Journal of Computer and Information Systems Ampera
Vol. 7, No. 3, September 2026 e-ISSN: 2775-2496
https://journal-computing.org/index.php/journal-cisa/index

Identification of Student Stress Levels Using XGBoost with Psychological, Physical, Academic, and Environmental Data

Muhamad Isa Firdaus[1], Hilmy Aliy Andra Putra[2], Muhammad Encep[3]
[1,2,3]Computer Science Study Program, Faculty of Computer Science, Djuanda University, Bogor, Indonesia
Email: [1]muhamadisafirdaus123@gmail.com, [2]hilmy.aliy@unida.ac.id, [3]ahmadpoac@unida.ac.id

Abstract

Student stress is a multidimensional problem influenced by psychological, physical, academic, and environmental factors. This study aims to identify student stress levels using the Extreme Gradient Boosting (XGBoost) algorithm, evaluate its classification performance, and determine the most influential variables. The study used the *StressLevelDataset* containing 1,100 observations, 20 predictor variables, and one target variable representing low, moderate, and high stress. The research stages followed data collection, data understanding, preprocessing, stratified data splitting, baseline modeling, Grid Search hyperparameter tuning, evaluation, and feature importance analysis. The tuned XGBoost model achieved 86.82% accuracy, 86.94% precision, 86.85% recall, and 86.79% F1-score on the test set. Hyperparameter tuning did not materially increase test accuracy, but reduced the training-test accuracy gap from 13.18 to 6.82 percentage points, indicating improved generalization and reduced overfitting. Feature importance identified *sleep_quality* (32.6%), *academic_performance* (29.8%), and *blood_pressure* (19.2%) as the three most influential variables. The results demonstrate that XGBoost can provide a stable data-driven approach for identifying student stress levels.

Keywords: stress, XGBoost, classification, machine learning, feature importance

1. INTRODUCTION

Stress is a non-specific response of the body to demands or pressure originating from physical and psychological sources. Among university students, stress is closely related to the transition into adulthood and the simultaneous demands of academic, social, and personal life. Heavy study loads, examination pressure, and poor time management are among the factors associated with academic stress, while unmanaged stress can also be accompanied by anxiety and depressive symptoms [1].

Student stress is multidimensional. Previous work using machine learning has reported that psychological and physical factors can become dominant predictors of student stress, while sleep quality, academic conditions, social pressure, and living conditions may also contribute to the overall stress level [2]. Because these factors interact and may exhibit nonlinear patterns, simple descriptive statistics or single-factor approaches may be insufficient for capturing the classification patterns contained in multidimensional student data.

The use of machine learning provides an opportunity to build a more systematic and objective classification model. In particular, XGBoost is an ensemble boosting algorithm that constructs decision trees sequentially and incorporates regularization mechanisms to control model complexity [3]. Research on student stress has shown that XGBoost can achieve competitive performance. Yeler and Sürücü reported 86.8% accuracy for XGBoost on a similar student-stress dataset, while Gedam et al. reported 96.17% accuracy when analyzing physiological data collected using wearable technology [4], [5]. Other studies have explored hybrid models, feature selection, dimensionality reduction, ensemble voting, stacking, or Random Forest optimization [6], [7].

Although these studies demonstrate the usefulness of machine learning for student stress classification, there remains an opportunity to examine the contribution of a single XGBoost model when all multidimensional psychological, physical, academic, and environmental variables are retained. This study therefore focuses on optimizing XGBoost directly through hyperparameter tuning without combining it with another classifier, feature-selection method, or ensemble scheme. The research questions concern how XGBoost can be applied to the *StressLevelDataset*, how well the tuned model performs, and which variables contribute most strongly to the classification result.

2. METHODS

2.1 Dataset

The study used the StressLevelDataset obtained from Kaggle. The dataset contains survey data describing student conditions across psychological, physical, academic, and environmental dimensions. It consists of 1,100 observations, 20 predictor variables, and one target variable, *stress_level*, with three classes: low, moderate, and high stress. The variables are numerical and primarily represent ordinal questionnaire responses, which reduces the need for categorical encoding [2].

2.2 Research Procedure

The research followed the *CRISP-DM* oriented workflow used in the thesis: data collection, data understanding, preprocessing, data splitting, model training, hyperparameter tuning, and evaluation [8], [9]. Data understanding included structural inspection, descriptive statistics, and Pearson correlation analysis. Preprocessing consisted of checking missing values, duplicate records, and outliers using the Interquartile Range (IQR). Because the predictors were already numerical, categorical encoding was not required, and feature scaling was not considered necessary for the tree-based XGBoost model [10], [3].

2.3 Data Splitting and XGBoost Modeling

The cleaned dataset was divided into training and testing sets using an 80:20 stratified split. The resulting sets contained 880 training observations and 220 testing observations, while preserving the class proportions. Two XGBoost models were developed. The first was a baseline model using the default configuration. The second was optimized using Grid Search with 5-fold cross-validation. The search space consisted of *n_estimators* = {100, 200, 300}, *max_depth* = {3, 4, 5, 6}, *learning_rate* = {0.01, 0.05, 0.1, 0.2}, and *subsample* = {0.7, 0.8, 1.0}. The 144 parameter combinations were evaluated across five folds, resulting in 720 model-training runs. The best configuration was selected according to the cross-validation result [11], [12].

2.4 Evaluation and Feature Importance

Model performance was evaluated using accuracy, precision, recall, and F1-score. A confusion matrix was used to inspect class-level prediction patterns. The final analysis also examined feature importance to identify the variables contributing most strongly to the classification decision. The

Vol. 7, No. 3, September 2026 e-ISSN: **2775-2496**
https://journal-computing.org/index.php/journal-cisa/index

thesis reports the resulting importance values for the three dominant variables as *sleep_quality*, *academic_performance*, and *blood_pressure*.

Table 1. Dataset and experimental configuration

Item	Configuration
Dataset	*StressLevelDataset*
Observations	1,100
Predictors	20 variables
Target	*stress_level* (Low, Moderate, High)
Training/testing split	80% / 20% (stratified)
Training/testing size	880 / 220
Baseline	XGBoost default configuration
Tuning	Grid Search + 5-fold cross-validation

3. RESULTS AND DISCUSSION

3.1 Data Understanding and Preprocessing

The dataset contained 1,100 rows and 21 columns, consisting of 20 predictors and one target. Descriptive analysis showed that the variables remained within their defined measurement ranges. The stress-level classes were relatively balanced: 373 observations (33.9%) were low stress, 358 (32.5%) were moderate stress, and 369 (33.5%) were high stress. No missing values or duplicate rows were found. IQR-based screening identified observations outside the statistical IQR limits for *noise_level* (173), *study_load* (165), and *living_conditions* (62); however, these observations remained within valid Likert-scale ranges and were therefore retained.

Pearson correlation analysis showed that most predictors had moderate to strong relationships with *stress_level.* The strongest correlations were *self_esteem* (-0.756), bullying (0.751), and *sleep_quality* (-0.749), whereas *blood_pressure* had the weakest reported correlation (0.394). The direction of the correlations was also consistent with the expected interpretation of several variables: higher *self_esteem*, sleep_quality, *safety*, and *academic_performance* were associated with lower stress, while higher *bullying*, *anxiety*, and *depression* were associated with higher stress.

Vol. 7, No. 3, September 2026 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

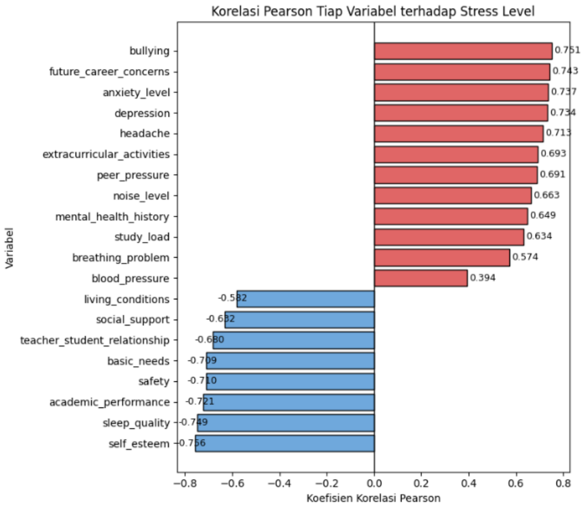

Figure 1. Pearson correlation of predictor variables with *stress_level*

3.2 Hyperparameter Tuning

Grid Search selected *n_estimators* = 100, *max_depth* = 3, *learning_rate* = 0.01, and *subsample* = 1.0 as the best configuration. This configuration achieved a mean cross-validation accuracy of 89.66%. The combination of a low learning rate and shallow trees indicates that a more conservative model configuration was suitable for the relatively limited dataset.

Table 2. Grid Search space and selected XGBoost configuration

Parameter	Search values	Best value
n_estimators	100, 200, 300	100
max_depth	3, 4, 5, 6	3
learning_rate	0.01, 0.05, 0.1, 0.2	0.01
subsample	0.7, 0.8, 1.0	1.0
CV accuracy	—	89.66%

3.3 Model Performance

The baseline model reached 100.00% on the training set for accuracy, precision, recall, and F1-score, but its test accuracy was 86.82%, indicating a 13.18-percentage-point training-test gap. After tuning, the training metrics decreased to 93.64%, while the test accuracy remained 86.82%. Test precision increased from 86.82% to 86.94%, recall from 86.84% to

86.85%, and F1-score changed slightly from 86.80% to 86.79%. Thus, tuning did not materially improve raw test accuracy, but it reduced the training-test accuracy gap to 6.82 percentage points, approximately 48% smaller than the baseline gap.

Table 3. Baseline and tuned XGBoost performance

Model	Dataset	Accuracy	Precision	Recall	F1-score
XGBoost Baseline	Training	100.00%	100.00%	100.00%	100.00%
XGBoost Baseline	Testing	86.82%	86.82%	86.84%	86.80%
XGBoost Tuned	Training	93.64%	93.64%	93.64%	93.64%
XGBoost Tuned	Testing	86.82%	86.94%	86.85%	86.79%

The tuned model classified 191 of 220 test observations correctly, corresponding to an error rate of 13.18%. Class-level results were relatively balanced: precision was 0.90, 0.87, and 0.85 for low, moderate, and high stress, respectively; recall was 0.81, 0.90, and 0.89; and F1-score was 0.85, 0.88, and 0.87. The errors were concentrated mainly between adjacent stress categories, such as low versus moderate and moderate versus high, while confusion between low and high stress was comparatively limited. This pattern is consistent with the ordinal nature of the stress-level categories.

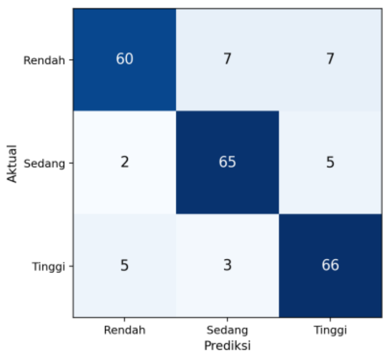

Figure 2. Confusion matrix of the tuned XGBoost model on the test set

Vol. 7, No. 3, September 2026 e-ISSN: **2775-2496**
https://journal-computing.org/index.php/journal-cisa/index

3.4 Feature Importance

Feature importance analysis identified *sleep_quality* (32.6%), *academic_performance* (29.8%), and *blood_pressure* (19.2%) as the three most influential variables. Together, these variables contributed more than 80% of the reported feature importance. The result suggests that physical factors, particularly sleep quality and blood pressure, together with academic performance, were the dominant predictors in this dataset. This finding is compatible with previous research reporting psychological and physical factors as important predictors of student stress, although the precise ranking of variables can vary across datasets and models [2].

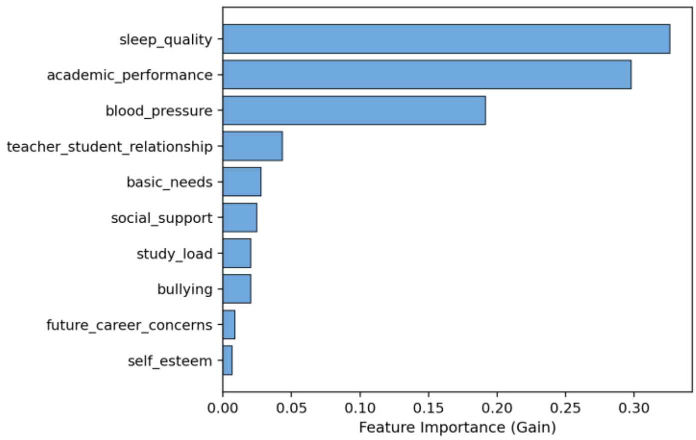

Figure 3. Top feature importance values in the tuned XGBoost model

3.5 Discussion

The results show that XGBoost can classify student stress levels effectively using multidimensional survey variables. The test metrics were all above 86% and differed by less than 0.2 percentage points, indicating balanced performance across the evaluation measures. The principal benefit of tuning in this experiment was not a higher test accuracy, but a substantial reduction in the training-test performance gap. This distinction is important because a model that performs perfectly on training data but substantially worse on unseen data can indicate overfitting. The tuned configuration therefore provides a more conservative and generalizable model for this dataset.

Vol. 7, No. 3, September 2026 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

Compared with previous studies, the test accuracy of 86.82% is close to the 86.8% reported for XGBoost by Yeler and Sürücü on a similar student-stress dataset [6]. The present study differs in emphasizing direct optimization of a single XGBoost model rather than combining XGBoost with feature selection, dimensionality reduction, or ensemble schemes. The feature-importance result also provides a practical interpretation of the model by highlighting sleep quality, academic performance, and blood pressure as the strongest contributors in the analyzed data.

4. CONCLUSION

This study demonstrated that XGBoost can be used to identify student stress levels from psychological, physical, academic, and environmental variables in the *StressLevelDataset*. The tuned model achieved 86.82% accuracy, 86.94% precision, 86.85% recall, and 86.79% F1-score on 220 unseen test observations. Although Grid Search did not materially improve test accuracy compared with the default model, it reduced the training-test accuracy gap from 13.18 to 6.82 percentage points, indicating lower overfitting and improved model stability. Feature importance showed that *sleep_quality, academic_performance, and blood_*pressure were the dominant predictors, contributing 32.6%, 29.8%, and 19.2%, respectively. Future research should consider larger and more diverse datasets, comparisons with other algorithms, broader hyperparameter search spaces, and longitudinal or derived features to improve the robustness and interpretability of student stress classification.

REFERENCES

[1] H. Selye, "The Stress Of Life," 1976.

[2] R. de Filippis and A. Al Foysal, "Comprehensive analysis of stress factors affecting students: a machine learning approach," *Discov. Artif. Intell.*, vol. 4, no. 1, 2024.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016.

[4] K. Yeler and S. Sürücü, "Araştırma Makalesi," pp. 210–214, 2025.

[5] S. Gedam, S. Dutta, and R. Jha, "Analyzing mental stress in Indian students through advanced machine learning and wearable technologies," *Sci. Rep.*, vol. 15, no. 1, pp. 1–18, 2025.

[6] V. Doma, A. A. Almisreb, E. Yaman, S. Amanzholova, and N. Ismail, "Classification of Student Stress Levels Using a Hybrid Machine Learning Model," pp. 77–104, 2025.

[7] S. D. Amalia, M. A. Barata, and P. E. Yuwita, "Optimization of Random Forest Algorithm with Backward Elimination Method in Classification of Academic Stress Levels," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 633–641, 2025.

[8] C. Pete *et al.*, "Crisp-Dm 1.0," *Cris. Consort.*, p. 76, 2000.

[9] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021.

[10] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Anal.*, vol. 1, no. 1, pp. 1–22, 2016.

[11] R. Kohavi and G. H. John, "Automatic Parameter Selection by Minimizing Estimated Error," *Proc. 12th Int. Conf. Mach. Learn. ICML 1995*, pp. 304–312, 1995.

[12] G. James, D. Witten, T. Hastie, and R. Tibshirani, *Novel Approaches for Designing 5-O-Ester Prodrugs of 3-Azido-2,3-dideoxythymidine (AZT).*, vol. 7, no. 10. 2013.