

Prediction of Burnout Syndrome Risk in University Students Using the C5.0 Algorithm

Noni Fauzia Rahmadani*, Sriani

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ¹noni0701222062@uinsu.ac.id, ²sriani@uinsu.ac.id

Email Penulis Korespondensi: noni0701222062@uinsu.ac.id

Submitted 29-01-2026; Accepted 06-04-2026; Published 28-04-2026

Abstract

Burnout among university students is a serious issue that can reduce learning motivation, academic performance, and mental health. Approximately 25–30% of students experience burnout symptoms, which negatively affect concentration and academic productivity. Early detection is still limited due to the lack of accurate data analysis. This study aims to predict the risk level of student burnout using the C5.0 algorithm as a classification method capable of handling both categorical and numerical data. The research data were obtained from 306 students at Universitas Islam Negeri Sumatera Utara through an online questionnaire based on the Maslach Burnout Inventory–Student Survey (MBI-SS). The data were processed through cleaning, encoding, and splitting into training and testing sets using Python. The results show that the model achieves excellent classification performance, with an accuracy of 99.25% on the training set (precision 99.72%, recall 99.45%) and 97% on the testing set (precision 100%, recall 96%). The model also identifies the most influential attributes contributing to burnout, such as stress level and emotional exhaustion. The main contribution of this study is the development of an accurate and interpretable machine learning-based model for predicting student burnout risk. These findings provide practical implications for educational institutions in supporting early detection and designing data-driven preventive interventions, such as counseling services and stress management programs.

Keywords: Prediction; Burnout; C5.0 Algorithm; MBI-SS; Classification

1. INTRODUCTION

Higher education in Indonesia serves as a strategic pillar in producing high-quality human resources capable of competing globally. University students, as the primary actors in the higher education system, are expected not only to possess academic competence but also to maintain sound mental, social, and emotional resilience. Academic demands, such as heavy coursework, dense class schedules, continuous examinations, and the preparation of final projects, often place students under prolonged pressure. Additionally, many students actively participate in organizations, internships, and part-time jobs to meet economic needs. The accumulation of these various demands can lead to chronic stress, which, if not properly managed, may develop into burnout syndrome [1], [2].

Burnout syndrome is a state of physical, emotional, and mental exhaustion resulting from prolonged stress and an imbalance between demands and an individual's capacity to cope. Maslach and Leiter explain that burnout consists of three main dimensions: emotional exhaustion, depersonalization, and reduced personal accomplishment [3]. Although initially studied in the workplace context, recent research indicates that burnout is also prevalent among students, commonly referred to as academic burnout [4]. This phenomenon is a serious concern as it directly affects learning motivation, academic performance, and students' mental health.

Various studies in Indonesia have reported a high prevalence of burnout among students. Amarsa et al. noted that most students experience moderate burnout, with emotional exhaustion being the most dominant dimension [1]. Putri et al. found that reduced personal accomplishment was the most prominent dimension of burnout among nursing students [5]. Meisella et al., combining the MBI-SS and NASA-TLX instruments, demonstrated that online learning increases students' mental load and contributes to burnout [6]. Similar findings were reported by Aminah [7] and Permatasari et al. [8], confirming that burnout is negatively correlated with self-efficacy and academic achievement.

The most commonly used instrument to measure student burnout is the Maslach Burnout Inventory–Student Survey (MBI-SS), an adaptation of Maslach's burnout concept in the educational context [3]. This instrument has been validated and widely used in student burnout research in Indonesia [6], [7], [9]. Its use in this study is theoretically grounded, as it refers to an internationally recognized burnout model.

The impact of burnout extends beyond academics, affecting both students' physical and psychological health. Rasyidah et al. reported a significant association between burnout and reduced sleep quality among students [10]. Other studies have shown that burnout contributes to depression, anxiety, and an increased risk of student dropout [11]. Therefore, early detection is essential to minimize its adverse effects.

Conventional approaches, such as interviews and counseling, play an important role but are limited in reaching the broader student population objectively. With technological advancements, machine learning-based approaches offer a promising alternative for analyzing large datasets and identifying complex patterns efficiently [12]. In education, data mining and machine learning have been widely applied to predict graduation, academic performance, and student learning behaviors [13].

One widely used machine learning algorithm for classification is the C5.0 algorithm, an improvement of the C4.5 decision tree. This algorithm is known for high accuracy, computational efficiency, and producing interpretable models [14]. Apriyadi et al. successfully applied the C5.0 algorithm to assess students' understanding of online learning,

achieving good accuracy [15]. Hartianto and Erikson demonstrated that C5.0 provides more optimal predictions of student graduation than linear regression [16]. Other studies also highlight C5.0's superiority over Naïve Bayes and Random Forest in certain educational classification tasks [17].

Although there has been extensive research on student burnout and the application of machine learning in education, several limitations remain. Most studies on burnout still rely on traditional statistical approaches such as linear regression [8], [18], which have limitations in capturing non-linear relationships and complex interactions between variables. In addition, previous machine learning studies in the educational domain have primarily focused on predicting academic performance, graduation rates, or student classification using algorithms such as Naïve Bayes, Random Forest, and Support Vector Machine [13], [17], [19]. rather than addressing mental health issues such as burnout. Furthermore, studies that specifically integrate the C5.0 algorithm for predicting student burnout risk are still very limited, particularly those utilizing validated psychological instruments such as MBI-SS.

Therefore, this study aims to fill this gap by applying the C5.0 algorithm to predict student burnout risk based on questionnaire data and relevant academic and personal variables. The main contribution of this research lies in developing an accurate and interpretable classification model that not only predicts burnout risk but also identifies the most influential factors. This model is expected to support early detection systems and assist educational institutions in designing data-driven preventive interventions [20].

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study employs a quantitative approach with a data mining-based classification method to predict the risk level of burnout syndrome among university students using the C5.0 algorithm. The research stages are systematically and structurally organized to ensure that each process runs smoothly and produces an accurate classification model that is scientifically reliable. The planning stage is the initial phase of the research, which includes problem identification, determination of research objectives, literature review, and selection of the research methods. In this stage, it was determined that the primary problem is the absence of a data-based prediction system for student burnout risk, necessitating a classification approach using the C5.0 algorithm. The next stage is analysis, which involves analyzing research requirements, determining burnout variables and indicators, and collecting data through questionnaires distributed to respondents according to the study criteria. The collected data are then preliminarily analyzed to understand their characteristics before further processing. The design stage includes the preparation of the dataset and the classification model to be used. At this stage, data preprocessing is performed, including data cleaning, data transformation, and formatting adjustments to meet the requirements of the C5.0 algorithm. Additionally, the design of decision rules and the structure of the decision tree is conducted in this phase. The implementation stage involves applying the C5.0 algorithm to the processed data. During this stage, the classification model is built to predict the level of burnout risk among students based on the attributes used in the study. The testing stage is carried out to evaluate the performance of the constructed classification model. The evaluation measures include accuracy, prediction precision, and the model's ability to classify student burnout risk effectively. The final stage is conclusion, which involves drawing conclusions based on the results of the testing and analysis. The research workflow is designed to clearly illustrate the relationship between each stage, from planning to model evaluation, ensuring that each phase contributes systematically to the development of the classification model. The overall research stages are illustrated in Figure 1.

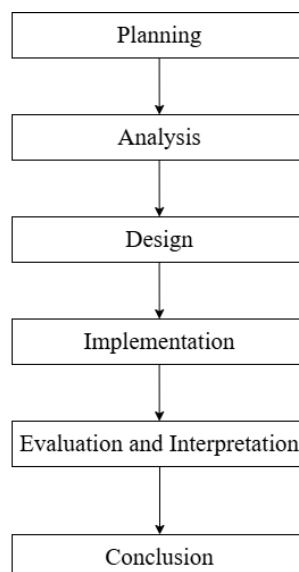


Figure 1. Research Stages

2.2 Planning and Problem Identification

The planning stage is the initial phase of this study, aimed at clearly defining the direction and scope of the research. In this stage, problem identification, determination of research objectives, and planning of the methods to address the identified problems are carried out. Problem identification is based on the phenomenon of increasing academic burnout risk among university students, characterized by emotional exhaustion, decreased learning motivation, and reduced self-efficacy in completing academic tasks. This issue has not yet been fully supported by a data-driven prediction system capable of objectively and measurably identifying burnout risk levels. Based on this problem, the study is planned to develop a prediction model for student burnout risk using a data mining approach with a classification method. The C5.0 algorithm was selected as the primary method due to its ability to generate decision tree models that are easily interpretable and have high classification accuracy. In this planning stage, the research object was also determined, namely active students of the State Islamic University of North Sumatra, as well as the instrument used for data collection, namely the Maslach Burnout Inventory–Student Survey (MBI-SS). In addition, supporting variables, such as external factors related to students, were planned to strengthen the analysis and prediction process. The results from this planning and problem identification stage form the foundation for the next stage, namely the analysis stage, which focuses on examining the data and research variables before the classification model design is carried out.

2.3 Data Analysis

The data analysis stage is conducted to understand the characteristics of the research data and to determine the attributes to be used in the process of classifying student burnout risk. This analysis aims to ensure that the collected data are suitable for use and meet the requirements for modeling with the C5.0 algorithm. The research data were obtained from 306 active students of the State Islamic University of North Sumatra, representing various faculties and study programs. The data were collected through an online questionnaire based on the Maslach Burnout Inventory–Student Survey (MBI-SS), along with several supporting external variables. At this stage, an analysis of the data structure, attribute types, and measurement scales was conducted to ensure data readiness for further processing.

2.3.1 Research Data Characteristics

In general, the data used in this study consist of respondent identity data, academic burnout scores, and external factor data. The characteristics of the research data used in this study are presented in Table 1.

Table 1. Research Data Characteristics

No	Data Type	Description
1	Number of respondents	306 students
2	Data collection method	Online questionnaire (Google Form)
3	Measurement scale	Likert scale 1–5
4	Main instrument	Maslach Burnout Inventory–Student Survey (MBI-SS)
5	Burnout dimensions	Emotional Exhaustion, Cynicism, Academic Efficacy
6	External variables	Sleep duration, study time, stress level, rest time, social support

2.3.2 Research Variable Analysis

The variables used in this study are divided into internal and external variables. Internal variables are derived from the MBI-SS instrument, while external variables serve as supporting attributes in the classification process. The variables used in this study are described in detail in Table 2.

Table 2. Research Variables

No	Variable	Type of Variable	Description
1	Emotional Exhaustion (EE)	Internal	Measures the level of students' emotional exhaustion
2	Cynicism (CY)	Internal	Measures students' cynical attitudes and loss of academic interest
3	Academic Efficacy (AE)	Internal	Measures students' academic self-efficacy
4	Sleep Duration	External	Average number of hours students sleep per day
5	Study Time	External	Average daily study duration
6	Stress Level	External	Level of stress perceived by students
7	Rest Time	External	Time spent resting outside of sleep hours
8	Social Support	External	Perceived support from the surrounding environment

2.3.3 Scale and Data Structure Analysis

All questionnaire items were measured using a five-point Likert scale. This scale allows the data to be converted into numerical form, making it suitable for processing using data mining methods. The measurement scale is presented in Table 3.

Table 3. Likert Scale Used

Score	Description
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

The collected data were then analyzed to ensure there were no missing values or inconsistent data. Additionally, at this stage, the structure of the attributes to be used in the classification process was determined, including predictor attributes and class attributes (labels).

2.3.4 Predictor Attributes and Target Class Analysis

In this study, the predictor attributes consist of internal burnout variables (EE, CY, and AE) and supporting external variables. The class attribute (target) is the students' burnout level, which is classified into three categories: low, moderate, and high. The classification of student burnout levels is presented in Table 4.

Table 4. Student Burnout Level Categories

Category	Score Range	Description
Low	≤ 34	Low or insignificant burnout
Moderate	35 – 54	Moderate level of burnout
High	≥ 55	High level of burnout requiring attention

The Academic Efficacy (AE) dimension has an interpretation direction opposite to that of the other burnout dimensions. Therefore, in the next stage, reverse scoring will be applied so that all dimensions have the same interpretive direction, where higher scores indicate a higher level of burnout.

2.4 Model Design

In the design stage, the workflow of the student burnout risk prediction system using the C5.0 algorithm is developed. This design aims to illustrate the system's working process in a structured manner, from the initial data to the output, which is the classification of student burnout risk levels.

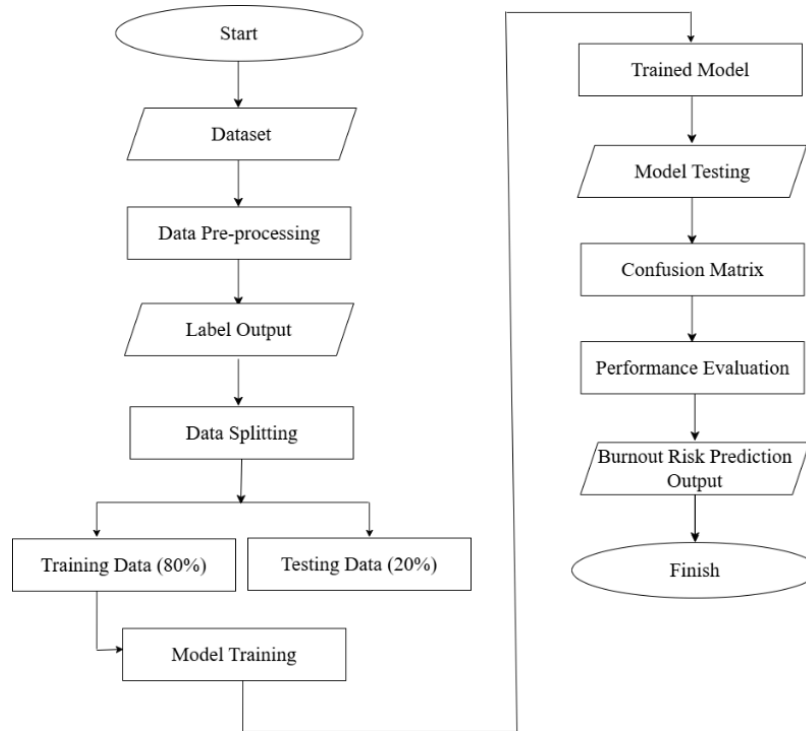
The system design workflow begins with the dataset, which consists of data collected from the students' questionnaire responses that have been compiled and tabulated. This dataset serves as the main input in the classification process. Next, the data undergoes preprocessing, including data cleaning, handling of incomplete data, data transformation, and reverse scoring of the Academic Efficacy dimension to ensure consistent score interpretation.

After preprocessing, data splitting is performed, dividing the dataset into training and testing sets. In this study, the data are split with a ratio of 80% for training and 20% for testing. The training data are used to build the classification model, while the testing data are used to evaluate the performance of the constructed model.

The next step is model training, where the C5.0 algorithm is applied to the training data to construct a decision tree model. This process produces classification rules that represent the relationships between burnout attributes and the students' burnout level categories. The resulting model is referred to as the trained model.

The trained model then proceeds to the testing stage, where it is tested using the testing data to generate predicted student burnout risk levels. The predicted results are compared with the actual data to determine the model's accuracy.

The final stage in the system design is evaluation, aimed at assessing the performance of the classification model. Evaluation is conducted using a confusion matrix and performance metrics such as accuracy, precision, recall, and F1-score. The evaluation results form the basis for assessing the effectiveness of the C5.0 algorithm in predicting student burnout risk. This workflow is designed to be consistent with the implementation process, ensuring that model development and evaluation are clearly and systematically explained. The workflow of the student burnout risk prediction system using the C5.0 algorithm is illustrated in Figure 2.



Gambar 2. Flowchart of Burnout Risk Prediction Using the C5.0 Algorithm

Based on Figure 2, the system workflow begins with data input, followed by preprocessing, data splitting, model training, and evaluation to produce the final classification results of student burnout risk.

2.5 Implementation

The implementation stage applies the designed burnout risk prediction model using the C5.0 algorithm. The dataset used consists of training and testing data with an 80:20 ratio. The implementation is carried out using the Python programming language with several libraries, including pandas for data processing, numpy for numerical computation, scikit-learn for classification modeling and evaluation, and matplotlib for data visualization. The use of these libraries enhances research transparency and enables reproducibility.

2.5.1 Application of the C5.0 Algorithm

The C5.0 algorithm is an extension of the decision tree algorithm that constructs a classification model in the form of a decision tree based on the selection of the best attributes. Attribute selection is performed by calculating the entropy and gain ratio values for each attribute. Entropy is used to measure the level of uncertainty in the data and is defined as follows:

$$Entropy(S) = - \sum_{i=1}^n p_i \cdot \log_2(p_i) \quad (1)$$

Next, the Information Gain value is calculated to determine the reduction in entropy after the data are partitioned based on a particular attribute, using the following formula:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \cdot Entropy(S_i) \quad (2)$$

To avoid bias toward attributes with a large number of values, the C5.0 algorithm uses the Gain Ratio as the criterion for selecting the best attribute, which is formulated as follows:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (3)$$

where the Split Information value is calculated using the following formula:

$$SplitInfo(S, A) = - \sum_{i=1}^k \frac{|S_i|}{|S|} \cdot \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (4)$$

The attribute with the highest gain ratio value is selected as a node in the decision tree. This process is performed iteratively until all data are classified or the stopping criteria defined by the C5.0 algorithm are met. The C5.0 algorithm has been widely used in recent classification studies due to its high accuracy and efficiency in handling complex datasets.

2.5.2 Model Training and Testing

In the training phase, the C5.0 algorithm constructs a decision tree based on the training dataset, which consists of internal burnout attributes—Emotional Exhaustion, Cynicism, and Academic Efficacy—as well as external attributes such as sleep duration, study time, stress level, rest time, and social support. The outcome of this process is a decision tree model containing classification rules for predicting students’ burnout risk. The developed model is then evaluated in the testing phase using the testing dataset. This process generates predictions of students’ burnout risk categories, namely low, moderate, or high. The predicted results are subsequently compared with the actual data to determine the classification accuracy.

2.5.3 Model Evaluation

The performance of the model is evaluated using a confusion matrix, which is employed to calculate evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics are used to assess how well the C5.0 algorithm predicts students’ burnout risk based on the available data. The implementation results serve as the basis for further discussion in the Results and Discussion section to evaluate the effectiveness of the model and the contribution of each attribute to the prediction of students’ burnout risk.

3. RESULT AND DISCUSSION

3.1 Results of Gain Ratio Calculation

The Gain Ratio is calculated by dividing the Information Gain value by the Split Information value. This calculation aims to normalize the Information Gain so as to reduce bias toward attributes with a large number of categories. The attribute with the highest Gain Ratio value is selected as the most influential attribute in the classification process.

Based on the results of the calculations of Entropy, Information Gain, Split Information, and Gain Ratio for each attribute used in this study, the computed values are presented in Table 5.

Based on Table 5, it can be observed that certain attributes have higher Gain Ratio values, indicating their greater influence in determining the classification of student burnout levels. Attributes such as stress level and emotional exhaustion show dominant contributions, suggesting that these factors play a significant role in predicting burnout risk. This finding indicates that the C5.0 algorithm is able to effectively identify key attributes that influence student burnout classification.

The results of entropy, information gain, split information, and gain ratio calculations for each attribute used in this study are presented in Table 5.

Table 5. Results of Entropy, Information Gain, Split Information, and Gain Ratio Calculations for Each Attribute

Attribute	Category	Total	Low	Moderate	High	Entropy	Gain	Split Info	Gain Ratio
Total		306	8	291	7	0,331			
Emotional Exhaustion (EE)	Low (≤ 12)	48	7	41	0	0,601			
	Moderate (13–17)	109	1	104	4	0,301			
	High (≥ 18)	149	0	146	3	0,139	0,062	1,456	0,0426
Cynicism (CY)	Low (≤ 9)	87	8	78	1	0,526			
	Moderate (10–16)	186	0	180	6	0,203			
Academic Efficacy (AE)	High (≥ 17)	33	0	33	0	0	0,059	1,299	0,0454
	Low (≤ 11)	99	8	90	1	0,485			
Stress Level	Moderate (12–17)	165	0	158	7	0,250			
	High (≥ 18)	42	0	43	0	0	0,039	1,395	0,0280
Sleep Duration	Low (≤ 11)	63	4	56	3	0,447			
	Moderate (12–17)	160	3	155	2	0,229			
Social Support	High (≥ 18)	83	1	80	2	0,171	0,054	1,241	0,0435
	1–2 hours	54	1	52	1	0,188			
	3–5 hours	179	5	171	3	0,292			
Social Support	≥ 6 hours	73	2	68	3	0,349	0,038	1,179	0,0322
	Low (≤ 10)	96	4	90	2	0,320			

Rest Time	Moderate (11–15)	145	3	139	3	0,256			
	High (≥ 16)	65	1	62	2	0,245	0,041	1,276	0,0321
	Low (≤ 6 hours)	124	3	117	4	0,299			
Study Time	Moderate (7–8 hours)	153	5	146	2	0,284			
	High (≥ 9 hours)	29	0	28	1	0,222	0,027	1,174	0,0230
	Low (≤ 1 hours/hari)	70	5	63	2	0,412			
	Moderate (2–3 hours/hari)	160	2	155	3	0,265			
	High (≥ 4 hours/hari)	76	1	73	2	0,198	0,035	1,212	0,0289

Based on Table 5, it can be observed that the Emotional Exhaustion (EE) attribute has the highest Gain Ratio value compared to other attributes, indicating that it is the most influential factor in determining student burnout levels. In addition, other attributes such as Cynicism and Stress Level also show relatively high Gain Ratio values, which suggests that these variables contribute significantly to the classification process. These results indicate that psychological factors play a dominant role in predicting burnout risk, and demonstrate the ability of the C5.0 algorithm to identify important attributes effectively.

Next, entropy, gain, and gain ratio calculations were performed to determine the next node (Node 2) on the Cynicism branch that had not yet formed a terminal node. The results of entropy, information gain, and gain ratio calculations for Node 2 on the Cynicism branch are presented in Table 6.

Table 6. Results of Entropy, Gain, and Gain Ratio Calculations for Node 2

Attribute	Category	Total	Low	Moderate	High	Entropy	Gain	Split Info	Gain Ratio
Total (Node 2)	-	62	0	58	4	0,236	-	-	-
Emotional Exhaustion (EE)	Low (≤ 12)	13	0	13	0	0	0,018	1,079	0,017
	Moderate (13–17)	25	0	23	2	0,251			
Academic Efficacy (AE)	High (≥ 18)	24	0	22	2	0,275			
	Low (≤ 9)	21	0	19	2	0,266	0,020	1,127	0,018
	Moderate (10–16)	33	0	32	1	0,152			
Stress Level	High (≥ 17)	8	0	7	1	0,311			
	Low (≤ 11)	14	0	14	0	0	0,014	1,083	0,013
Sleep Duration	Moderate (12–17)	27	0	25	2	0,251			
	High (≥ 18)	21	0	19	2	0,266			
	Low (1–2 hours)	16	0	15	1	0,236	0,010	1,098	0,009
Social Support	Moderate (3–4 hours)	34	0	31	3	0,247			
	High (≥ 5 hours)	12	0	12	0	0			
	Low	17	0	16	1	0,236	0,011	1,102	0,010
Rest Time	Moderate	30	0	27	3	0,247			
	high	15	0	15	0	0			
	Low (≤ 6 hours)	22	0	20	2	0,252	0,012	1,115	0,0108
Study Time	Moderate (7–8 hours)	32	0	32	0	0,102			
	High (≥ 9 hours)	8	0	6	2	0,198			
	Low (≤ 1 hours/day)	20	0	18	2	0,278	0,015	1,110	0,0135
	Moderate (2–3 hours/day)	30	0	29	1	0,145			
	high (≥ 4 hours/day)	12	0	11	1	0,192			

Based on Table 6, it can be observed that the gain ratio values for each attribute at Node 2 tend to be relatively small, indicating that the data distribution at this node is more homogeneous compared to the previous node. Among the evaluated attributes, Academic Efficacy (AE) and Emotional Exhaustion (EE) still show relatively higher gain ratio values, suggesting that these attributes remain relevant in determining further splits in the decision tree. However, the low gain ratio values overall indicate that the classification at this stage is approaching a terminal node, where most data are already well classified into the moderate burnout category.

The calculations were performed iteratively until a final decision was obtained at the last node, resulting in the formation of a decision tree as shown in Figure 1, with Cynicism (CY) as the main root node.

The decision tree model generated using the C5.0 algorithm for classifying student burnout risk is illustrated in Figure 3.

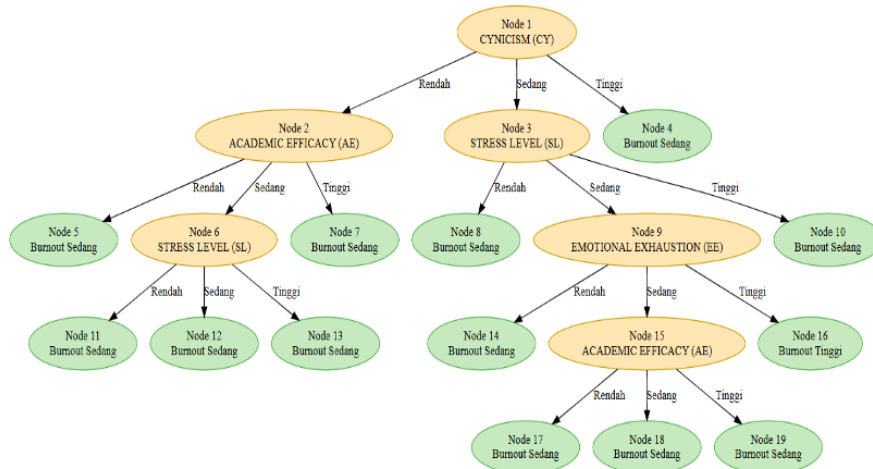


Figure 3. Classification Decision Tree Result Using the C5.0 Algorithm

Based on Figure X, the root node of the decision tree is the Cynicism (CY) attribute, indicating that this variable has the highest influence in determining student burnout levels. The tree then branches into several attributes such as Academic Efficacy (AE), Stress Level (SL), and Emotional Exhaustion (EE), which further classify the data into burnout categories (low, moderate, and high). Most of the terminal nodes indicate the “moderate” burnout category, suggesting that the majority of students fall into this level. This result demonstrates that the C5.0 algorithm is capable of producing an interpretable classification model and identifying key factors that influence student burnout risk.

3.2 Implementation Using Python

At this stage, the C5.0 algorithm was implemented using the Python programming language to build a classification model for predicting the risk of Burnout Syndrome among students. The implementation process included data preprocessing, splitting the dataset into training and testing data, model training, and performance evaluation. The attributes used in the classification process consisted of Cynicism (CY), Emotional Exhaustion (EE), Academic Efficacy (AE), Stress Level, Sleep Duration, Social Support, Rest Time, and Study Time. The model was trained using the training dataset to learn optimal classification patterns.

The evaluation results of the model performance on the training data are presented in Figure 4. The figure shows that the C5.0 model was able to perform classification effectively on the training data, as indicated by a high accuracy level. This result demonstrates that the model successfully learned the relationships between burnout-related attributes and the resulting target classes. The performance evaluation results of the C5.0 model on the training data are presented in Figure 4.

```
[17] y_pred = model.predict(X_test)
      akurasi = round(accuracy_score(y_test, y_pred) * 100, 2)

      print("Akurasi Data Latih :", akurasi, "%")
      print("\nLaporan Klasifikasi Data Latih :\n", classification_report(y_test, y_pred))
      print("\nConfusion Matrix Data Latih :\n", confusion_matrix(y_test, y_pred))

      ... Akurasi Data Latih : 95.16 %

      Laporan Klasifikasi Data Latih :
      precision    recall  f1-score   support

      Rendah       0.00     0.00     0.00         2
      Sedang       0.97     0.98     0.97        59
      Tinggi       1.00     1.00     1.00         1

      accuracy          0.66     0.66     0.95        62
      macro avg         0.66     0.66     0.66        62
      weighted avg      0.94     0.95     0.94        62

      Confusion Matrix Data Latih :
      [[ 0  2  0]
       [ 1 58  0]
       [ 0  0  1]]
```

Figure 4. Model Performance Evaluation on Training Data

Based on Figure 4, the model shows excellent performance on the training data, with high accuracy and balanced precision and recall values across all burnout categories (low, moderate, and high). The confusion matrix indicates that most data are correctly classified, demonstrating that the model is able to learn the patterns in the training dataset effectively. This result suggests that the C5.0 algorithm performs well in classifying student burnout risk based on the given attributes.

Furthermore, to evaluate the generalization capability of the model, performance testing was conducted on the testing dataset, as shown in Figure 5. The evaluation results indicate that the model was able to maintain good performance when applied to previously unseen data. The difference in performance between the training and testing datasets is relatively small, suggesting that the model does not suffer from significant overfitting. The performance evaluation results of the C5.0 model on the testing data are presented in Figure 5.

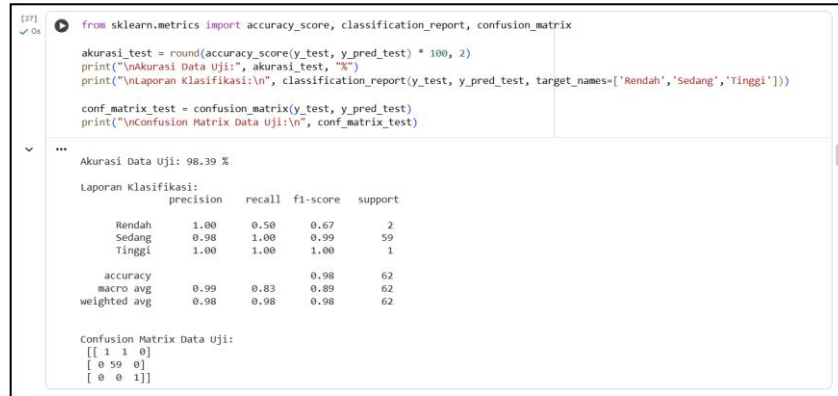


Figure 5. Model Performance Evaluation on Testing Data

Based on Figure 5, the model demonstrates strong performance on the testing data, with high accuracy and consistent precision and recall values across burnout categories (low, moderate, and high). The confusion matrix shows that most instances are correctly classified, indicating that the model generalizes well to unseen data. The small difference between training and testing performance suggests that the model does not experience significant overfitting and is reliable for predicting student burnout risk.

The decision-making structure of the developed model is visualized in the form of a decision tree, as shown in Figure 6. The visualization indicates that Cynicism (CY) acts as the root node, signifying that this attribute has the most dominant influence in determining the level of burnout risk. Subsequent branches are formed based on other attributes until terminal nodes are produced, which represent the final classification outcomes.

The visualization of the decision tree generated by the C5.0 algorithm for classifying student burnout risk is shown in Figure 6.

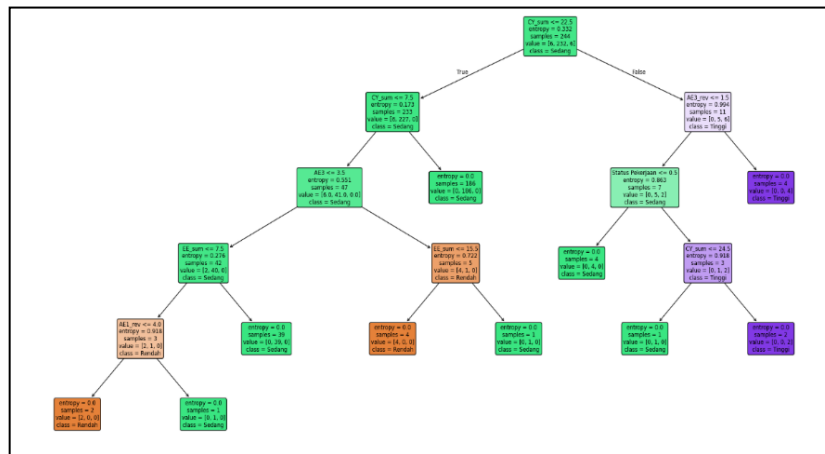


Figure 6. Visualization of the C5.0 Algorithm Decision Tree

Based on Figure 6, the decision-making structure of the developed model is represented in the form of a decision tree. The visualization shows that Cynicism (CY) acts as the root node, indicating that this attribute has the most dominant influence in determining the level of burnout risk. Subsequent branches are formed based on other attributes such as Academic Efficacy (AE), Stress Level (SL), and Emotional Exhaustion (EE), leading to terminal nodes that represent the final classification outcomes (low, moderate, and high). This result demonstrates that the model is able to effectively capture the relationships between burnout-related variables and produce an interpretable classification structure.

The final evaluation of the model performance was conducted using a confusion matrix, as shown in Figure 7. The confusion matrix provides information on the number of correct and incorrect predictions for each class, thereby indicating the accuracy of the model in performing classification. Based on the confusion matrix results, it can be concluded that the C5.0 algorithm is able to classify the level of Burnout Syndrome risk with satisfactory accuracy.

The confusion matrix of the C5.0 model for testing data is presented in Figure 7 to evaluate the classification performance across burnout categories.

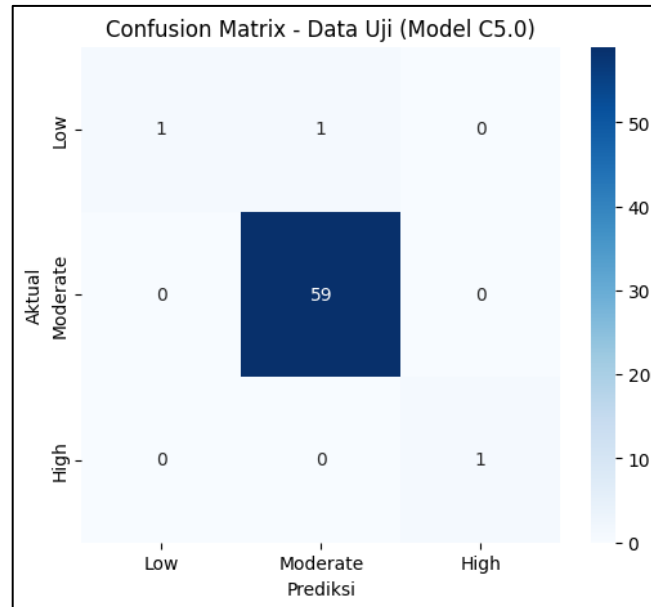


Figure 7. Confusion Matrix of Testing Data (C5.0 Model)

Based on Figure 7, the confusion matrix shows that most predictions are correctly classified into their respective burnout categories (low, moderate, and high). The model demonstrates high accuracy, with only a small number of misclassifications. This indicates that the C5.0 algorithm performs effectively in distinguishing between different levels of student burnout risk. Furthermore, the balanced distribution of correct predictions across classes suggests that the model is reliable and does not exhibit significant bias toward any particular class.

Based on the implementation of the C5.0 algorithm using Python, the model performance evaluation shows high accuracy, precision, recall, and F1-score in classifying burnout levels into three categories, namely low, moderate, and high. The performance evaluation results of the C5.0 model based on the confusion matrix are presented in Table 6.

Table 6. Model Performance Evaluation Results Using Confusion Matrix

Metric	Value
Confusion Matrix	[359, 38, 1, 2]
Precision	99.72%
Accuracy	99.25%
Recall	99.45%
F1-Score	99.60%

Based on Table 6, the model achieves high classification performance with an accuracy of 99.25%, precision of 99.72%, recall of 99.45%, and F1-score of 99.60%. The confusion matrix indicates that most data instances are correctly classified into their respective burnout categories (low, moderate, and high), with only a small number of misclassifications. This demonstrates that the C5.0 algorithm is highly effective in predicting student burnout risk based on the given attributes.

However, it is observed that certain categories, particularly the “Low” class, appear less frequently in the prediction results. This may be influenced by the imbalance of the dataset, where the number of samples in this class is significantly smaller compared to others. Despite this limitation, the model still shows strong overall performance.

Despite these promising results, this study has certain limitations. The dataset used only involves respondents from Universitas Islam Negeri Sumatera Utara, which may limit the generalizability of the findings. In addition, external factors that were not included in this study may also influence the classification results. Nevertheless, the findings of this study are consistent with previous research, which shows that the C5.0 algorithm is effective for classification tasks. Therefore, this study contributes to the application of the C5.0 algorithm in predicting student burnout and can serve as a reference for future research..

4. CONCLUSION

This study successfully developed a predictive model for student burnout risk using the C5.0 algorithm, which demonstrated strong classification performance and the ability to identify key factors influencing burnout, such as Cynicism (CY) and Emotional Exhaustion (EE). The results indicate that the model is effective in classifying burnout

levels into low, moderate, and high categories and can support early detection of student burnout. The main contribution of this study lies in providing an interpretable and accurate classification model based on both psychological and external variables. However, this study is limited by the dataset, which only includes students from Universitas Islam Negeri Sumatera Utara, and may not fully represent broader populations. Future research is recommended to expand the dataset, include additional relevant variables, and compare the performance of the C5.0 algorithm with other classification methods to improve prediction accuracy and generalization.

REFERENCES

- [1] R. Rajfalni Amarsa, R. Nurhudi Ramdhani, A. Taufiq, And I. Saripah, "Kecenderungan Academic Burnout Pada Mahasiswa Bimbingan Dan Konseling Universitas Pendidikan Indonesia," *Jurnal Bimbingan Dan Konseling*, Vol. 7, No. 2, 2023.
- [2] T. Harlia Putri Et Al., "Faktor Yang Mempengaruhi Academic Burnout Pada Mahasiswa Keperawatan," 2023.
- [3] C. Maslach And M. P. Leiter, "Understanding The Burnout Experience: Recent Research And Its Implications For Psychiatry," *World Psychiatry*, Vol. 15, No. 2, Pp. 103–111, Jun. 2021, Doi: 10.1002/Wps.20311.
- [4] S. Aminah, "Analisis Burnout Akademik Mahasiswa Pada Pembelajaran Daring : Analisis Pemodelan Rasch," *Psychocentrum Review*, Vol. 4, No. 2, Pp. 228–237, Jul. 2022, Doi: 10.26539/Pcr.421099.
- [5] R. Aulia And M. Fadli, "Perbandingan Algoritme C5.0 Dan Random Forest Menggunakan Data Bank Marketing," 2024.
- [6] V. Meisella, R. Rahmahwati, And Y. E. Prawatya, "Pengukuran Beban Mental Dan Tingkat Burnout Mahasiswa Selama Melakukan Perkuliahan Online Menggunakan Metode Nasa-Tlx Dan Maslach Burnout Inventory-Student Survey (Mbi-Ss)," 2022.
- [7] D. Permatasari, L. Latifah, And P. R. Pambudi, "Studi Academic Burnout Dan Self-Efficacy Mahasiswa," *Jurnal Prakarsa Paedagogia*, Vol. 4, No. 2, Dec. 2021, Doi: 10.24176/Jpp.V4i2.7418.
- [8] A. A. Wasito And H. N. Yoenanto, "Pengaruh Academic Self-Efficacy Terhadap Academic Burnout Pada Mahasiswa Yang Sedang Mengerjakan Skripsi," 2021. [Online]. Available: [Http://E-Journal.Unair.Ac.Id/Index.Php/Brpkm](http://E-Journal.Unair.Ac.Id/Index.Php/Brpkm)
- [9] N. Made, M. I. Sari, G. Ayu, A. Antari, And I. A. Sanjiwani, "Hubungan Tipe Kepribadian Dengan Burnout Pada Mahasiswa Tahun Pertama Pssikpn Fk Unud," 2021.
- [10] Q. Rasyidah, R. Fitriyari, And A. S. Wahyudi, "Hubungan Tingkat Stres Dan Burnout Syndrome Dengan Kualitas Tidur Pada Mahasiswa Profesi Ners Fakultas Keperawatan Universitas Airlangga (A Relationship Between Stress Level And Burnout Syndrome With Sleep Quality On Clinical Nursing Students In Faculty Of Nursing Universitas Airlangga)," 2020, Doi: 10.20473/Pnj.V1i1.19209.
- [11] J. Pranoto And D. Wibowo, "Peran Self Efficacy Terhadap Burnout Yang Dimoderatori Jenis Kelamin Pada Mahasiswa Bekerja Paruh Waktu," *Jurnal Psikologi Udayana* 2023, Vol. 10, No. 1, Pp. 286–296, 2023, Doi: 10.24843/Jpu/2023.V10.I01.P08.
- [12] I. G. Sudipa Et Al., *Data Mining*. 2023. [Online]. Available: [Www.Globaleksekitifteknologi.Co.Id](http://www.Globaleksekitifteknologi.Co.Id)
- [13] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Jurnal Simetris*, Vol. 11, No. 2, 2020.
- [14] F. N. Umma, B. Warsito, D. Asih, And I. Maruddani, "Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 Di Kabupaten Pematang," Vol. 10, No. 2, Pp. 221–229, 2021, [Online]. Available: [Https://Ejournal3.Undip.Ac.Id/Index.Php/Gaussian/](https://Ejournal3.Undip.Ac.Id/Index.Php/Gaussian/)
- [15] R. L. M. Apriyadi, B. Efendi Damanik, S. Informasi, And S. Tunas Bangsa Pematangsiantar, "Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring," *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, Vol. 11, No. 1, 2022.
- [16] D. Hartianto And H. Erikson, "Optimasi Prediksi Kelulusan Mahasiswa Dengan Menggunakan C5.0 Dan Regresi Linear," 2022.
- [17] M. Zainuri, M. H. Fahmi, And R. A. Hamdhana, "Komparasi Metode Klasifikasi Algoritma C5.0 Dan Naïve Bayes Untuk Menentukan Jurusan Siswa," Vol. 1, No. 1, Pp. 1–8, 2022.
- [18] S. Ayumida, M. S. Azis, And Z. G. Fiano, "Implementasi Program Administrasi Pembayaran Berbasis Dekstop (Studi Kasus: Sma Negeri 1 Cikampek)," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, Vol. 15, No. 2, Pp. 30–41, Jul. 2020, Doi: 10.35969/Interkom.V15i2.70.
- [19] P. N. Amir, M. Fatchan, And Edora, "Prediction Of Student Graduation Using The C5.0 Algorithm Approach At Sman 2 South Cikarang," *Jurnal Ilmiah Intech : Information Technology Journal Of Umus*, Vol. 5, No. 01, Pp. 19–31, 2023.
- [20] J. R. Fauzi, "Algoritma Dan Flowchart Dalam Menyelesaikan Suatu Masalah Disusun Oleh Universitas Janabadra Yogyakarta 2020," 2020.