Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 3, November 2025, pp. 579 – 591
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Prediction of Student Academic Stress Levels Using the Decision Tree Algorithm and Particle Swarm Optimization

[1*]Syifa Ghina Dzakiyyah, [2]Deni Mahdiana
[1]Department of Computer Science (Master's Program), Faculty of Information Technology
Universitas Budi Luhur, Jakarta, Indonesia
[2]Department of Computer Science, Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia
Email: [1]syifaghina2505@gmail.com, [2]deni.mahdiana@budiluhur.ac.id

Article Info

Article history:
Received Jul 30th, 2025
Revised Sep 14th, 2025
Accepted Sep 25th, 2025

Keyword:
Academic Stress
Decision Tree
Machine Learning
Particle Swarm Optimization
Perception of Academic Stress

ABSTRACT

Academic stress was recognized as a major challenge for university students because it negatively affected learning outcomes, mental health, and overall well-being. The purpose of this research was to develop and validate a predictive model of student academic stress levels and to evaluate whether optimization techniques improved the performance of a baseline classifier. Data were collected from 413 active students of Universitas Sapta Mandiri from the 2022 and 2023 cohorts using the Perception of Academic Stress (PAS) scale, which consisted of 18 indicators, together with demographic, academic, and psychosocial attributes. The Decision Tree (DT) algorithm was selected for its interpretability and transparency in multi-class classification. To improve generalization, its parameters were optimized using Particle Swarm Optimization (PSO) with 10 particles and 20 iterations. The baseline model achieved an accuracy of 93 percent, with the highest recall observed in the low-stress group. After optimization, the accuracy increased to 95 percent, and the recall for the high-stress group reached 0.96, indicating greater sensitivity to students at risk. These results confirmed that the research objectives were achieved, as the integration of DT with PSO enhanced both accuracy and class balance. The proposed model was consistent with the intended purpose of supporting early detection and timely academic and psychological interventions in higher education institutions.

Corresponding Author:
Syifa Ghina Dzakiyyah,
Department of Computer Science (Master's Program), Faculty of Information Technology
Universitas Budi Luhur,
Jalan Ciledug Raya Blok Haji Haisin No.22, RT.10/ RW.3, Petukangan Utara, Pesanggrahan,
Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12260, Indonesia.
Email: syifaghina2505@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.38081

1. INTRODUCTION

Academic stress has become an increasingly prevalent phenomenon among university students, particularly in higher education environments that demand rapid adaptation to academic workloads, social pressures, and complex time management [1], [2]. Such conditions may trigger mental health issues, including anxiety, depression, and even burnout [1], [2], [18]. Data from the American College Health Association indicate that 45% of students worldwide experience high levels of stress. In Indonesia, a survey conducted by the Indonesian Clinical Psychologists Association reported that more than 50% of students exhibit stress symptoms during their academic journey [19]. Moreover, an internal survey conducted at Universitas Sapta Mandiri in 2025 revealed that most students experienced moderate to high levels of stress, particularly during examinations and assignment deadlines.

Although various efforts had been made to identify stress levels, most approaches remained manual and subjective, such as observation, interviews, and direct surveys [1], [3]. These methods had limitations in terms of efficiency, accuracy, and their ability to support early detection [20]. The urgency of this research lay in the fact that academic stress had become a widespread phenomenon, with surveys showing that more than half of students in Indonesia experienced moderate to high stress levels, often leading to decreased academic performance and psychological problems. Early identification using reliable computational tools was therefore essential, not only to prevent the escalation of stress into severe conditions but also to enable universities to design timely interventions in counseling, workload adjustment, and support services.

In addressing this problem, machine learning offered a promising solution because of its capacity to analyze diverse variables simultaneously and generate reproducible results. Among various algorithms, the Decision Tree (DT) was chosen because it produced interpretable rules that could be directly understood by academic staff and policy makers, making it highly suitable for educational settings [3], [4]. Previous studies demonstrated the effectiveness of DT in classifying academic outcomes and student stress [3], [4]. However, DT models were often vulnerable to overfitting, which reduced their generalization to new data [4]. To overcome this issue, optimization methods such as Particle Swarm Optimization (PSO) have been successfully applied in related domains to improve accuracy and parameter tuning [5], [8]. For example, Hendra et al. [4] showed that PSO improved DT performance in predicting graduation, while Elham and Mehrabi [8] applied DT–PSO to mental health diagnosis with promising results. Building on these findings, this study combined DT with PSO to ensure higher accuracy and balanced detection across stress categories, particularly the high-stress group that required urgent attention.

Several previous studies have demonstrated the effectiveness of DT in predicting student stress [3], [4], but few have integrated optimization techniques such as PSO to enhance model performance [4], [5]. Pankajavalli and Karthick, for example, showed that PSO improved the accuracy of Support Vector Machine models in stress detection based on physiological signals [5]. Nevertheless, its application to stress prediction using questionnaire-based academic, social, and psychological features remains relatively underexplored [5]. This research aims to fill this gap while strengthening the scientific contribution to the development of data mining–based early stress detection systems [3], [4], [5].

While existing studies have applied various machine learning algorithms, such as Support Vector Machine, Naïve Bayes, K-Nearest Neighbor, and ensemble methods for stress prediction [9],[10], [11], [12], most of them focused on binary classification or relied on physiological signals rather than academic and psychosocial survey data. Moreover, only a few studies explicitly integrated DT with metaheuristic optimization techniques. For instance, Hendra et al. [4] demonstrated that PSO improved DT performance in predicting academic outcomes, and Elham and Mehrabi [8] reported promising results for mental health prediction using a DT–PSO approach. However, little attention has been given to the application of DT–PSO for predicting academic stress based on the Perception of Academic Stress (PAS) questionnaire in a multi-class setting. The novelty of this research lies in combining an interpretable algorithm, namely DT, with PSO-based parameter tuning to enhance both accuracy and recall across stress categories, particularly the high-stress group. In addition, the deployment of the optimized model in a web-based prototype represents a practical contribution, bridging the gap between theoretical modeling and real-world application in higher education environments.

The objective of this study is to develop a predictive model of student academic stress levels by applying a DT algorithm enhanced through PSO [4], [5]. The model is expected to improve classification accuracy and balance across stress categories, particularly in detecting students at high risk of stress more effectively and at an earlier stage [4], [5].

2. LITERATURE REVIEW

2.1. Data Mining and Prediction

Data mining is an important process in extracting hidden knowledge from large datasets through various analytical techniques. In the context of classification, algorithms such as Decision Tree (DTs) and naive Bayes are often used to build predictive models based on both survey and sensor data. One systematic approach in data mining projects is the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which includes six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [13].

In the context of higher education, data mining plays a significant role in evaluating academic performance, detecting dropout risks, and predicting students' academic stress levels using attributes such as Grade Point Average (GPA), sleep quality, and workload [1], [2], [3]. To enhance predictive accuracy, the integration of optimization techniques such as PSO has been increasingly explored [4], [5], [8].

2.2. Decision Tree (DT) and Particle Swarm Optimization (PSO)

The DT algorithm is one of the machine learning methods used to construct predictive models based on decisions represented in a tree-like structure [6]. This algorithm works by recursively partitioning a dataset into several subgroups according to the most significant attributes until a final decision is reached. Each branch of the DT represents a specific rule or condition, while each terminal node or leaf corresponds to a final outcome or decision. DTs are frequently applied in classification and regression tasks due to their ability to separate data effectively through a series of simple yet powerful rules [7]

A DT is composed of two main types of nodes, namely decision nodes and leaf nodes. A decision node represents a feature from the dataset and is used to make a decision, while a leaf node represents the final output of that decision and does not contain any further branches. The basic structure of a DT is illustrated in Figure 1.

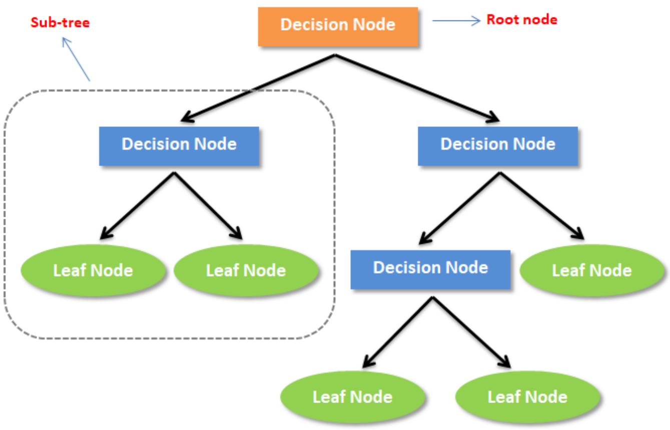

Figure 1. Flowchart Decision Tree

Figure 1 depicts the fundamental structure of a DT, which includes core components such as the root node, decision nodes, and leaf nodes. The root node represents the complete dataset and determines the most suitable attribute for data partitioning using criteria like information gain or the Gini index.

A DT is a rule-based classification algorithm that partitions data using metrics such as information gain or the Gini index [5]. Although DTs are effective for classification tasks, they are vulnerable to overfitting. Therefore, optimization techniques like PSO are applied to tune parameters such as maximum depth and minimum samples split to enhance model performance [8].

PSO is a population-based optimization method that simulates the collective behavior of swarms, where each particle represents a candidate solution and is updated according to its own experience (pbest) as well as the group's best experience (gbest). The combination of DT and PSO has been proven to enhance accuracy in various classification tasks [21], including graduation prediction [4] and mental health diagnosis [8].

The PSO procedure in Figure 2 can be outlined in the following steps:

1. Start: The PSO process begins.
2. Initialize algorithm constants: The algorithm begins by setting its constants, which include parameters such as inertia weight (w), acceleration coefficients (c1 and c2), and, if required, velocity or position constraints.
3. Set t = 1: The initial iteration *t* is set to 1.
4. Initialize particles with random positions and velocities: several particles are generated within the search space, each assigned random positions and velocities, with every particle representing a potential solution.
5. Evaluate the fitness function for each particle: the fitness of each particle is determined using the objective function that needs to be optimized. Fitness indicates the quality of the solution represented by the particle's position.
6. Is the current fitness better than the previous pbest? The particle's current fitness is compared to its previous personal best (pbest) to determine if it represents an improvement.
   a. If Yes, the previous pbest is replaced with the current fitness value.
   b. If No, the previous pbest value is retained.

7. Set the best pbest as gbest: The best personal best (pbest) among all particles is selected as the global best (gbest), which serves as a guide for the swarm in finding the optimal solution.
8. Update particle velocity:
   The velocity of each particle is revised according to the following equation.:

$$vi(t + 1) = w \cdot vi(t) + c1 \cdot r1 \cdot (pbest - xi(t)) + c2 \cdot r2 \cdot (gbest - xi(t)) \quad (1)$$

9. Update particle position:
   Each particle's position is then updated according to the following equation:

$$xi(t + 1) = xi(t) + vi(t + 1) \quad (2)$$

10. Are the stopping criteria satisfied or has the maximum iteration limit been achieved ?: Checking whether the stopping criteria have been met or the maximum number of iterations has been reached.
    a. If Yes, the algorithm terminates, and the output gbest represents the best solution obtained.
    b. If No, the iteration counter is incremented (t = t + 1), and the algorithm returns to the fitness evaluation step for the next iteration.

11. End: The algorithm terminates once the stopping condition
    Overall, the diagram illustrates an iterative process in which particles move within the search space, interacting through pbest and gbest to converge toward the optimal solution.

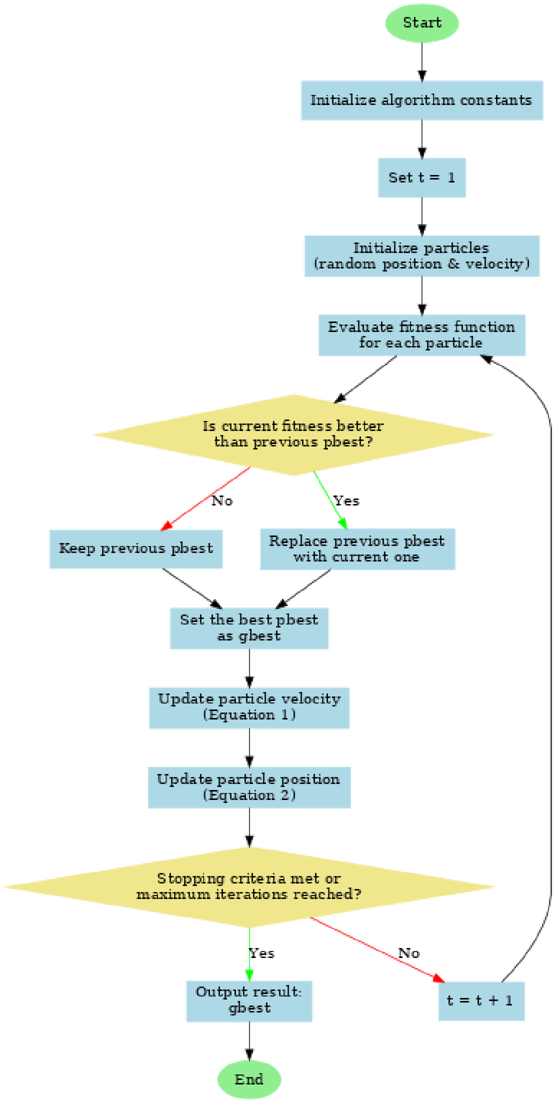

Figure 2. Flowchart PSO

3. RESEARCH METHOD

This research applies the CRISP-DM [13], which is structured into six key phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The selection of this method is based on its ability to provide a systematic, flexible, and adaptable workflow for the development of data-driven classification models.

3.1. Business Understanding

The main problem in this study is the high level of stress arising among students as a result of academic, social, and psychological demands. A predictive model was developed to classify student stress levels based on demographic, academic, and psychosocial data. The DT algorithm was selected because it is transparent, easy to interpret, and effective for multi-class classification [8]. To further improve model performance, parameter optimization was conducted using PSO [5], a metaheuristic algorithm that simulates swarm behavior to search for optimal solutions [5], [8].

Data were collected from 413 active students of Universitas Sapta Mandiri from the 2022 and 2023 cohorts using a purposive sampling method [14]. The target variable was defined based on the total score of 18 items from the PAS scale [15], which was subsequently categorized into three levels: low stress (<48), moderate stress (48–60), and high stress (>60).

3.2. Data Understanding

The collected data consisted of three groups of variables: (1) demographic (gender, residence, study program, semester), (2) academic (GPA and Semester Performance Index/SPI), and (3) psychosocial (sleep quality, frequency of social interaction, and weekly independent study hours). The selection of these variables was based on previous studies that demonstrated their significant influence on students' stress conditions [1], [2], [3].

The data distribution, variable correlations, and possible outliers were examined through descriptive statistics and visual tools such as heatmaps and bar charts. The initial structure of the dataset used is presented in Figure 3.

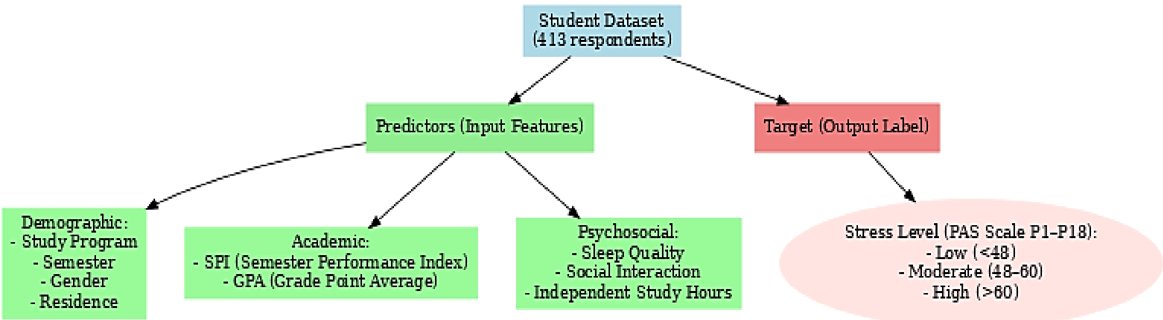

Figure 3. Structure of The Student Dataset

3.3. Data Preparation

The data preparation steps included removing duplicate records, handling missing values, and transforming categorical features using label encoding. The total PAS score was calculated by summing the values of 18 items [15]. The target labels were classified into three stress categories and encoded as numerical variables. The dataset was divided into training and testing subsets in an 80:20 ratio, employing stratified sampling to preserve the balance of target class distribution. The entire data preparation process is illustrated in Figure. 4.

3.4. Modelling

Modeling began by constructing a baseline model using the DT Classifier algorithm [8]. Subsequently, hyperparameters such as max_depth, min_samples_split, and criterion were optimized using PSO [5], [8], [22] . PSO functions by generating a set of particles in the search space and iteratively updating their positions according to individual and collective experiences to determine the optimal parameter configuration.

The fitness value in the optimization process was calculated using the average macro F1-score from 5-fold cross-validation . The flowchart of the modeling process is presented in Figure 5.

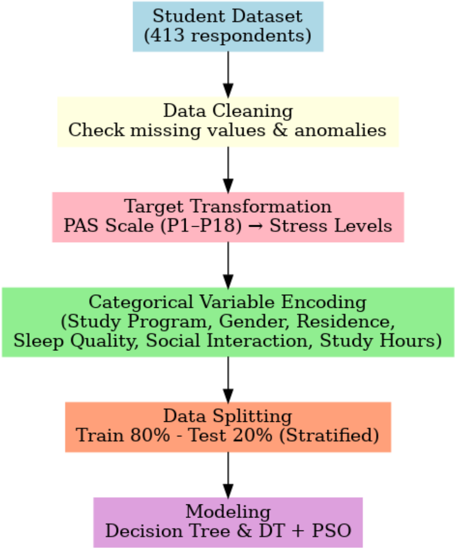

Figure 4. Data Preprocessing and Modeling Steps

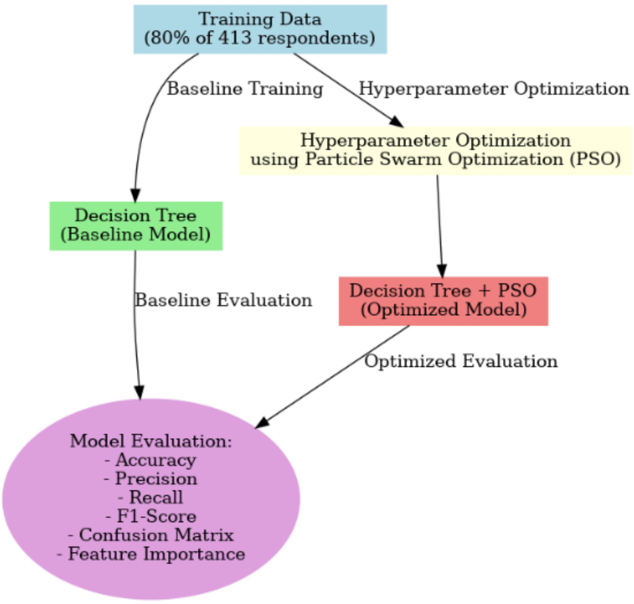

Figure 5. Modeling and Evaluation Flowchart

3.5. Evaluation

The model was assessed using classification metrics, including accuracy, precision, recall, and F1-score. Additionally, a confusion matrix was used to evaluate the distribution of classifications across target classes. An analysis of feature importance was performed to determine which variables had the greatest influence on the model's predictions [8]. A performance comparison was conducted between the baseline model and the PSO-optimized model to demonstrate the effectiveness of the parameter tuning process.

3.6. Implementation

The optimized model was deployed in a web-based interface built with the Streamlit framework [16], [17]. The system provides three main functions: data description, model testing, and manual prediction based on user input. The visualizations include classification results, the confusion matrix, the DT, and feature contribution graphs.

4. RESULTS AND ANALYSIS

This study produced two main models, namely the baseline DT and the DT optimized using PSO [8], [5]. All experiments were conducted using a dataset collected from a survey of 413 students from

Universitas Sapta Mandiri, comprising the 2022 and 2023 cohorts. The dataset consisted of 18 indicators from the PAS scale [15], combined with demographic, academic, and psychosocial variables.

4.1. General Overview

The research process began with data preprocessing to ensure the quality of the dataset. Missing values were cleaned, and categorical variables were encoded into numerical form. Subsequently, a labeling stage was carried out to define the target variable of the study, namely the students' stress levels. The labeling was based on the total score of the PAS scale [15], obtained by summing the values of 18 items. Following the categorization proposed by Bedewy and Gabriel (2015) [15], stress levels were classified into three groups: low stress (<48), indicating minimal academic strain with limited disruption to daily functioning; moderate stress (48–60), reflecting recurring strain that affects time management, study efficiency, and psychological balance; and high stress (>60), representing pervasive symptoms such as persistent worry, sleep disturbance, and academic overload, which often require professional counseling or workload adjustment. The initial distribution of the data showed that the three categories were relatively balanced, with each group representing approximately one-third of the respondents. Figure 6 presents this distribution, confirming that the dataset was suitable for multi-class classification.

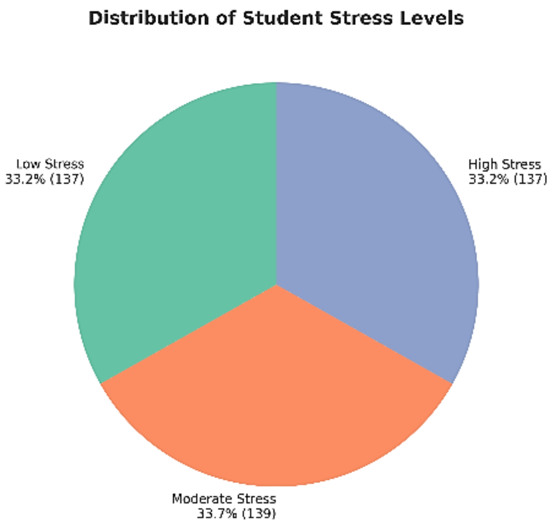

Figure 6. Distribution of Student Stress Levels

In the modeling stage, the DT algorithm was employed as the baseline due to its capability to process mixed data and provide visual interpretation in the form of a DT [8]. The baseline model was trained using default parameters, while the second model was optimized with PSO [5], [8] to determine the parameter combination that maximizes performance. PSO was selected for its ability to explore the solution space globally and avoid local optima, which are often encountered in conventional optimization algorithms [5].

To verify the effectiveness of the proposed algorithm, this study compared the performance of the DT with two widely recognized classification methods, Naïve Bayes and K-Nearest Neighbors (KNN) (see Table 1) [9], [23]. The experimental findings revealed that the DT obtained the best performance, reaching 93% in accuracy, precision, recall, and F1-score. In contrast, Naïve Bayes achieved an accuracy of only 84.34%, while KNN reached 86.75% [24]. Although KNN recorded the highest AUC value, the difference was not substantial when compared to the DT.

Table 1. Comparison of Algorithms

Evaluation Aspect	Decision Tree	Naive Bayes	KNN
Final Accuracy	93%	85%	87%
Precision	93%	85%	87%
Recall	93%	84%	87%
F1-Score	93%	84%	87%
AUC	94,58%	94,72%	96,66%

This difference reinforces the superiority of the DT compared to probabilistic and instance-based learning models, thereby justifying its selection as the basis for modeling.

4.2. Baseline Model Results

The evaluation results for the baseline DT model are presented in Table 2.

Table 2. Classification Report Model DT

Category	Precision	Recall	F1-Score	Support
Low Stress	0,87	0,96	0,92	28
Moderate Stress	1,00	0,89	0,94	28
High Stress	0,93	0,93	0,93	27
Accuracy			0,93	83
Macro Avg	0,93	0,93	0,93	
Weighted Avg	0,93	0,93	0,93	

This model achieved an accuracy of 93%, with the best performance observed in the low-stress category, which obtained a recall of 0.96. This indicates that almost all students with low stress were correctly predicted. However, the recall for the moderate-stress category was only 0.89, suggesting that a number of students who should have been classified in this category were misclassified, primarily into the low-stress group. Nevertheless, the precision and F1-score across all three categories remained high (≥0.92), indicating that the baseline DT model performed fairly well, although there is still room for improvement, particularly in detecting students with moderate stress levels.

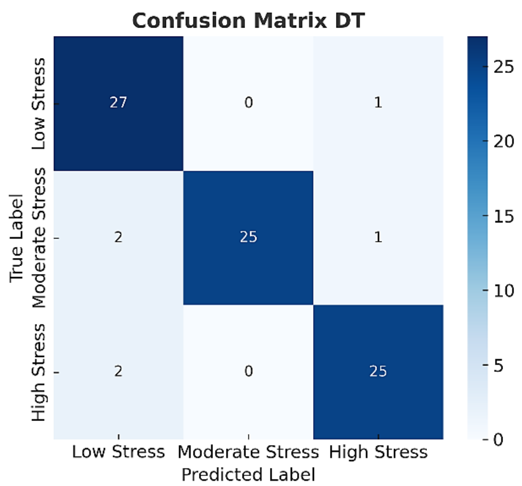

Figure 7. Confusion Matrix Decision Tree

Figure 7 illustrates the confusion matrix of the baseline DT model. Most predictions are located along the main diagonal, indicating fairly good classification performance, with the majority of students correctly classified into their actual categories. In particular, the low-stress category achieved the best result with a recall of 0.96. However, some misclassifications were still observed in the moderate- and high-stress categories, although the numbers were relatively small. To provide further clarity, Table 3 presents the detailed breakdown of TP, FP, FN, and TN for each category.

Table 3. Breakdown Confusion Matrix Decision Tree

Category	TP	FP	FN	TN
Low Stress	27	4	1	51
Moderate Stress	25	0	3	55
High Stress	25	2	2	54

In a multiclass setting, the confusion matrix is structured as a 3×3 table, with rows indicating the true labels and columns showing the model's predicted labels. True Positive (TP) is indicated by the values on the main diagonal, namely the number of students correctly predicted according to their actual condition. False Negative (FN) occurs when students of a given category are misclassified into another category, while False Positive (FP) arises when students from other categories are incorrectly classified into the category under consideration. True Negative (TN) refers to all cases from other categories that are correctly identified as not belonging to the category being analyzed.

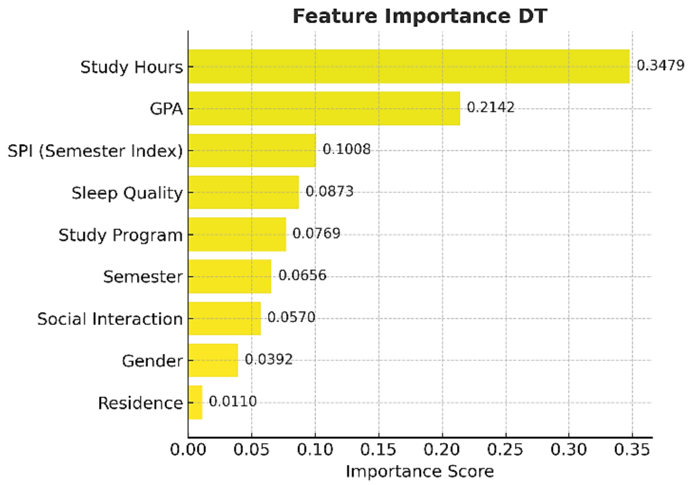

Figure 8. Feature Importance Decision Tree

Figure 8 presents the feature importance analysis of the DT model. The results indicate that independent study hours contributed the most to predicting student stress levels, with a weight exceeding 0.34. This finding is consistent with the literature, which emphasizes that academic workload significantly influences stress levels [1], [2], [3].

The next influential variable was the GPA, with a weight of 0.21, reaffirming the strong relationship between academic achievement and stress levels. This was followed by sleep quality and study program, each contributing considerably. These findings are consistent with previous literature, which highlights that academic and psychosocial factors such as performance, sleep quality, and the learning environment are key determinants of student stress levels.

Meanwhile, the variables of residence and gender were found to have the lowest contribution. Nevertheless, they remain relevant as they may indirectly influence psychological conditions, for instance, through the conduciveness of the living environment or the social interactions within the campus setting.

Overall, this feature importance visualization demonstrates that the DT model not only provides predictive outcomes but also offers insights into the dominant factors influencing student stress. These findings can provide a basis for formulating more focused intervention strategies.

The ranking of features in Figure 8 was determined using the Gini importance, also referred to as mean decrease in impurity. Each time a feature was used to split a node, the decrease in impurity was recorded and summed across the entire tree. These values were then normalized so that the total contribution equaled one, and the features were ordered in descending order. Consequently, independent study hours appeared at the top because it provided the greatest overall reduction in impurity, followed by GPA and sleep quality, while residence and gender showed the smallest contributions.

4.3. Baseline Model Results with PSO

The optimization process was carried out by forming 10 particles that moved within the search space for 20 iterations. The optimization results indicated the best parameter combination of max_depth = 5, min_samples_split = 5, and criterion = entropy. After optimization using PSO, the evaluation results are presented in Table 4. The optimized model achieved an accuracy of 95%, representing a 2% improvement compared to the baseline. The greatest enhancement was found in the recall value of the high-stress category, which increased to 0.96, indicating that the model became more sensitive in identifying students at high risk. This finding is crucial, as this group requires earlier interventions to prevent negative impacts on mental health and academic performance.

Table 4. Classification Decision Tree with PSO

Category	Precision	Recall	F1-Score	Support
Low Stress	1,00	0,96	0,98	28
Moderate Stress	0,96	0,93	0,95	28
High Stress	0,90	0,96	0,93	27
Accuracy			0,95	83
Macro Avg	0,95	0,95	0,95	
Weighted Avg	0,95	0,95	0,95	

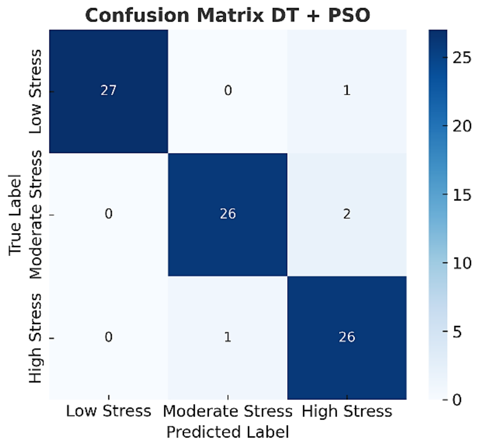

Figure 9. Confusion Matrix DT with PSO

Figure 9 shows the confusion matrix after optimization. Most predictions lie along the main diagonal, with fewer misclassifications compared to the baseline. For instance, almost all high-stress students were correctly predicted, with only one case misclassified as moderate stress. In the multiclass context, this pattern demonstrates that TP values dominate across all three categories, while the number of FN decreased significantly (see Table 5).

Table 5. Classification Decision Tree with PSO

Category	TP	FP	FN	TN
Low Stress	27	0	1	55
Moderate Stress	26	1	2	54
High Stress	26	2	1	54

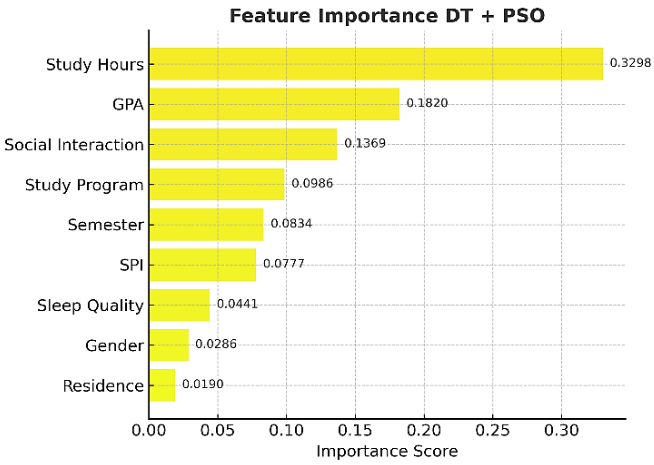

Figure 10. Feature Importance DT with PSO

The feature importance analysis revealed a more balanced distribution of weights. Independent study hours remained the most dominant factor (0.33), followed by GPA (0.18) and social interaction (0.13). The PSO optimization strengthened the role of academic and psychosocial variables in the model while minimizing the contribution of less influential variables such as residence. These results confirm that PSO successfully structured a more optimal DT to improve accuracy [8].

Similarly, the ranking of features in Figure 10 followed the same procedure, with the importance values sorted from highest to lowest based on their contribution to impurity reduction. This ranking confirmed that independent study hours, GPA, and social interaction were the dominant factors after optimization.

Overall, the evaluation results confirmed that optimizing the DT with PSO provided a significant improvement in accuracy and recall, particularly in the high-stress category. This finding is crucial for supporting early detection and more targeted interventions for students at high risk.

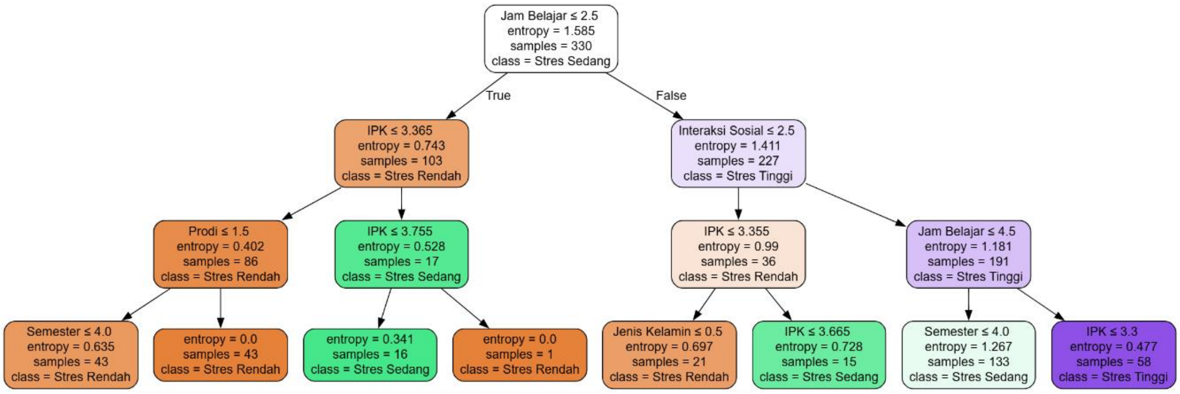

Figure 11. Decision Tree Visualization

Figure 11 presents the visualization of the DT constructed using the DT algorithm after the optimization process. The tree illustrates the classification rules based on the main predictor variables. The top-level nodes are dominated by GPA and independent study hours, indicating that these two factors are the most decisive in classifying student stress levels.

Each branch of the tree represents the outcome of testing a specific condition of a variable. For example, students with a low GPA and high study hours tend to be classified into the high-stress category, while students with a stable GPA and moderate study hours are more likely to fall into the moderate-stress category. This pattern is consistent with the feature importance findings, in which GPA and independent study hours play a dominant role.

4.4. Discussion

The 2% improvement in accuracy after applying PSO confirmed that parameter optimization enhanced the generalization capability of the DT. This finding was in line with Elham and Mehrabi [8], who combined DT and PSO for mental health prediction, and Hendra et al. [4], who demonstrated improved accuracy in academic outcome prediction. The increased recall in the high-stress group indicated the practical importance of optimization, as reliable identification of high-risk students is essential for early prevention and intervention.

Independent study hours and GPA were identified as the most influential predictors, supporting previous studies that associated academic workload and performance pressure with student stress [1], [3]. The strengthened role of social interaction also aligned with literature emphasizing the protective effect of peer support in reducing psychological burdens [1], [2]. These findings suggest that higher education institutions can develop policies and programs focusing on study management, academic counseling, and social support services.

The main strengths of this study include the use of an interpretable model, the integration of optimization to improve class balance, and the deployment of a prototype system, which increased the practical relevance of the findings. Nonetheless, the study was limited to a single institution and relied on self-reported data, which may introduce bias. Future work should expand the dataset to multiple institutions, examine ensemble methods optimized with PSO, and incorporate additional features such as physiological or behavioral data. Integration of the model into institutional academic information systems is also recommended to strengthen its applicability in student support services.

5. CONCLUSION

This research successfully constructed a predictive model for student academic stress levels by employing the DT algorithm enhanced with PSO. The analysis revealed that independent study hours (33%), GPA (18%), and social interaction (14%) were the three primary factors influencing student stress levels. These findings underscore that time management, academic achievement, and the quality of social support make significant contributions to students' mental health.

The DT model optimized with PSO achieved an accuracy of 95%, which is higher than the 93% obtained by the baseline model. This result provides evidence that swarm intelligence–based optimization is effective in improving the generalization capability of the model. From a practical perspective, the proposed system has the potential to serve as an early detection tool for identifying student stress levels, which may be integrated into institutional mental health services to support both preventive and curative interventions.

The clinical implication of this study lies in its ability to assist universities in identifying students at high risk of experiencing stress, thereby enabling timely interventions such as academic counseling, stress management programs, and adjustments to academic workload [25].

Future research is recommended to expand the dataset by involving students from multiple institutions, explore ensemble algorithms optimized with PSO, and integrate additional variables such as physiological data or text mining analysis to enrich prediction accuracy. For practical application, it is recommended to design a web-based platform that can be integrated with the university's student information system, accompanied by staff training and the formulation of data usage policies that prioritize ethics and privacy.

REFERENCES

[1] N. M. Bayantari, S. H. Indonesiani, and P. I. B. Apsari, "Regulasi Diri dalam Belajar dan Hubungannya dengan Stres Akademik Pada Mahasiswa Tingkat Pertama Fakultas Kedokteran," *J. Imiah Pendidik. dan Pembelajaran*, vol. 6, no. 3, pp. 609–618, Dec. 2022, doi: 10.23887/jipp.v6i3.51175.

[2] M. F. S. Hadi and N. S. Tajudin, "Pengaruh Stress Terhadap Prestasi Belajar Mahasiswa Fakultas Kedokteran Universitas Tarumanagara Angkatan 2021," *J. Kesehat. Tambusai*, vol. 4, no. 4, pp. 6694–6701, Dec. 2023, doi: 10.31004/jkt.v4i4.22712.

[3] A. Anjani and Y. Yamasari, "Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree," *J. Informatics Comput. Sci.*, vol. 05, pp. 83–89, 2023.

[4] H. Hendra, M. A. Azis, and S. Suhardjono, "Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decission Tree Berbasis Particle Swarm Optimization," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 1, pp. 102–107, Mar. 2020, doi: 10.32736/sisfokom.v9i1.756.

[5] P. B. Pankajavalli and G. S. Karthick, "Improved Stress Prediction Using Differential Boosting Particle Swarm Optimization-Based Support Vector Machine Classifier," 2022, pp. 233–244. doi: 10.1007/978-981-16-2422-3_20.

[6] P. Sharma and S. Kumar, "Retracted: Analyzing Customer Behavior on Online Products Using Decision Tree Machine Learning," in *2022 International Conference on Advances in Computing, Communication and Materials (ICACCM)*, IEEE, Nov. 2022, pp. 1–8. doi: 10.1109/ICACCM56405.2022.10009622.

[7] A. B. Raharjo, Y. A. Budiyono, and D. Purwitasari, "Analisis Laporan Gangguan Pelanggan Menggunakan Kombinasi Regresi Linear Berganda dan Klasifikasi Decision Tree Untuk Penentuan Komposisi Tim Yantek (Studi Kasus PLN UP3 Makassar Selatan)," *Briliant J. Ris. dan Konseptual*, vol. 7, no. 4, p. 1076, 2022, doi: 10.28926/briliant.v7i4.1037.

[8] E. Salehi, "Applying Particle Swarm Optimization-Base Decision Tree Classifier for Mental Illnesses," *J. Inf. Eng. Appl.*, vol. 9, no. 7, pp. 23–30, 2019, doi: 10.7176/jiea/9-7-03.

[9] V. R. P V, A. R. B, Harshitha, N. S, and S. S. Thingalaya, "Classification of Stress in Students using Machine Learning Algorithms," in *2022 International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, IEEE, Oct. 2022, pp. 229–233. doi: 10.1109/DISCOVER55800.2022.9974872.

[10] C. Ding, Y. Zhang, and T. Ding, "A systematic hybrid machine learning approach for stress prediction," *PeerJ Comput. Sci.*, vol. 9, p. e1154, Feb. 2023, doi: 10.7717/peerj-cs.1154.

[11] A. Kene and S. Thakare, "Mental Stress Level Prediction and Classification based on Machine Learning," in *2021 Smart Technologies, Communication and Robotics (STCR)*, IEEE, Oct. 2021, pp. 1–7. doi: 10.1109/STCR51658.2021.9588803.

[12] U. Ananthanagu and P. Agarwal, "Fostering Resilience: Machine Learning Models for Student Stress Prediction in Education," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/I2CT61223.2024.10543492.

[13] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," *Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, no. 24959, pp. 29–39, 2000, [Online]. Available: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining

[14] P. D. Sugiyono, "Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R&D dan Penelitian Pendidikan)," p. 18, 2019.

[15] D. Bedewy and A. Gabriel, "Examining perceptions of academic stress and its sources among university students: The Perception of Academic Stress Scale," *Heal. Psychol. Open*, vol. 2, no. 2, Jul. 2015, doi: 10.1177/2055102915596714.

[16] Wes McKinney, *Python for Data Analysis, 3rd Edition*. Sebastopol, CA: O'Reilly Media, 2022.

[17] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Sep. 2011, [Online]. Available: https://ehp.niehs.nih.gov/doi/10.1289/EHP4713

[18] K. R. Nadyandra and S. R. Nio, "Hubungan stress akademik dengan motivasi akademik pada mahasiswa yang sedang mengerjakan skripsi," *J. Pendidik. Tambusa*, vol. 7, no. 2, pp. 3614–3622, 2023.

[19] Y. Savira, E. Marhayuni, S. Widodo, and S. M. P. Lestrari, "Hubungan Manajemen Waktu Dengan Stres Akademik Pada Mahasiswa Kedokteran Universitas Malahayati Angkatan 2020," *J. Med. Malahayati*, vol. 7, no. 2, pp. 627–634, 2023, doi: 10.33024/jmm.v7i2.10120.

[20] S. Suhandiah, A. Ayuningtyas, and P. Sudarmaningtyas, "Tugas Akhir dan Faktor Stres Mahasiswa," *JAS-PT (Jurnal Anal. Sist. Pendidik. Tinggi Indones.*, vol. 5, no. 1, p. 65, 2021, doi: 10.36339/jaspt.v5i1.424.

[21] J. B. E. Putry, A. T. Sasongko, and W. Hadikristanto, "Optimization of Decision Tree Using Particle Swarm Optimization for Credit Risk of KMG Bank DKI," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. October, pp. 1403–1410, 2024.

[22] D. D. S. Fatimah and E. Rahmawati, "Penggunaan Metode Decision Tree dalam Rancang Bangun Sistem Prediksi untuk Kelulusan Mahasiswa," *J. Algoritm.*, vol. 18, no. 2, pp. 553–561, 2022, doi: 10.33364/algoritma/v.18-2.932.

[23] D. Wulandari, W. Roessali, and K. Budiraharjo, "Willingness to pay of the native chicken eggs in urban supermarket: evidence from Semarang, Indonesia," *J. Socioecon. Dev.*, vol. 3, no. 2, p. 79, Aug. 2020, doi: 10.31328/jsed.v3i2.1418.

[24] V. Manurung and A. F. Rozi, "Analisis Perbandingan Algoritma K-Nearest Neighbor dan Decision Tree Pada Klasifikasi Tingkat Stress Individu," *TIN Terap. Inform. Nusant.*, vol. 5, no. 1, pp. 73–80, 2024, doi: 10.47065/tin.v5i1.5286.

[25] Y. Yamasari *et al.*, "Exploring the tree algorithms to generate the optimal detection system of students' stress levels," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 1, pp. 548–558, 2024, doi: 10.11591/ijeecs.v36.i1.pp548-558.

BIBLIOGRAPHY OF AUTHORS

Syifa Ghina Dzakiyyah, S.Psi., is a Master's student in Information Systems at the Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia. She earned her Bachelor's degree in Psychology from Lambung Mangkurat University in 2024. Her academic interests include data analysis, research methodology, and psychological tes6ing. She is proficient in Microsoft Office applications, SPSS, and has strong skills in public speaking. Her professional experience includes roles as a Psychology Assistant at Lambung Mangkurat University, Teaching Assistant at Sapta Mandiri University, and Head of Secretariat at Sapta Mandiri University.

Deni Mahdiana is a lecturer at the Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia. The author has completed various research, particularly in the fields of Information Systems, Decision Support Systems, and Data Mining.