R.E., Ukwetang, J.O., Ideba, M.A., Cornelius-Ukpepi, B. and Ndifon, R.A. (2022) 'The socio-economic rank of parents and students' academic and cognitive outcomes: Examining the physical, psychological and social mediators', *Frontiers in Education*, 7, p. 938078. Available at: https://doi.org/10.3389/feduc.2022.938078

Öztekin, G.G., Gómez-Salgado, J. and Yıldırım, M. (2025) 'Future anxiety, depression and stress among undergraduate students: psychological flexibility and emotion regulation as mediators', *Frontiers in Psychology*, 16, p. 1517441. Available at: https://doi.org/10.3389/fpsyg.2025.1517441

Ribeiro, Í.J.S., Pereira, R., Freire, I.V., de Oliveira, B.G., Casotti, C.A. and Boery, E.N. (2018) 'Stress and quality of life among university students: A systematic literature review', *Health Professions Education*, 4(2), pp. 70–77. Available at: https://doi.org/10.1016/j.hpe.2017.03.002

DOI: https://doi.org/10.5526/esj.453

Shawe-Taylor, J. and Sun, S. (2011) 'A review of optimization methodologies in support vector machines', *Neurocomputing*, 74(17), pp. 3609–3618. Available at: https://doi.org/10.1016/j.neucom.2011.06.026

Wang, R. (2012) 'AdaBoost for feature selection, classification and its relation with SVM, a review', *Physics Procedia*, 25, pp. 800–807. Available at: https://doi.org/10.1016/j.phpro.2012.03.160

Wickramasinghe, I. and Kalutarage, H. (2021) 'Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation', *Soft Computing*, 25, pp. 2277–2293. Available at: https://doi.org/10.1007/s00500-020-05297-6

Yılmaz Koğar, E. and Koğar, H. (2024) 'A systematic review and meta-analytic confirmatory factor analysis of the perceived stress scale (PSS-10 and PSS-14)', *Stress and Health*, 40(1), p. e3285. Available at: https://doi.org/10.1002/smi.3285

Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., Lin, Y.D., Xu, Z., Daniels, A.D., Liu, Y. et al. (2025) 'Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students', *Journal of Counseling & Development*, 103(1), pp. 110–125.

Copyright Statement

DOI: https://doi.org/10.5526/esj.453