JSiI | Jurnal Sistem Informasi
Vol. 13 No. 1 | March 2026, P. 126-129

p-ISSN: 2406-7768
e-ISSN: 2581-2181

CLASSIFICATION OF STUDENTS' ACADEMIC STRESS LEVELS THROUGH DATA MINING TECHNIQUES BASED ON THE C4.5 ALGORITHM

Raihan Apriandi Putra Pratama[1], Sumiati[2], Hendry Gunawan[3]

[1,2,3] Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Serang Raya
Jln. Raya Cilegon Serang – Drangong Kota Serang

[1] rhnapp24@gmail.com

[2] sumiatiunsera82@gmail.com

[3] hendrygunawanan@unsera.ac.id

Abstract

Academic stress is a common psychological problem experienced by students during their education at university. This condition can be influenced by various factors, including high academic load, limited time to complete assignments, difficulty understanding lecture material, and an unsupportive learning environment. If not managed effectively, academic stress has the potential to negatively impact students' academic achievement, learning motivation, and mental health. Along with the development of information technology, the application of data mining techniques can be utilized to help identify and classify students' academic stress levels more objectively and accurately. This study aims to classify the academic stress levels of using the C4.5 algorithm. Research data were obtained through distributing questionnaires to students from the Informatics Engineering, Information Systems, and Computer Engineering Study Programs. The research stages include data preprocessing, the formation of a classification model, and evaluation of model performance. Algorithm was chosen because it has the ability to build a classification model in the form of a decision tree that is easy to interpret and is able to determine the most influential attributes in the decision-making process. The resulting classification model is expected to be a tool for universities, especially lecturers and student affairs, in conducting early detection of students who have the potential to experience academic stress. Thus, appropriate mentoring and intervention steps can be provided to support the academic success and psychological well-being of students.

Keywords: Academic Stress ,C4.5 Algorithm, Classification,Data Mining, Students.

I. INTRODUCTION

Academic stress is a common problem experienced by students during their studies at university. This condition arises from various academic demands, such as completing assignments, lab work, course projects, and preparing for final assignments, which require adaptability and good time management. If not managed optimally, academic stress can impact students' concentration, learning motivation, emotional stability, and academic achievement [1].

The Faculty of Information Technology, which comprises the Informatics Engineering, Information Systems, and Computer Systems study programs, has learning characteristics that require intensive mastery of theory and practice. The high academic load in information technology has the potential to increase student stress levels. Therefore, a method is needed to objectively identify and classify academic stress levels as a basis for decision-making in efforts to improve student well-being[2].

Furthermore, the attribute selection mechanism using the gain ratio allows for the development of an effective model for identifying factors influencing academic stress levels[3].

Previous studies have shown that the C4.5 algorithm has been widely applied in education for classification and decision-making. However, research specifically examining the classification of students' academic stress levels within the Faculty of Information Technology is still relatively limited [4]. The results are expected to serve as a basis for the faculty in early detection of students at risk of experiencing academic stress and support decision-making in developing student coaching and mentoring programs [5].

II. RESEARCH METHOD

Received: Feb 13th, 2026 | Published: March 30th, 2026

p-ISSN: 2406-7768
e-ISSN: 2581-2181

The type of research used in this thesis is applied research. Applied research aims to implement and develop certain technologies or methods to solve real problems that occur in certain communities or environments. In the context of this research, the problem raised is the classification of students' academic stress levels.

A. Research Stages

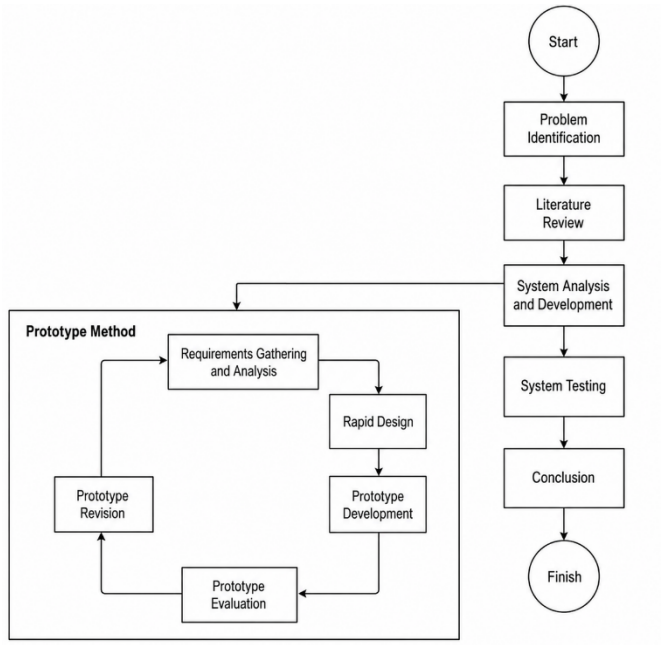


Figure 1. Research Stages

B. System Design

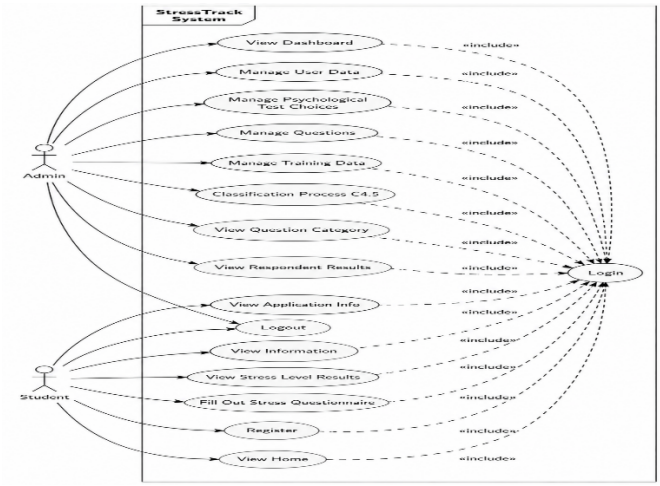


Figure 2 Use Case Diagram

C. Decision Tree method C.45

$$Information\ Gain\ (S, F_j) = Entropy\ (S) - \sum_{V_i \in V_{F_j}} \frac{|S_{V_i}|}{|S|} . Entropy\ (S_{V_i})$$

The calculation of the Entropy value can be done using the Shannon Entropy equation as follows:

$$Entropy\ (S) = \sum_{i=1}^{c} -p_i . log_2(p_i)$$

D. Model Evaluation

Success Rate (SR)

	Predictive	
	YES	NO
ACTUAL CLASS YES	TRUE POSSITIVE (TP)	FALSE NEGATIVE (FN)
NO	FALSE POSTIVE (FP)	TUE NEGATIVE (TN)

BAB III. RESULT AND DISCUSSION

A. Implementation Results

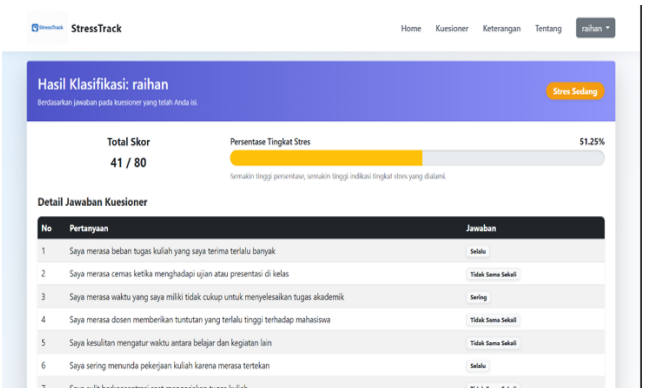


Figure 3. Classification Results Interface

The page in the image above is the Stress Level Classification Results page in the StressTrack application. This page displays the analysis results based on the questionnaire answers completed by students. The top section displays the user's identity and the stress level category obtained, for example, Moderate Stress. Additionally, the Total Score and Stress Level Percentage are displayed in the form of a progress bar, indicating the level of stress experienced. The higher the percentage, the higher the stress level. At the bottom, the Questionnaire Answer Details are presented, a list of questions and the answers selected by the student. This information helps users understand which aspects most influence the stress level classification results obtained.

B. Model Evaluation

Received: Feb 13th, 2026 | Published: March 30th, 2026

p-ISSN: 2406-7768
e-ISSN: 2581-2181

Model evaluation on the system is performed by comparing the system's classification results.

Confusion Matrix

		Classification		
		Low	Medium	High
Current	**Low**	12	0	0
	Medium	1	5	0
	High	0	0	2

The confusion matrix shows the comparison between the actual class and the model's predicted class. One Low class data item was predicted as Medium, while all High class data items were correctly classified.

Calculation of TP, FP, TN, FN

Low Class
TP = 12
FP = 1
FN = 0
TN = 20 – (12 + 1 + 0) = 7
Medium Class
TP = 5
FP = 0
FN = 1
TN = 20 – (5 + 0 + 1) = 14
High Class
TP = 2
FP = 0
FN = 0
TN = 20 – (2 + 0 + 0) = 18

Table 1.TP, FP, TN, FN Value Data

Class	TP	FP	TN	FN
Low	12	1	7	0
Mediun	5	0	14	1
Hight	2	0	18	0
Total	19	1	39	1

Precision, Recall, and F1-Score Values for Each Class

Precision, recall, and F1-score are used to evaluate model performance for each class in more detail. Precision indicates the accuracy of the model's predictions for a class, recall indicates the model's ability to find data belonging to that class, while the F1-score is the harmonic mean between precision and recall, reflecting the balance between the two.

Table 2. Precision, Recall, and F1-Score Values for Each Class

Class	*Precision* $\left(\frac{TP}{TP+FP}\right)$	*Recall* $\left(\frac{TP}{TP+FN}\right)$	*F1-Score* $\left(2\times\frac{Precision\times Recall}{Precision\times Recall}\right)$
Low	12/(12+1) = 0.923	12/(12+0) = 1.000	2 × (0.923 × 1.000) / (0.923 + 1.000) = 0.960
medium	5/(5+0) = 1.000	5/(5+1) = 0.833	2 × (1.000 × 0.833) / (1.000 + 0.833) = 0.909
Hight	2/(2+0) = 1.000	2/(2+0) = 1	2 × (1.000 × 1.000) / (1.000 + 1.000) = 1.000

Table 3.Macro F1 Value

Class	***F1-Score***	***Macro-Averaged F1 Score***
Low	0.960	$F_1 Macro = \frac{\sum F1 - Score}{\sum Kelas} = 0.956$
Medium	0.909	
Hight	1.000	

Table 4. Micro F1 Value

Class	**TP**	**FP**	**FN**	***Micro-Averaged F1 Score***
Low	12	1	0	$F_1 Micro = \frac{TP}{TP+\frac{1}{2}(FP+FN)} = \frac{19}{19+\frac{1}{2}(1+1)} = 0.950$
Medium	5	0	1	
Hight	2	0	0	
Total	**19**	**1**	**1**	

The assessment using the confusion matrix, precision, recall, and F1-score revealed that the C4.5 algorithm achieved an accuracy of 95%. These high evaluation scores reflect minimal misclassification and excellent classification performance, confirming the model's suitability for the research objectives.

BAB IV. CONCLUSION

The classification process is carried out by constructing a decision tree by calculating entropy, information gain, split info, and gain ratio to determine the most influential attributes in the decision-making process. The resulting model is able to categorize students' stress levels into low, medium, and high categories, thus serving as a basis for identifying students' academic stress levels.

The system provides various features to support the classification process, from data management, C4.5 algorithm calculations, decision tree visualization, and presentation of classification results. The simple and user-friendly interface allows users to process data and obtain information on academic stress levels more quickly and efficiently.

Test results indicate that the developed classification model has a fairly good level of accuracy in predicting students' academic stress levels. However, several cases of misclassification were found due to similar data characteristics between categories, particularly in the low and medium stress categories.

Received: Feb 13th, 2026 | Published: March 30th, 2026

p-ISSN: 2406-7768
e-ISSN: 2581-2181

V. RECOMMENDATIONS

This research still has several limitations that require attention in further development. The classification model performance evaluation was conducted using training data without cross-validation, so the accuracy obtained does not fully reflect the model's ability to generalize to new data. The application of the k-fold cross-validation method is recommended to produce a more objective and consistent evaluation process and minimize testing bias.

Furthermore, this research only implemented the C4.5 algorithm without conducting a comparative analysis with other classification algorithms, such as ID3, CART, Naive Bayes, or K-Nearest Neighbor (KNN). Comparing several algorithms could provide a more comprehensive picture of the accuracy, stability, and effectiveness of each method in classifying students' academic stress levels.

REFERENCES

[1] Anjani. (2023). Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree. *Journal of Informatics and Computer Science*, *05*, 83–89.

[2] Apandi et al. (2023). Pembuatan Website Penjualan Toko Baju Biazra-Store Menggunakan Php Dan Mysql. *Jurnal Teknik Dan Science*, *2*(3), 80–91. https://doi.org/10.56127/jts.v2i3.998

[3] Azkiya & Ifriza. (2025). Analysis of Coding Stress Impact on Students Programming Skills with Random Forest and C4.5 Algorithms. *Jurnal Ilmu Komputer Dan Informasi*, *18*(2), 167–180. https://doi.org/10.21609/jiki.v18i2.1487

[4] Desyanti, D. (2021). Implementasi Metode C.45 dalam Menganalisa Tingkat Kecemasan Mahasiswa Menyusun Tugas Akhir. *Jurnal Unitek*, *14*(1), 17–29. https://doi.org/10.52072/unitek.v14i1.175

[5] Efendi, E. P., & Barata, M. A. (2024). *KOMPARASI ALGORITMA SVM DAN C4 . 5 DENGAN BACKWARD ELIMINATION UNTUK KLASIFIKASI TINGKAT STRES MAHASISWA COMPARATIVE ANALYSIS OF SVM AND C4 . 5 ALGORITHMS WITH BACKWARD*. *3*, 539–549. https://doi.org/10.26798/jiko.v9i3.2130

[6] Elisa, E., Simanjuntak, P., & Maslan, A. (2025). Analisis Diagnosis Tingkat Kesehatan Mental Dengan Teknik. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)*, *7*(1), 77–86.

[7] Kurniah, R., & Diana, E. (2025). Klasifikasi Tingkat Stres Pelajar Bedasarkan Pola Gaya Hidup Menggunakan Algoritma Decission Tre C4.5. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *8*(2), 742–747. https://doi.org/10.32672/jnkti.v8i2.8888

[8] Ledoh et al. (2023). C4.5 Algorithm Implementation to Predict Student Satisfaction Level of Lecturer's Performance in the Covid-19 Pandemic. *Komputasi: Jurnal Ilmiah Ilmu Komputer Dan Matematika*, *20*(2), 126–134. https://doi.org/10.33751/komputasi.v20i2.8284

[9] Muhammad Reza. (2024). *PENGEMBANGAN SISTEM PREDIKSI KESEHATAN MENTAL TINGKAT STRES PADA MAHASISWA DI LINGKUNGAN UIN AR-RANIRY MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus: Fakultas Tarbiyah Dan Keguruan) SKRIPSI Diajukan Oleh* (Vol. 5).

[10] Renaningtias et al. (2024). Studi Komparasi Algoritma Decision Tree C4.5 dan K-Nearest Neighbor pada Klasifikasi Masa Studi dan Tingkat Stres Mahasiswa. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, *13*(3), 1776–1785. https://doi.org/10.35889/jutisi.v13i3.2272

[11] Sumiati, Repi, V. V. R., Hendriyati, P., Anharudin, Yusta, A., & Triayudi, A. (2023). Classification of cardiac disorders based on electrocardiogram data using a Decision Tree classification approach with the C45 algorithm. IAES International Journal of Artificial Intelligence, 12(3), 1128–1138. https://doi.org/10.11591/ijai.v12.i3.pp1128-1138

Received: Feb 13th, 2026 | Published: March 30th, 2026