Journal of Information Systems Engineering and Management
2025, 10(13s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Learning Student Stress Intention: An Interdisciplinary Analysis of Psychological, Environmental, and Academic Contributors using Machine Learning

Varsha D. Jadhav[1], Kirti Wanjale[2], Amar Buchade[3] Avinash Raut[4] Shailesh Kulkarni Dhananjay R. Dolas[4]
[1]*Artificial Intelligence and Data Science Department, Vishwakarma Institute of Technology, Pune, Maharashtra, India.*
[1]*drvarshajadhav22@gmail.com*
[2] *Department of Computer Engineering,, Vishwakarma Institute of Technology, Pune, Maharashtra, India., kirti.wanjale@vit,edu*
[3]*Artificial Intelligence and Data Science Department, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India amar.buchade@viit.ac.in*
4 Department of Sciences Engineering and Humanities, , Vishwakarma Institute of Technology, Pune, Maharashtra, India. avinash.raut@vit.edu
5 Department of Electronics and Telecommunication, , Vishwakarma Institute of Technology, Pune, Maharashtra, India. shailesh.kulkarni@vit.edu
[6] *Department of Mechanical Engineering, Jawaharlal Nehru Engineering College, MGM University, Chhatrapati Sambhaji Nagar, Maharashtra, India.*

ARTICLE INFO

Received: 30 Nov 2024

Revised: 12 Jan 2025

Accepted: 30 Jan 2025

ABSTRACT

The research work focuses into all of the factors which contribute to student stress intention, through a dataset with 20 features spanning psychological, physiological, environmental, academic, and social domains. The research utilizes statistical tests and machine learning models such as Decision Tree, Random Forest, K Neighbors, and Gaussian Naive Bayes to find major stress predictors and evaluate model performance. Key findings show important relationships between every attribute and stress levels, with Gaussian Naive Bayes having the highest test accuracy. Feature importance analysis and dividing with Variational Autoencoders (VAE) and K-means reveal distinctive stress profiles, particularly in clusters 1 and 4, which might profit from centered mental health interventions. Association rule mining shows much stronger correlations between mental health signs like depression and anxiety. The study highlights the need for more data, ideal models, and customized interventions that better comprehend and manage student stress intention.

Keywords: Stress intention, Annova, Chi-square-test, decision tree, random forest-nearest neighbors, variational autoencoders, K-means clustering.

1. INTRODUCTION

In today's educational environment, recognizing and handling student stress has become increasingly important. This study examines numerous factors which contribute to student stress, covering psychological [1], physiological [2,3,4], environmental [5], academic [6,7], and social factors [8]. The objective of this research is to illuminate the complex factors driving student stress levels by using a large dataset with 20 distinct features. The study utilizes statistical analysis and machine learning models such as Decision Tree, Random Forest, K Neighbors, and Gaussian Naive Bayes to determine major stress predictors and contrast the predictive power of different classification algorithm design [9]. Additionally, clustering algorithms and association rule mining are utilized to find patterns and correlations in the data [10]. This multimodal approach aims to offer significant understanding into the factors that affect student stress and to support the development of specific remedies and support strategies.

2. LITERATURE SURVEY

Understanding student stress intention and its various origins has been the topic of important research, emphasizing its significance in educational and psychological studies. The literature recognizes a number of stress-related elements, which are roughly classified as psychological, physiological, environmental, academic, and social domains, and this study followed pace. Recent studies for the student stress have shown the significance of psychological factors. The authors [11] examines the development and evaluation of mobile app designed to help students to manage their time and reduce the stress, It uses convincing methods to encourage enhanced time management skills which reduces the stress. The study helps to understand and improve the use of technology to improve mental health and efficiency in educational environment. The results of the research [12] imply that feeling both good and mentally and emotionally is crucial in determining how self-compassion, stress, and self-esteem are connected with each other. It suggests that these factors help students to grown and build good self-esteem. The study further suggests that emphasizing on mindfulness and self-compassion can help in developing methods to improve the mental health of students. The study discloses [13] that stress, especially workplace, social, economic, and discrimination-related stress, is an important contributor to hypertension. Research is now concentrating on both stress response and delayed recovery as factors contributing to hypertension. This review [14] explores how stress impacts physical activity, that frequently gets diminished, just as it affects other health habits. However, other research imply that stress may improve activity as a coping mechanism. Future study should focus on why some people stay active in stressful situations, as well as the need for more deeply studies and theories to guide interventions, especially in at-risk groups such as older persons. This research [15] indicates that noise, especially traffic noise, has an adverse effect on health through causing stress, which can lead to heart disease, stroke, and hypertension. It also has an effect on mental health, increasing anxiety, depression, and harmful habits such as smoking. Noise boosts stress hormones and inflammation, which damages blood vessels and neurons. Social support at work can boost relationships, performance, and stress management [16]. However, research experiences inconsistent definitions, broken thoughts, and imprecise measurements. Stress affects students' [17] academic performance. A study of 284 students revealed that stress from higher education, finances, and relationships affects performance. Universities should improve educational environments and offer therapy to help students manage with stress. The study [18] indicated that restlessness, sadness, and concern are significant symptoms for depression and anxiety. Guilt is linked with job stress. No gender distinctions were discovered, but Wuhan grads endured less symptoms than others, likely due to geographical impact. During the COVID-19 pandemic, people from South Africa suffered major mental health worries [19]. The ability to cope offered some protection against stress, while social support had little impact. This suggests that, although coping resources were successful, stressed social ties hindered the effects of social support. The authors [20] present a method for categorizing student stress using ensemble learning. It achieved 93.48% accuracy by analysing factors like sleep and screen time. The model helps identify stress levels and suggests improvements for better academic well-being. The research investigates stress factors among students at Tribhuvan University in Nepal and using a variety of machine learning models to detect and prevent stress [21]. The Naive Bayes model obtained 90% accuracy, while the SVM achieved 85.45%. The survey discovered that the academic term is the most stressful for students. Recent study combines both traditional and advanced techniques for comprehending and controlling student stress. This study applies progressed machine learning approaches to analyse and reduce student stress issues. The authors proposed a machine – learning based model to identify detrimental content and enhance accuracy for effective deterrence [22]. The authors have demonstrated AI-powered chatbot developed using SVM which improve mental health by accurately classifying the intentions in terms of emotions, making sure the real-time responses [23].

3. DATASET

A comprehensive dataset of student stress which enhances the hidden dynamics is downloaded from Kaggle repository [24]. 20 selected features related to student stress are examined which contribute to teal life factors. The features are grouped into five major categories, confirming the factors that influence stress in a complete understanding. The five factors are as follows:

i. Psychological
ii. Physiological
iii. Environmental
iv. Academic
v. Social

Psychological factor consists of four subfactors. Anxiety level which measures how oftenly the students feel anxious. It is represented by numerical value indicating the level of anxiety. Higher count shows frequent anxiety among the students. Self-Esteem factor estimates the self-worth and confidence of the students. It is a score which reflects the self-esteem level of individual, with higher numbers indicating high self-esteem. Any prior mental health issues are considering through the factor mental health history. It is represented as binary or categorical value (0 or 1) indicating that whether the student is having history of mental health issue or not. The signs of depression among the students are assessed by the depression factor. Higher score of depression represents severity of depression.

The physiological factor comprises of four subfactors. A common stress indicator, headache tracks the frequency of headaches. It is represented by a score indicating severity or frequency of headache. Student blood pressure levels are monitored by blood pressure. It is given by a score related to blood pressure, indicating risk levels. The quality and duration of sleep is evaluated by the factor sleep quality. Lower scores reflect poor sleep quality. Breathing difficulties are recorded by breathing problems. It is a score indicating the frequency of breathing problem,

The four subfactors of the environmental factors shows the surrounding conditions of the students. The noise in student living and learning environment is measured by the noise level factor. It is a score that indicates that indicates the noise level of the students living conditions. The students living arrangements quality is assessed through living conditions. The safety of students in their surrounding is evaluated by safety factor. The basic need of students is food and shelter which is ensured through the facto basic need. It is a score which represents how well the students' basic requirements are met.

The four subfactors of the academic factors reflects the academic performance of the students. Students grades and overall academic achievements is recorded by academic performance. It is given a score which represents the students' performance in the academic activities. The amount of academic work students should complete is measured by study load. It is represented by a numeric value showing the study load or the study burden. Teacher student relationship factor evaluates the quality of student teacher interactions. It is a score representing the relating the quality of relationship the teacher and the student is having. The career prospects and higher education worries are tracked through future career concerns. It is represented by a score indicating the students concern about the future career.

The four subfactors of the social factors measures the social touch of the students. The support received by the students from friend, family and nobles are measured through social support in terms of a numerical value. The influence of nobles on students' behaviour and stress are evaluated by peer pressure is indicated by numeric score. Participation of students in the classroom outside activities is recorded through extracurricular activities is given by a score. Students get experience of bullying both in person and online which is tracked by the factor bullying. It is represented by a score indicating how much bullying is experienced by an individual. The overall stress experienced by an individual is represented by a score.

The dataset consists of interesting and understandable stress factors affecting the students and was created through a survey. Each number represents a response with higher numbers indicating higher intensity, frequency or quality depending on the context of the variable. It helps to uncover the insights and builds a better support system and healthy learning environment.

Each column in the dataset has unique values of stress intention resented in table 1.

Table 1. Unique values of Stress Intention.

Stress Intention	Count of Unique entries
anxiety_level	22
self_esteem	31
mental_health_history	28
depression	6
headache	3
blood_pressur	6
sleep_quality	6
breathing_problem	6
noise-level	6
living_conditions	6
safety	6
basic_needs	6
academic_performance	6

study_load	6
teacher_student_relationship	6
future_career_concerns	6
social_support	4
peer_pressure	6
extracurricular_activities	6
bullying	6
stress_level	3

The distribution of each variable in the dataset is by visualization in figure 1. KDE plots show the continuous variables with more than 6 unique values, which displays smooth distribution. The categorical or ordinal variables with 6 or less unique values are pictured with count plots to illustrate the frequency of each category.

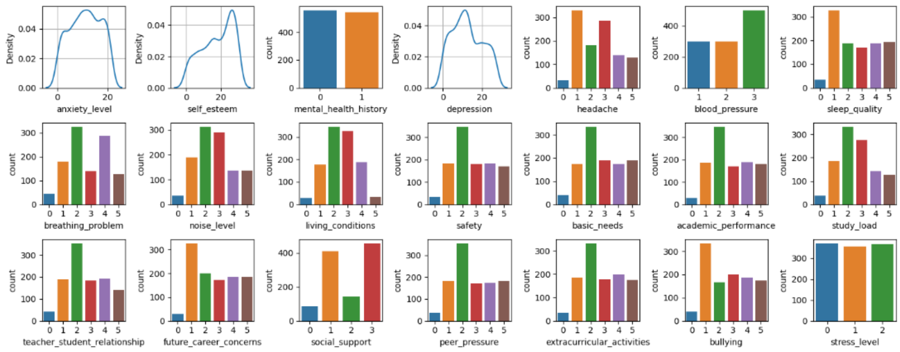

Fig 1. Distribution of each variable in the stress intention dataset.

Statistical Analysis

Based on the nature of data in each column, correct statistical analysis is performed. An array of features displays showing the statistically significant differences where $p < 0.05$ across the different levels of strees_level. For example if it displays anxiety_level,depression and sleep_quality, it means that they are significantly associated with stress_level. This helps in recognizing which factors are sturdily connected to stress levels in dataset. The experimentation shows the following array.

array(['anxiety_level', 'self_esteem', 'mental_health_history', 'depression', 'headache', 'blood_pressure', 'sleep_quality', 'breathing_problem', 'noise_level', 'living_conditions', 'safety', 'basic_needs', 'academic_performance', 'study_load', 'teacher_student_relationship', 'future_career_concerns', 'social_support', 'peer_pressure', 'extracurricular_activities', 'bullying'], dtype=object)

It shows that all the features are significantly associated with the stress_level. The p-values are as shown in table 2

Table 2. p-values related to Stress Intention

Stress Intention	**Count of Unique entries**
anxiety_level	5.2967e-188
self_esteem	1.2685e-210
mental_health_history	8.2021e-131
depression	1.9204e-187
headache	1.6343e-170
blood_pressur	9.2773e-264
sleep_quality	1.3973e-198
breathing_problem	1.7740e-98
noise-level	3.1688e-141
living_conditions	3.7522e-99
safety	1.3970e-179
basic_needs	8.5433e-177
academic_performance	9.0600e-185

study_load	3.2897e-124
teacher_student_relationship	5.1737e-158
future_career_concerns	1.2285e-193
social_support	1.1160e-138
peer_pressure	7.6330e-164
extracurricular_activities	4.0076e-166
bullying	4.2270e-199

Statistical analysis is carried out using ANNOVA and Chi-square test [25].ANNOVA (Analysis of variance) compares the means of three or more groups to identify whether there is statistically significant difference between three or more groups. The key components of ANNOVA are:

Total sum of squares (SST): Total variability in the data is measured by SST. It equals the total of the squared differences between each observation and the overall mean.

$$SST = \sum_{i=1}^{N}(y_i - \bar{y}_i)^2 \qquad (1)$$

Between Groups Sum of Squares (SSB): Measures the variability caused by variations in group means.

$$SSB = \sum_{j=1}^{k} n_j\left(\bar{y}_j - \bar{y}\right)^2 \qquad (2)$$

The degree of freedom (dfB) is calculated as k-1.

Within Groups Sum of Squares (SSW): Variability within each group is measured.

$$SSW = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)^2 \qquad (3)$$

Degree of freedom (dfW) is calculated as N-k

Mean Squares: The mean squares is calculated as follows:

$$Between\ Groups\ Mean\ Square\ (MSB) = \frac{SSB}{dfB}$$

And

$$Within\ Groups\ Mean\ Squares\ (MSW) = \frac{SSW}{dfW}$$

F-Statistics: Compares the variance between groups to the variation within groups to see if the differences in group averages are statistically significant.

$$F = \frac{MSB}{MSW}$$

p-value is associated with the F-statistics. If p-value is less than significance level of 0.05 then there is significant difference between the means of the groups.

The Chi-Square Test is used to determine is there is a significant relationship between two category variables. The test compares the observed frequencies in each category to the frequencies that would be predicted if the variables were not associated. The test test statistics is computed as

$$X^2 \sum_i \sum_j \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad (4)$$

Where Oij is the observed frequency in the cell (i,j), Eij is the expected frequency in the cell (i,j), the summation is the overall cells in the contingency table. The degree of freedom is calculated as

$df = (r-1) \times (c-1)$ where r and c are the rows and columns in the contingency table. The p-value is determined by comparing the calculated X2 statistic to the Chi-square distribution with df degrees of freedom.

The Chi-Square Test determines if the observed distribution of frequencies differs significantly from what would be predicted if the variables were independent. If the p-value is less than the significance level of 0.05, it suggests a meaningful relationship between the two category variables.

4.1. Splitting the dataset into train and test

This is very important step in machine learning to evaluate the performance of the model on unseen data. In this case 30 percent of the data is used for testing and 70 percent will be used for training.

Machine Learning Models used for Classification

Four machine learning models are used for classification. Decision Tree, Random Forest, K Neighbors and gaussian Naïve Bayes.

5.1. Decision Tree Classifier

Works by recursively dividing the data based on feature values to maximize information gain or decrease impurity, leading to a model that is easy to understand and applicable to a wide range of data formats. It can handle both numerical and categorical data. Information gain is used to determine which feature to split on [26]. It computes the drop in entropy following a split. The information gain for a feature (A) and a dataset (D) is as follows:

$$Information\ Gain(D,A) = Entropy(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} Entropy(D_v) \tag{5}$$

Entropy measures impurity or randomness in data. The definition of entropy for a dataset (D) and classes (C) is as follows:

$$Entropy(D) = -\sum_{i=1}^{n} p_i log_2(p_i) \tag{6}$$

Where pi is the probability of class i in the dataset.

5.2. Decision Tree Classifier

The Random Forest classifier use ensemble methods to combine many decision trees, each trained on a bootstrap sample and a subset of features [27]. Classification includes voting by majority among trees, while regression involves averaging all trees' predictions. This technique enhances model accuracy and robustness. The classifier calculates and sorts feature importances. Sorting feature importances enables you to observe which features are the most important for making predictions, which can help guide feature selection and engineering.

5.3K Neighbors Classifiers (KNN)

KNN classifies new data points using the majority class of their k nearest neighbors, using distance metrics such as Euclidean or Manhattan distance. It's a straightforward strategy that works well with little datasets and complex limits on decisions [28]. The most commonly used distance metric in KNN. In an n-dimensional space, the Euclidean distance (d) between two points (xi and xj) is as follows:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{kj})^2} \tag{7}$$

5.3 Gaussian Naive Bayes

It is a simple, probabilistic method for classifying data that assumes features follow a normal distribution and are independent of one another [29]. It computes the probabilities for each class and selects the class with the highest probability. The probability of a data point belonging to a class is calculated as

$$Probability\ of\ a\ Class = \frac{Probability\ of\ Features\ Given\ Class \times Probability\ of\ Class}{Probability\ of\ Features} \tag{8}$$

The likelihood of it belonging to a class is calculated as:

$$Likelihood = \frac{1}{Size\ of\ the\ Curve} \times Exponent\ Term \tag{9}$$

Naïve assumption is that features are independent of each other within the same class, so as to calculate which class has the highest probability based on the features.

5.4 Variational Autoencoders (VAE) and K-means Clustering

This strategy is particularly useful for detecting patterns in high-dimensional data by lowering its dimensionality prior to categorization. The data is normalized and transformed into PyTorch tensors for batch processing. The VAE compresses data into a latent space, which is subsequently reconstructed by a decoder [30]. The encoder computes the mean and log-variance of the latent space, from which a latent variable (z) is selected. The VAE is trained for 100 epochs using Mean Squared Error (MSE) loss and the Adam optimizer. Upon training, the encoder generates latent space representations of the input data. K-means are used to cluster data based on latent space representations.

Using the Apriori algorithm, association rule mining analyzes correlations between mental health features [31]. Anxiety levels are estimated based on self-esteem and academic achievement, offering a predictive indicator for future analysis or decision-making using linear regression. The ANOVA test determines whether anxiety levels differ significantly between clusters, providing insight into how anxiety varies between the identified groups. It classifies anxiety levels by cluster and computes the F-statistic and p-value. A significant p-value suggests that anxiety levels differ among clusters.

RESULTS AND DISCUSSION

As shown in table 2, the extremely small p-values indicate that there are statistically significant differences or associations for all the features tested against stress-level. These factors are strongly related to or affected by stress level, making them crucial in understanding how stress levels vary with different factors in the dataset. The train and test accuracies of the illustrated in table 3.

Table 3. Train and Test Accuracies of the Machine Learning Models.

Classifiers	Decision Tree Classifier	Random Forest Classifier	K Neighbors Classifiers (KNN)	Gaussian Naive Bayes
Train Accuracies	1.0	1.0	0.8766	0.8753
Test Accuracies	0.8909	0.8939	0.8878	0.9030

Figure 2 shows the pictorial representation of accuracies.

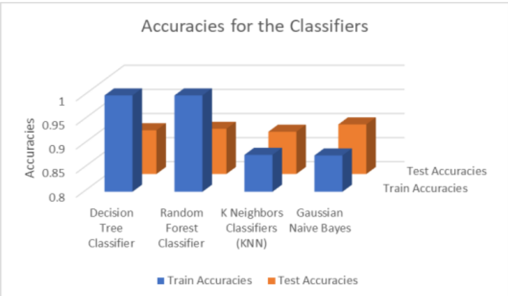

Fig. 1 Accuracies for Classifiers.

Decision tree classifier and Random Forest classifier has best train accuracy. These classifiers match the training data perfectly, although they may overfit. Gaussian Naïve Bayes has best test accuracy. It achieves the best accuracy on the test set, indicating good generalization performance. K Neighbors classifier shows balance between training and test performance, having poor training accuracy but high-test accuracy.

Important features should be considered so as to reduce complexity and potentially improve performance. Figure 3 shows the feature importances from a trained random forest model. The values related with each feature designates their comparative importance in models' prediction.

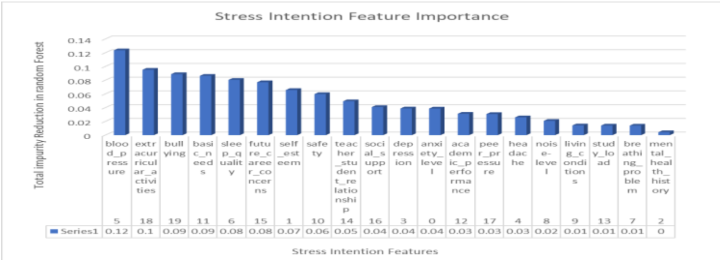

Fig. 3 Stress intention Feature Importance.

The value 0.124 for feature 5 indicates that feature 5 is responsible for 12.4% of the total impurity reduction in the random Forest model, making it more significant feature in expecting the target variable. The value is related to other characteristics, demonstrating how much each feature contributes to the model's decision-making process. Assists with feature selection, model interpretation, and understanding of underlying data patterns. Applying K-means clustering on the latent representation for n clusters produces a scatter plot in the latent space. In the experimentation n=8 clusters are considered. Figure 4 shows the clustering in VAE latent space.

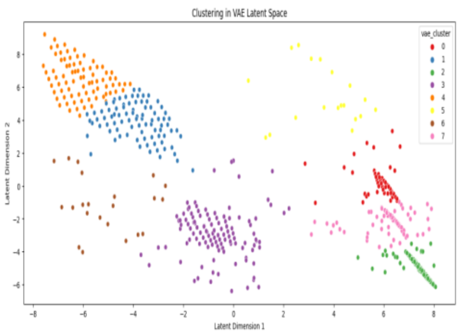

Fig. 4 Clustering in VAE Space.

Clusters from VAE and K-means clustering are evaluated to identify those that may require focused interventions. It calculates the average anxiety and self-esteem levels for each cluster and then assesses whether any cluster surpasses an anxiety threshold but falls below a self-esteem threshold. If both criteria are met, the code suggests that the cluster may benefit from targeted mental health interventions. The results shows that cluster 1 and 4 may benefit from targeted mental health intervention. The Silhouette Score calculates how well each data point is grouped. It ranges from -1 to 1.

1: Points are appropriately matched to their cluster.

0: Points are on the border between clusters.

-1: Points may be in the incorrect cluster. A higher score suggests more clearly defined clusters.

The results show the number of clusters and the Silhouette Score in table 4

Table 1. Title of the table

Number of Clusters	Silhouette Score
2	0.6130
3	0.6631

4	0.6704
5	0.6040
6	0.6178
7	0.5954
8	0.5844
9	0.5694

The Apriori algorithm detects patterns in a binary matrix comprising features such as depression, anxiety, and self-esteem. It converts these attributes into binary numbers based on whether they are higher or lower than the median. The algorithm finds frequently used itemsets and generates association rules that determine feature correlations. The results contribute to a better understanding of how different levels of these traits interact.

The result displays association rules between mental health features. Antecedents and Consequents: Rules-related features. Support: The frequency with which both the antecedent and the consequent occur together. Confidence: The likelihood of the consequence occurring given the antecedent. Lift: Determines the strength of the correlation; values greater than 1 indicate a strong relationship. Leverage is the difference between observed and expected support. Conviction: Degree of implication; higher values suggest more stringent regulations. Zhang's Metric: An alternative measure of rule strength. For example, Rule 1: If a person suffers from depression, they are more likely to experience high levels of anxiety, with a confidence of 77.5% and a lift of 1.60, indicating a strong link. Rule 2: If a person has a high level of anxiety, they are more likely to suffer from depression, with a confidence of 74.2% and a lift of 1.60, indicating a strong link.

ANNOVA test results are: F-statistic=704.05 and p-value =1.18 × 10-222 . A high F-statistic implies that the variance in anxiety levels between clusters is much greater than the variance within clusters. The p-value is extremely modest, close to zero, indicating that anxiety levels differ significantly between clusters. The test findings reveal that anxiety levels differ significantly amongst clusters.

CONCLUSIONS AND FUTURE WORK

The evaluation of student stress demonstrates that all 20 studied factors spanning psychological, physiological, environmental, academic, and social domains have a significant impact on stress levels. Statistical tests confirm the importance of these characteristics, with Gaussian Naive Bayes attaining the highest test accuracy and Decision Tree and Random Forest models outperforming in training accuracy. The Random Forest model's feature importance analysis uncovers key stress variables and guides targeted actions. Clustering with VAE and K-means indicated particular stress profiles, especially in clusters 1 and 4, which may benefit from targeted mental health treatment. Furthermore, association rules found substantial connections between traits such as sadness and anxiety, which improved our knowledge of stress dynamics. Future research should include expanding the dataset to enhance model accuracy, maximizing methods for machine learning, carrying out long-term research to track stress over time, developing interventions based on stress profiles, and examining additional factors such as social media influence to provide a more accurate picture of student stress.

REFERENCES

[1] Rayees, Ahmad & Dar, Rayees & Bilal, Ahmad. (2023). Impact of depression, stress and anxiety on academic achievement of Higher Secondary students. 8. 115-125. 10.31426/ijamsr.2019.2.1.222378.

[2] Yüksel Doğan, R., Metin, E.N. Exploring the Relationship between Mindfulness and Life Satisfaction in Adolescents: The Role of Social Competence and Self-Esteem. Child Ind Res 16, 1453–1473 (2023). https://doi.org/10.1007/s12187-023-10028-y

[3] Francisco J, Golshan F, Morrison TG, Mickleborough M. Stress and headaches in university students during the COVID-19 pandemic. PLoS One. 2023 Nov 22;18(11):e0288745. doi: 10.1371/journal.pone.0288745. PMID: 37992022; PMCID: PMC10664899.

[4] Schmickler JM, Blaschke S, Robbins R, Mess F. Determinants of Sleep Quality: A Cross-Sectional Study in University Students. Int J Environ Res Public Health. 2023 Jan 23;20(3):2019. doi: 10.3390/ijerph20032019. PMID: 36767422; PMCID: PMC9915447.

[5] Arregi, A., Vegas, O., Lertxundi, A. et al. Road traffic noise exposure and its impact on health: evidence from animal and human studies—chronic stress, inflammation, and oxidative stress as key components of the

complex downstream pathway underlying noise-induced non-auditory health effects. Environ Sci Pollut Res 31, 46820–46839 (2024). https://doi.org/10.1007/s11356-024-33973-9

[6] Kristensen SM, Larsen TMB, Urke HB, Danielsen AG. Academic Stress, Academic Self-efficacy, and Psychological Distress: A Moderated Mediation of Within-person Effects. J Youth Adolesc. 2023 Jul;52(7):1512-1529. doi: 10.1007/s10964-023-01770-1. Epub 2023 Mar 30. PMID: 36995523; PMCID: PMC10175374.

[7] Keane, Kevin & Evans, Retta & Orihuela, Catheryn & Mrug, Sylvie. (2023). Teacher–student relationships, stress, and psychosocial functioning during early adolescence. Psychology in the Schools. 60. 10.1002/pits.23020.

[8] Abdul Aziz NA, Baharudin NS, Alias NA. Association between Stress and Social Support Perceived among Undergraduate Health Sciences Student. Malays J Med Sci. 2023 Jun;30(3):176-183. Epub 2023 Jun 27. PMID: 37425381; PMCID: PMC10325121. doi: 10.21315/mjms2023.30.3.16

[9] Khairy, D., Alharbi, N., Amasha, M.A. et al. Prediction of student exam performance using data mining classification algorithms. Educ Inf Technol (2024). https://doi.org/10.1007/s10639-024-12619-w

[10] Genga L, Allodi L, Zannone N. Association Rule Mining Meets Regression Analysis: An Automated Approach to Unveil Systematic Biases in Decision-Making Processes. Journal of Cybersecurity and Privacy. 2022; 2(1):191-219. https://doi.org/10.3390/jcp2010011

[11] Mona Alhasani & Rita Orji (08 Jan 2024): Promoting Stress Managementamong Students in Higher Education: Evaluating the Effectiveness of a Persuasive Time Management Mobile App, International Journal of Human–Computer Interaction, DOI: https://doi.org/10.1080/10447318.2023.2297330

[12] Tran, M.A.Q., Vo-Thanh, T., Soliman, M. et al. Self-compassion, Mindfulness, Stress, and Self-esteem Among Vietnamese University Students: Psychological Well-being and Positive Emotion as Mediators. Mindfulness 13, 2574–2586 (2022). https://doi.org/10.1007/s12671-022-01980-x

[13] Spruill TM. Chronic psychosocial stress and hypertension. Curr Hypertens Rep. 2010 Feb;12(1):10-6. PMID: 20425153; PMCID: PMC3694268.doi: 10.1007/s11906-009-0084-8

[14] Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med. 2014 Jan;44(1):81-121. PMID: 24030837; PMCID: PMC3894304.doi: 10.1007/s40279-013-0090-5

[15] Hahad O, Prochaska JH, Daiber A, Muenzel T. Environmental Noise-Induced Effects on Stress Hormones, Oxidative Stress, and Vascular Dysfunction: Key Factors in the Relationship between Cerebrocardiovascular and Psychological Disorders. Oxid Med Cell Longev. 2019 Nov 11;2019:4623109. PMID: 31814877; PMCID: PMC6878772.doi: 10.1155/2019/4623109

[16] Jolly, Phillip & Kong, Dejun Tony & Kim, Kyoung Yong. (2020). Social Support at Work: An Integrative Review. Journal of Organizational Behavior. 42. 10.1002/job.2485.

[17] Abdullah, Siti & Shah, Norliana & Mohamad Idaris, Rohaiza. (2020). Stress and Its Relationship with the Academic Performance of Higher Institution Students. 2. 61-73.

[18] Tang Q, Lei G, Zhang Y, Shi H. The relation between mental health and career-related stress among prospective graduates in higher education stage during the COVID-19 pandemic: an evidence based on network analysis. Front Psychol. 2024 Jul 8;15:1381846. PMID: 39040967; PMCID: PMC11260747. doi:10.3389/fpsyg.2024.1381846

[19] Tang Q, Lei G, Zhang Y, Shi H. The relation between mental health and career-related stress among prospective graduates in higher education stage during the COVID-19 pandemic: an evidence based on network analysis. Front Psychol. 2024 Jul 8;15:1381846. PMID: 39040967; PMCID: PMC11260747. doi: 10.3389/fpsyg.2024.1381846

[20] Anand RV, Md AQ, Urooj S, Mohan S, Alawad MA, C. A. Enhancing Diagnostic Decision-Making: Ensemble Learning Techniques for Reliable Stress Level Classification. Diagnostics. 2023; 13(22):3455. https://doi.org/10.3390/diagnostics1322345

[21] Arya S, Anju, Ramli NA. Predicting the Stress level of students using Supervised Machine Learning and Artificial Neural Network (ANN). Indian Journal of Engineering, 2024, 21, e9ije1684 doi: https://doi.org/10.54905/disssi.v21i55.e9ije1684

[22] F. A. H. Ambreen and V. D. Jadhav, "Novel Model for Detection of Cyber-Aggressive Comments on Social Media Platforms a Review," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), Aurangabad, India, 2020, pp. 118-121, doi: 10.1109/ICSIDEMPC49020.2020.9299608.

[23] Jadhav, Varsha. (2025). MindIntent: A Chatbot Using SVM for Identifying Mental Health Intervention

Intentions. Journal of Information Systems Engineering and Management. 10. 535-542. 10.52783/jisem.v10i4s.564.

[24] /kaggle/input/student-stress-factors-a-comprehensive-analysis/StressLevelDataset.csv

[25] Fisher, R.A. (1992). Statistical Methods for Research Workers. In: Kotz, S., Johnson, N.L. (eds) Breakthroughs in Statistics. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4380-9_6

[26] Breiman, L., Friedman, J., Olshen, R.A., & Stone, C.J. (1984). Classification and Regression Trees (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315139470

[27] Chen X, Ishwaran H. Random forests for genomic data analysis. Genomics. 2012 Jun;99(6):323-9. doi: 10.1016/j.ygeno.2012.04.003 Epub 2012 Apr 21. PMID: 22546560; PMCID: PMC3387489.

[28] Li Z, Zhang Q, Zhao X. Performance analysis of K-nearest neighbor, support vector machine, and artificial neural network classifiers for driver drowsiness detection with different road geometries. International Journal of Distributed Sensor Networks. 2017;13(9). doi: 10.1177/1550147717733391

[29] Vikramkumar, & B, Vijaykumar & Tripathy, Trilochan. (2014). Bayes and Naive Bayes Classifier.

[30] Doersch, Carl. (2016). Tutorial on Variational Autoencoders.

[31] Shi, Xianjin & Wang, Wanwan & Zhang, Chongsheng. (2017). An Empirical Comparison of Latest Data Clustering Algorithms with State-of-the-Art. Indonesian Journal of Electrical Engineering and Computer Science. 5. 410. 10.11591/ijeecs.v5.i2.pp410-415.