



Implementation of the C4.5 Algorithm to Build A Prediction Model for Student Success in Database Courses

Nanda Pratama Alfyardri¹, Hari Antoni Musril², Sarwo Derta³

^{1,2,3} Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia

Article Information

Article History:

Submitted: September 26, 2025

Revision: November 23, 2025

Accepted: December 3, 2025

Published: December 11, 2025

Keywords

C4.5 Algorithm
 Education Data Mining
 Prediction Model
 Decision Tree
 Student Success

Correspondence

E-mail: nandapraalf@gmail.com

A B S T R A C T

This study aims to implement the C4.5 algorithm to build a model for predicting student success in database system courses in the Informatics and Computer Engineering Education study program at UIN Sjech M. Djamil Djambek Bukittinggi. Using the Knowledge Discovery in Database (KDD) approach, this study includes the stages of data selection, cleaning, transformation, modeling, and evaluation. Secondary data from the academic information system of students enrolled from 2018 to 2023 included 1,177 entries, which after cleaning resulted in 1,030 valid data. Predictor attributes consisted of academic factors such as Algorithm Logic scores, 1st semester Grade Point Average (GPA), attendance, and credit load, as well as non-academic factors such as gender and UKT (Tuition Fee Category). The target variable was student success status. Modeling was performed using Altair RapidMiner 2025 software with the C4.5 algorithm, resulting in a decision tree model. Evaluation showed an accuracy of 82.10%, recall of 69.58%, and precision of 62.51%, indicating the algorithm's effectiveness in classifying students as potentially successful or unsuccessful. This model identifies the most influential attributes, both academic and non-academic, on student success. Overall, the application of the C4.5 algorithm supports Educational Data Mining (EDM) in higher education, helping study programs improve the quality of learning and the effectiveness of data-based academic interventions.

This is an open access article under the CC-BY-SA license



1. Introduction

Student success in completing their studies on time is a key indicator of higher education quality. This success not only impacts students individually in terms of academic achievement and readiness for the workforce but also contributes to the reputation and accreditation of the study program. However, many higher education institutions still face challenges related to high rates of delayed study completion and student

failure in core courses. This situation poses a serious challenge for educational institutions to provide adaptive, inclusive, and targeted academic services that effectively support student success.

One factor influencing student success is the institution's ability to detect the risk of failure early. Early detection is crucial so that students potentially facing learning difficulties can promptly receive mentoring or academic intervention. Unfortunately, these detection efforts often rely solely on traditional evaluations – such as final exams or attendance reports – which are inherently reactive. In fact, by leveraging existing academic data within the campus information system, study programs can conduct predictive analysis to identify at-risk students even before failure occurs [1][2][3].

As information technology advances, higher education institutions have gained access to vast amounts of data encompassing aspects such as attendance, grades, academic activities, and student socioeconomic backgrounds. Unfortunately, this abundant data is rarely utilized optimally to support strategies for improving learning success. Typically, the data is merely stored as administrative archives without further processing. However, through an analytical approach, this data can become an invaluable source of new knowledge for enhancing the quality of learning [4].

One relevant analytical approach is Educational Data Mining (EDM), which is the application of data mining techniques in the field of education. EDM aims to discover hidden patterns in student academic data that can be used to predict learning performance, study success, and the risk of failure [5]. With the aid of EDM, both lecturers and study program administrators can design more data-driven and personalized learning strategies [6]. Several studies have demonstrated that EDM is effective in building early warning systems and supporting more accurate academic decisions.

In the implementation of EDM, the selection of a classification algorithm is a critical success factor. One widely used algorithm is C4.5, which generates a decision tree model based on the principles of entropy and information gain ratio [7]. The C4.5 algorithm offers advantages in handling both numerical and categorical data, managing missing values, and producing highly interpretable models [8]. These characteristics are particularly relevant for educational research, as the resulting model is not only accurate but also explainable to non-technical stakeholders, such as lecturers and academic administrators.

The Database Systems course in the Informatics and Computer Engineering Education Study Program is a core subject that significantly influences student academic success. This course provides the foundational knowledge required for advanced subjects such as Software Engineering and Information Systems. However, data indicates that a number of students still encounter difficulties and fail to complete this course. This not only impedes students' academic progress but also adversely affects the overall effectiveness of the study program. Consequently, a predictive system capable of early detection of students at risk of failing this course is essential. This research was conducted specifically at the State Islamic University of Sjech M. Djamil Djambek Bukittinggi within the Informatics and Computer Engineering Education Study Program. Student data from several academic cohorts was utilized to enable a more comprehensive analysis.

The previous researches have proved that the C4.5 algorithm can provide good results in predicting students' academic success. One study reported an accuracy of up to 93.94% using this algorithm [9]. Meanwhile, another study comparing C4.5 with K-Nearest Neighbor (KNN) showed that C4.5 achieved an accuracy of 73.49%, a precision of 64.62%, and a recall of 41.89% [10]. This variation in results indicates that the algorithm's performance is influenced by the characteristics of the analyzed dataset, yet it still affirms the potential of C4.5 as a reliable classification method in the educational context.

The primary distinction of this research lies in its specific focus on the Database Systems course, which presents a compelling case for investigation. Furthermore, it utilizes a substantial dataset of students from the research location. Unlike previous studies that often employed smaller or more generic datasets, this study analyzes historical data spanning multiple years to enhance the model's accuracy and generalizability.

The main reason for conducting this research is to provide data-based solutions to the problem of student failure in Database Systems courses. By building a prediction model, study programs can perform early detection and more targeted academic intervention. In addition, this research also aims to demonstrate the use

of RapidMiner software as a user-friendly and effective analysis tool. RapidMiner enables data processing and implementation of the C4.5 algorithm without requiring advanced programming skills, making the research results easier to replicate and implement by universities [11]. This is in line with the position of the research as a continuation and development of previous studies, with practical contributions to the application of EDM in Indonesian higher education, which is rarely explored in international literature. This position places the research as a bridge between EDM theory and local applications, filling a gap in data-driven academic intervention strategies.

Based on the aforementioned rationale, the objectives of this study are to implement the C4.5 algorithm in building a predictive model for student success in the Database Systems course using RapidMiner software. The resulting model is expected to assist the study program in early identification of students at risk of failing, thereby enabling the design of more effective academic intervention strategies. Consequently, this research aims to provide not only a theoretical contribution to the development of predictive modeling in education but also a practical contribution through the enhancement of instructional quality, the improvement of on-time graduation rates, and the strengthening of the study program's academic reputation.

Data Mining

Data mining is the process of extracting patterns, knowledge, or valuable information from large data sets using analytical techniques, statistics, and artificial intelligence algorithms [12]. The goal is to transform raw data into meaningful information to support decision making [13][14].

Educational Data Mining

Educational Data Mining (EDM) is a field that utilizes data mining and machine learning techniques to analyze educational data and predict academic performance [15]. EDM contributes to the improvement of learning quality by identifying learning patterns and factors influencing educational outcomes [16][17]. Consequently, stakeholders in academic institutions can benefit from EDM in implementing early preventive measures and making data-driven decisions based on patterns identified within the data.

Algoritma C4.5 Decision Tree

The C4.5 algorithm is an extension of ID3, designed to construct classification models based on decision trees [18][19]. These algorithms have improved C4.5's performance to some extent, with each possessing distinct advantages and limitations [20]. Generally, the fundamental concept of this algorithm has not undergone significant changes, and no groundbreaking innovations have been introduced [21]. This method not only classifies data into specific categories but also generates a decision structure that can support analytical processes and problem-solving, such as in credit evaluation or risk assessment [22]. It performs the process of separating data into more homogeneous groups by determining the attributes that have the highest level of information using measures such as entropy and information gain [23].

Entropy is expressed in equation (1) as follows:

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

Description:

S : set of cases or data

n : number of partitions or classes

p_i : proportion of the number of cases S_i to the total S

The stages of decision tree formation using the Decision Tree algorithm can be explained through the following Equation (2):

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

Description:

S : case set or overall data

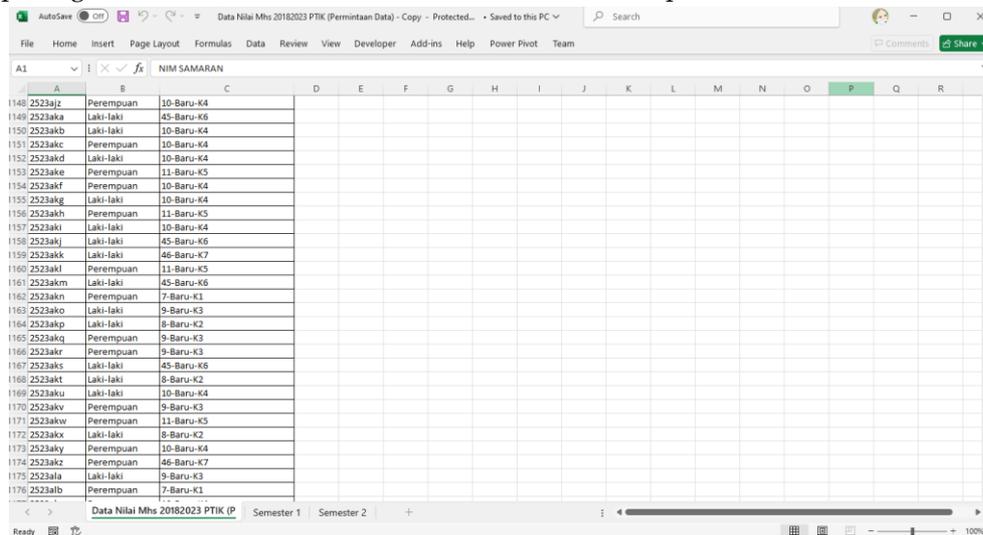
- A : attributes currently being tested
- n : number the partitions of the A attribute
- $|S_i|$: number of cases on partition i
- $|S|$: total number of cases in the set S

2. Method

This reserach uses the Knowledge Discovery in Database (KDD) approach, which includes the stages of data collection, data selection, cleaning and transformation, as well as modeling and evaluation. The type of research is applied research with the aim of producing a model for predicting student success using the C4.5 algorithm in RapidMiner software.

1.1 Data and Attribute Collection Process

The data used in this study consisted of secondary data obtained from the academic information system of UIN Sjech M. Djamil Djambek Bukittinggi, managed by the Information Technology and Database Division of UIN Sjech M. Djamil Djambek Bukittinggi. The dataset included student records from the Informatics and Computer Engineering Education Study Program for the academic years 2018–2023, comprising a total of 1,177 entries distributed across three separate sheets.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1148	2523ajz	Perempuan	10-Baru-K4														
1149	2523aka	Laki-laki	45-Baru-K6														
1150	2523akb	Laki-laki	10-Baru-K4														
1151	2523akc	Perempuan	10-Baru-K4														
1152	2523akd	Laki-laki	10-Baru-K4														
1153	2523ake	Perempuan	11-Baru-K5														
1154	2523akf	Perempuan	10-Baru-K4														
1155	2523akg	Laki-laki	10-Baru-K4														
1156	2523akh	Perempuan	11-Baru-K5														
1157	2523aki	Laki-laki	10-Baru-K4														
1158	2523akj	Laki-laki	45-Baru-K6														
1159	2523akk	Laki-laki	46-Baru-K7														
1160	2523akl	Perempuan	11-Baru-K5														
1161	2523akm	Laki-laki	45-Baru-K6														
1162	2523akn	Perempuan	7-Baru-K3														
1163	2523ako	Laki-laki	9-Baru-K3														
1164	2523akp	Laki-laki	8-Baru-K2														
1165	2523akq	Perempuan	9-Baru-K3														
1166	2523akr	Perempuan	9-Baru-K3														
1167	2523aks	Laki-laki	45-Baru-K6														
1168	2523akt	Laki-laki	8-Baru-K2														
1169	2523aku	Laki-laki	10-Baru-K4														
1170	2523akv	Perempuan	9-Baru-K3														
1171	2523akw	Perempuan	11-Baru-K5														
1172	2523akx	Laki-laki	8-Baru-K2														
1173	2523aky	Perempuan	10-Baru-K4														
1174	2523akz	Perempuan	46-Baru-K7														
1175	2523ala	Laki-laki	9-Baru-K3														
1176	2523alb	Perempuan	7-Baru-K3														

Gambar 1 Sheet 1 Data Mahasiswa PTIK TAahun 2018-2023

Figure 1 displays Sheet 1, which contains student data consisting of three attributes: Student Identification Number (NIM), gender, and tuition fee group.

Gambar 2 Nilai Semester 1 Mahasiswa PTIK tahun 2018-2023

Figure 2 shows the first semester grades of PTIK students for the 2018-2023 academic year, consisting of several attributes, namely Student ID Number, Course Code, Lecturer, Semester, Academic Year, Grade, Letter Grade, GPA, Credits, Attendance, Detailed Grades, and Total Credits.

Gambar 3 Nilai Semester 2 Mahasiswa PTIK 2018-2023

Figure 3 shows the second semester grades of PTIK students for the 2018-2023 academic year, which have the same attributes as the first semester grades, namely Student ID Number, Course Code, Lecturer, Semester, Academic Year, Grade, Letter Grade, GPA, Credits, Attendance, Detailed Grades, and Total Credits.

1.2 Data Selection

The initial step in the KDD methodology is data selection. This stage involves identifying and extracting the relevant data fields required to construct a predictive model [24][25]. These fields are selected because some existing fields are not needed. This process is carried out in order to select the appropriate attributes for building a model to predict student success in database system courses. Table 1 below shows the fields that will be used in processing data in the RapidMiner application.

Table 1 Field Terpilih

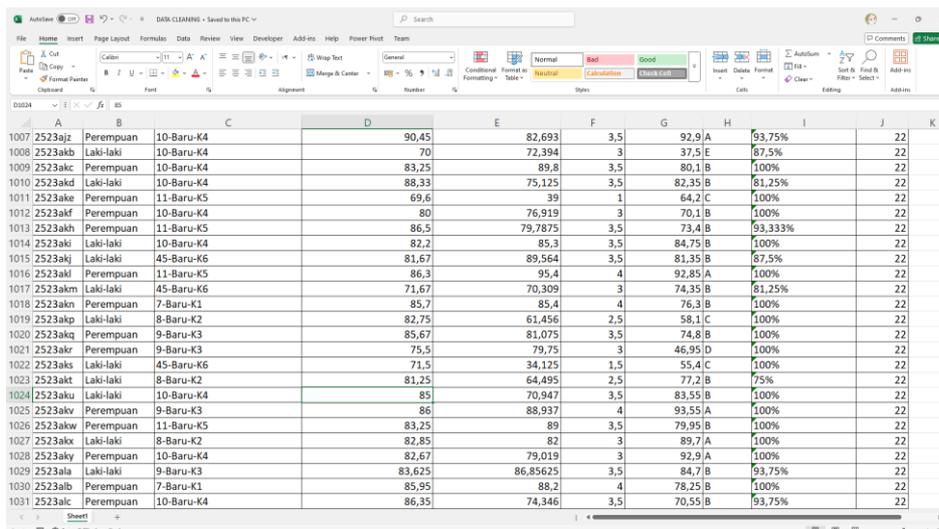
NO	Field	Keterangan
1	Jenis Kelamin	Field Fitur
2	Golongan UKT	Field Fitur

3	Nilai Paket Pemograman	Field Fitur
4	Nilai Paket Logika Algoritma	Field Fitur
5	IP Semester 1	Field Fitur
6	Nilai Sistem Basis Data	Field Target

The utilized fields comprise academic factors (Programming Package score, Logic and Algorithms score, first-semester GPA, credit load, attendance) and non-academic factors (gender, tuition fee group/UKT category). The target field is the success status in the Database Systems course.

1.3 Data Cleaning and Data Transformation

The second step in KDD is data cleaning. Data cleaning is the process of cleansing data aimed at selecting which data is needed to build a predictive model [26]. This includes handling empty (missing) data and redundant data.



Gambar 4 Hasil Data Cleaning

Figure 4 shows a reduction in the number of data points from 1,177 to 1,030 after the data cleaning process. This reduction of 147 data points occurred due to the presence of irrelevant or problematic data that could potentially compromise the accuracy of the predictive model. Consequently, only 1,030 data points were used in building the model to ensure more valid results.

The subsequent step is data transformation. Data transformation is the process of converting raw data into a suitable format to facilitate analysis. This stage typically includes normalization, encoding, and data simplification, enabling the algorithm to operate more effectively and produce a more accurate model [27].

Table 2 Data Transformation

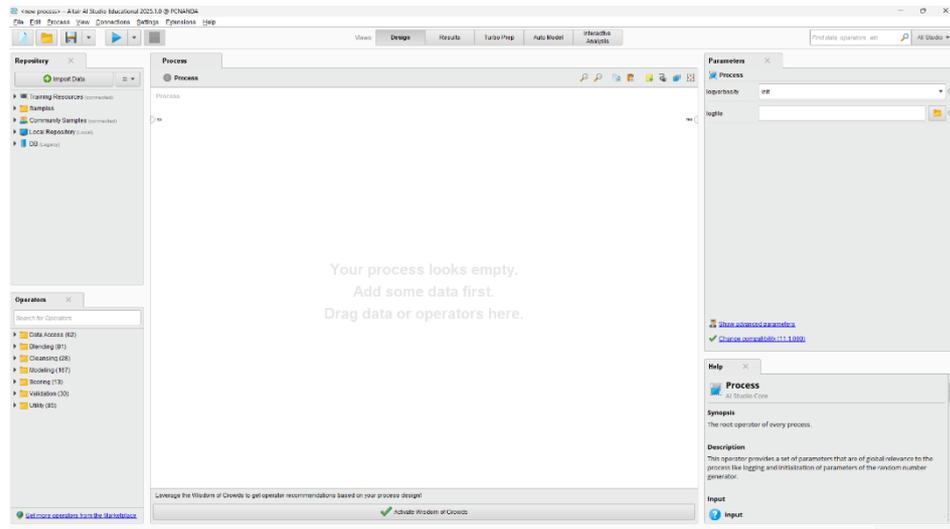
Kelompok	Atribut	Nilai Asli	Transformasi
Non Akademik	Jenis kelamin	Laki-Laki	L
		Perempuan	P
	Gol ukt	K1 dan K2	Rendah
		K3 dan K4	Sedang
		> K5	Tinggi
	Nilai PP	Beasiswa	Beasiswa
		85-100	A
		70-84	B
		55-69	C
		40-54	D
Akademik	Nilai LA	0-39	E
		85-100	A
		70-84	B
	IP smt 1	55-69	C
		40-54	D
		0-39	E
		3.5-4.0	Tinggi
		2.75-3.49	Baik
	BEBAN SKS	<2.75	Kurang
		23-24	Penuh
Kehadiran SBD	<22	Tidak penuh	
	100%	Penuh	
	75%-99%	cukup	
	<75%	kurang	
Nilai SBD	85-100	A	
	70-84	B	
	55-69	C	
	40-54	D	
	0-39	E	

1.4 Modeling and Evaluation

Model formation constitutes the stage of implementing the C4.5 algorithm in this research. Manual calculations were performed using Microsoft Excel, while the construction of the decision tree and model testing were conducted using Altair RapidMiner. A total of 1,030 cleaned data points were used for model development. This dataset was subsequently partitioned into training data and testing data. The split was performed with a proportion where the volume of training data exceeded that of the testing data. This approach ensures the resulting model possesses strong generalization capabilities and a higher accuracy rate. The subsequent procedure involves building the model using the Altair RapidMiner 2025 application, following these steps:

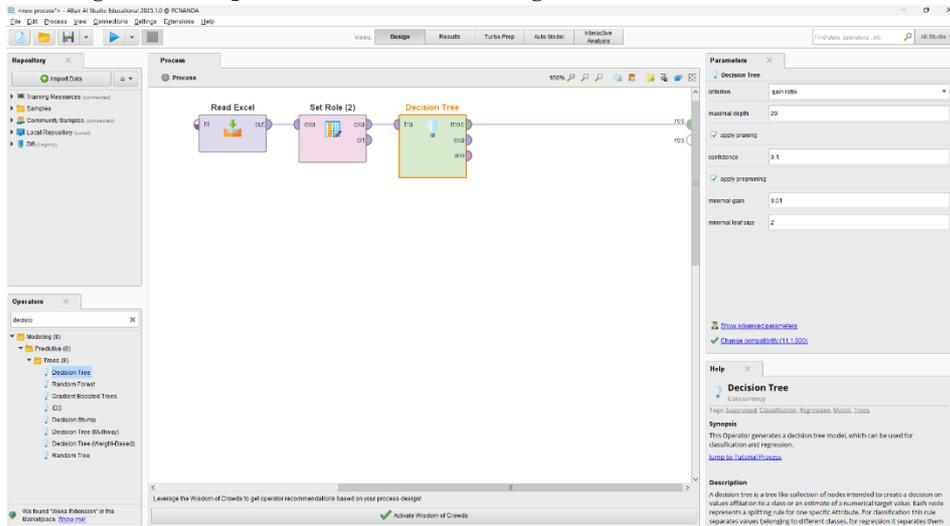
1.4.1 Processing training data

During model development, only the training data will be used. The first step is to launch Altair RapidMiner 2025. This is followed by selecting 'New Process,' with the initial interface appearing as shown in Figure 5.



Gambar 5 Tampilan Awal Rapid Miner

This was followed by arranging several operators, including read Excel, Set Role, and Decision Tree. The arrangement of operators can be seen in Figure 6.



Gambar 6 Susunan operator pembangunan model

1.4.2 Model Evaluations

The performance of the model is measured using three primary metrics derived from the confusion matrix: accuracy, precision, and recall. Each metric provides distinct insights into the model's capability to perform correct classifications [28]. The respective equations for these metrics are as follows [29].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

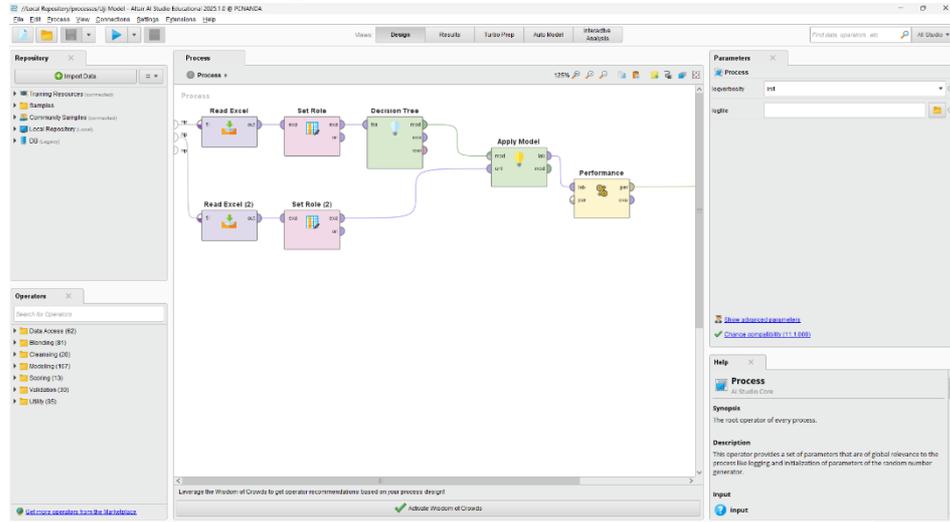
$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

Description:

TP (True Positive) : number of students predicted to succeed and actually succeeding

TN (True Negative) : number of students predicted to fail and those who actually failed
 FP (False Positive) : students who were predicted to succeed, but did not succeed
 FN (False Negative) : students who are actually successful, but are not detected by the model

Evaluation of the constructed model serves to assess the performance quality of the developed model[30]. The steps employed for model evaluation using Altair RapidMiner 2025 are illustrated in the configuration shown in Figure 7.



Gambar 7 Operator Evaluasi Model

3. Results and Discussion

The model was building use the C4.5 algorithm with the assistance of Altair RapidMiner 2025 software. The attributes used consisted of academic factors, namely Algorithm Logic score, first semester GPA, credit load, and attendance in the Database Systems course, as well as non-academic factors including gender and university tuition fee category. The target attribute was defined as the student's success status in the Database Systems course. The resulting model was a decision tree that clearly identified the most influential attributes on student success.

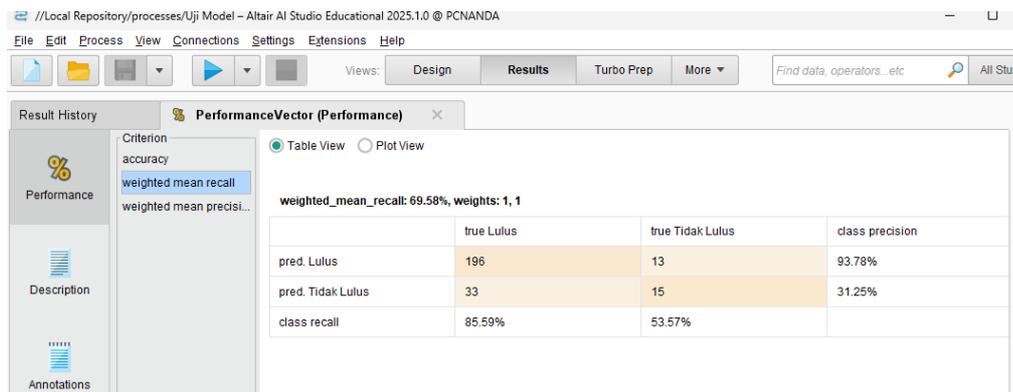
	true Lulus	true Tidak Lulus	class precision
pred. Lulus	196	13	93.78%
pred. Tidak Lulus	33	15	31.25%
class recall	85.59%	53.57%	

Gambar 8 Akurasi

Figure 8 shows that the prediction model built using the C4.5 algorithm demonstrates strong performance in identifying students who pass. The model successfully detected 85.59% of all students who actually passed (recall), and when predicting "Pass," its accuracy reached 93.78% (precision). This indicates the reliability of the model in classifying students who are likely to succeed in the Database course.

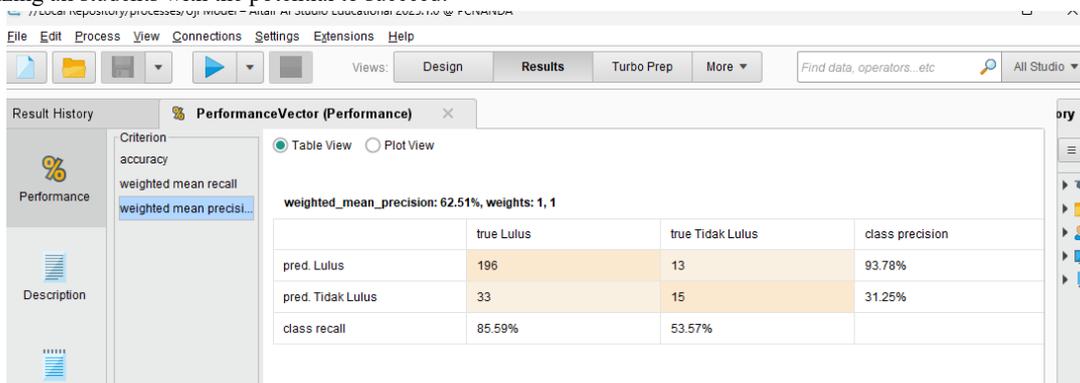
However, the model has limitations in predicting the "Fail" category. Only 53.57% of students who actually failed were successfully identified (recall), and of all the "Fail" predictions generated, only 31.25% were accurate (precision).

With an overall **accuracy of 82.10%**, it can be concluded that the model is more effective as a tool for identifying students with the potential to pass but is not yet optimal in detecting students at risk of failing.



Gambar 9 Recall

Furthermore, **Figure 9** shows a **recall value of 69.58%**. This recall, which is relatively lower than the accuracy, indicates that there are still some positive data points that were not correctly identified by the model. In other words, the model is not yet fully optimal in recognizing all students with the potential to succeed.



Gambar 10 Presisi

The precision value of 62.51%, as shown in Figure 10, indicates that of all positive predictions generated, only about 62.51% were actually correct. This indicates that there are still errors in predicting students who are considered successful when in fact they are not.

This study shows that the C4.5 algorithm is capable of achieving an accuracy of 82.10%, with a recall value of 69.58% and a precision of 62.51%. These values indicate that most of the test data can be classified correctly by the model, so that the model can be categorized as having a fairly good ability to recognize students who have the potential to succeed in the Database Systems course. This accuracy shows that the C4.5 algorithm has successfully learned the relationship patterns between attributes, such as algorithm logic values, first semester GPA, attendance rate, credit load, and non-academic factors, such as gender and UKT (Tuition Fee Category). Thus, the resulting model is able to represent the relationship between variables logically and consistently to predict student success.

When compared to previous studies that achieved higher accuracy levels, this difference in results can be explained by several factors, such as differences in the amount of data used, variations in predictor attributes, and differences in the data preprocessing stage [9]. Processes such as data cleaning, handling of missing values, normalization, and the division of training and test data significantly influence the performance of the final model. Additionally, an imbalanced class distribution between successful and unsuccessful students can also cause the model to be more inclined to recognize certain classes better, ultimately reducing the overall accuracy value. Nevertheless, an accuracy of 82.10% is still considered good in the context of educational classification, as it indicates that the model has a reasonably reliable predictive capability for use in academic decision-making.

When compared to other algorithms such as K-Nearest Neighbor (KNN), the C4.5 model in this study shows superiority in recall value, which is 69.58% compared to 41.89% in the KNN model [10] This shows that the model in this study has a better ability to recognize students who are truly successful. A high recall value also indicates that the model is more sensitive to positive cases, namely students who achieve success. Thus, even though its total accuracy is not as high as other models, C4.5 is superior in identifying students with the potential for success, which is an important aspect in the context of education because it can be used to support academic policies, such as providing additional guidance for students who are not yet performing optimally.

Furthermore, a precision value of 62.51% indicates that of all students predicted to succeed, approximately 62.51% were indeed successful according to the actual results. This value suggests that there are still some prediction errors in the positive category, where the model misclassifies unsuccessful students as successful. However, this error rate remains within acceptable limits and is not significant enough to substantially reduce the model's reliability. The relatively stable precision demonstrates the model's good generalization capability while maintaining a balance between correct and incorrect predictions. This precision value also indicates that the model not only focuses on high sensitivity but also maintains accuracy in its positive predictions.

The differences in accuracy, recall, and precision values between this study and previous research confirm that the performance of the C4.5 algorithm is significantly influenced by the characteristics of the dataset used. Factors such as variable diversity, data quality, and sample size play crucial roles in determining the model's effectiveness. Data with diverse, clean, and balanced variables increases the model's ability to produce more accurate classifications. Conversely, heterogeneous data or data containing class imbalances tend to make it difficult for the model to consistently recognize certain patterns.

4. Conclusion

This research successfully implemented the C4.5 algorithm to build a predictive model for student success in the Database Systems course using Altair RapidMiner 2025 software. The results obtained demonstrate that the model achieved an accuracy of 82.10%, a recall of 69.58%, and a precision of 62.51%. These values indicate that the C4.5 algorithm is reasonably effective in classifying students with potential success or risk of failure, although limitations in precision and recall remain. The resulting decision tree model was also able to identify the most influential attributes, both academic and non-academic, providing clear insights for lecturers and program administrators in determining learning strategies.

When compared to previous research, the performance of the model in this study shows variations in results influenced by data characteristics, sample size, and preprocessing stages. Although the accuracy level is not as high as in earlier research, the recall value obtained is relatively better compared to other studies, thereby strengthening the position of the C4.5 algorithm as a reliable classification method in the field of education. Thus, this research not only contributes to the theoretical development related to the application of Educational Data Mining but also provides practical benefits in the form of a predictive system that can help study programs conduct early detection of at-risk students, enabling more targeted academic interventions.

This research proves that the developed C4.5 model is a powerful classification tool. However, its performance can still be improved, particularly in detecting minority classes with imbalanced representation. Therefore, subsequent research is highly recommended to apply techniques such as SMOTE, ensemble methods, or other algorithms to enhance the model's sensitivity. This would allow interventions to be carried out more quickly and precisely. Ultimately, this research not only contributes theoretical insights to the field of Educational Data Mining but also offers practical solutions to improve the quality of learning.

References

- [1] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, 2022, doi: 10.1186/s40561-022-00192-z.
- [2] I. Iddrus and D. W. Sari, "Penerapan Data Mining Menggunakan Algoritma Decision Tree C4.5 Untuk Memprediksi Mahasiswa Drop Out Di Universitas Wiraraja," *J. Adv. Res. Inform.*, vol. 1, no. 02, pp. 1-7, 2023, doi:

- 10.24929/jars.v1i02.2684.
- [3] Y. E. Yuspita, R. Okra, and M. Rezeki, "PENERAPAN ALGORITMA KLASIFIKASI UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA MENGGUNAKAN RAPPIDMINER," *Djtechno J. Teknol. Inf.*, vol. 6, no. 1, pp. 376-388, 2025.
- [4] M. P. Colpo, T. Thompsen Primo, M. S. de Aguiar, and C. Cechinel, *Educational Data Mining for Dropout Prediction: Trends, Opportunities, and Challenges*, vol. 32. 2024. doi: 10.5753/rbie.2024.3559.
- [5] V. Fitriyanti, G. Testiana, and C. Eri Gunawan, "Klasifikasi Predikat Kelulusan Mahasiswa Menggunakan Algoritma C4.5," *J. SAINTEKOM*, vol. 14, no. 2, pp. 217-232, 2024, doi: 10.33020/saintekom.v14i2.626.
- [6] S. Sarker, M. K. Paul, S. T. H. Thasin, and M. A. M. Hasan, "Analyzing students' academic performance using educational data mining," *Comput. Educ. Artif. Intell.*, vol. 7, no. December 2023, p. 100263, 2024, doi: 10.1016/j.caeai.2024.100263.
- [7] A. W. Wicaksono and T. Setiadi, "Penerapan Klasifikasi Decision Tree (C4.5) untuk Memprediksi Kelulusan Siswa Sekolah Dasar di Kecamatan Juai," *Format J. Ilm. Tek. Inform.*, vol. 12, no. 2, p. 151, 2023, doi: 10.22441/format.2023.v12.i2.008.
- [8] M. Algoritma *et al.*, "Jurnal Processor Klasifikasi Pemilihan Siswa untuk Rekomendasi Beasiswa," vol. 20, no. 1, pp. 17-29, 2025.
- [9] I. W. M. Putri, R. Rusdah, L. Suryadi, and D. Anubhakti, "Prediction of Graduation for Students at the ISB Atma Luhur Faculty of Information Technology Using the C4.5 Algorithm," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, pp. 363-369, 2023, doi: 10.32736/sisfokom.v12i3.1731.
- [10] V. N. Juli, D. A. Putri, N. S. Prianita, and E. Rilvani, "Penerapan Metode C4 . 5 dan K-Nearest Neighbor untuk Klasifikasi Kelulusan Mahasiswa Berdasarkan Data Akademik," vol. 3, no. 4, pp. 226-267, 2025, doi: <https://doi.org/10.61132/jupiter.v3i4.1032>.
- [11] S. Aminah and Y. Yadi, "Application of the C4.5 Algorithm for Predicting Students' Learning Styles Based on Somatic, Auditory, Visual, and Intellectual Models," *Sinkron*, vol. 8, no. 4, pp. 2163-2171, 2024, doi: 10.33395/sinkron.v8i4.14032.
- [12] U. Al Faruq, M. Ainun Naja Fauzi, I. Fatayasya, E. Daniati, A. Ristyawan, and N. PGRI Kediri, "Prediksi Data Kelulusan Mahasiswa Dengan Metode Decision Tree menggunakan Rapidminer," in *Agustus*, 2023, pp. 131-138. [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/%0APrediksi>
- [13] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 1, pp. 193-199, 2023, doi: 10.37859/coscitech.v4i1.4755.
- [14] Y. Duan, "A Study of Prediction Accuracy of English Test Performance Using Data Mining and Analysis," *Ann. Emerg. Technol. Comput.*, vol. 7, no. 2, pp. 1-8, 2023, doi: 10.33166/AETiC.2023.02.001.
- [15] S. S.-L. Chen and X. Lin, "Application of Decision Tree Algorithm in Educational Data Mining," *Curric. Teach. Methodol.*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:259020691>
- [16] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Inf.*, vol. 14, no. 1, 2023, doi: 10.3390/info14010054.
- [17] M. A. Jassim, "Analysis of the Performance of the Main Algorithms for Educational Data Mining: A Review," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1090, no. 1, p. 012084, 2021, doi: 10.1088/1757-899x/1090/1/012084.
- [18] Yadi, "Implementation Algorithm C4.5 Classification Of Prospective Scholarship Recipients," *J. SimanteC*, vol. 11, no. 1, pp. 27-32, 2022.
- [19] L. Barbeiro, A. Gomes, F. B. Correia, and J. Bernardino, "A Review of Educational Data Mining Trends," *Procedia Comput. Sci.*, vol. 237, pp. 88-95, 2024, doi: 10.1016/j.procs.2024.05.083.
- [20] R. Akbar and S. 'Uyun, "Penentuan Bantuan Siswa Miskin Menggunakan Fuzzy Tsukamoto Dengan Perbandingan Rule Pakar dan Decision Tree (Studi Kasus : SDN 37 Bengkulu Selatan)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 651-662, 2021, doi: 10.25126/jtiik.0813191.
- [21] P. Li, F. Xiong, X. Huang, and X. Wen, "Construction and optimization of vending machine decision support system based on improved C4.5 decision tree," *Heliyon*, vol. 10, no. 3, p. e25024, 2024, doi: 10.1016/j.heliyon.2024.e25024.
- [22] Dhoni Hanif Supriyadi, F. B. Siahaan, S. Anwar, and F. W Handono, "Prediction of Customer Creditworthiness with the C4.5 Algorithm at PT Menara Indonesia Company," *JSAI (Journal Sci. Appl. Informatics)*, vol. 7, no. 3, pp. 512-521, 2024, doi: 10.36085/jsai.v7i3.7237.
- [23] D. Hendri, D. Nadha, F. K. Basri, M. F. Wajdi, and N. Nadhirah, "Comparison of Decision Tree Algorithm, Naive Bayes, K-Nearest Neighbors on Spotify Music Genre," *IJATIS Indones. J. Appl. Technol. Innov. Sci.*, vol. 1, no. 1, pp. 47-53, 2024, doi: 10.57152/ijatiss.v1i1.1219.
- [24] M. Daud, R. Juita, and C. D. Suhendra, "Penerapan Metode Algoritma C4.5 Untuk Klasifikasi Kelayakan Penerima Program Bantuan Pada Dinas Sosial Kabupaten Manokwari," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 271-278, 2025, doi: 10.51454/decode.v5i1.1057.
- [25] R. Ordoñez-Avila, N. Salgado Reyes, J. Meza, and S. Ventura, "Data mining techniques for predicting teacher evaluation in higher education: A systematic literature review," *Heliyon*, vol. 9, no. 3, 2023, doi: 10.1016/j.heliyon.2023.e13939.

- [26] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32-39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [27] T. Tukino, "Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia," *J. Sist. Inf. Bisnis*, vol. 9, no. 1, p. 39, 2019, doi: 10.21456/vol9iss1pp39-46.
- [28] D. Perez-Guaita, G. Quintas, Z. Farhane, R. Tauler, and H. J. Byrne, "Corrigendum: 'Data mining Raman microspectroscopic responses of cells to drugs in vitro using multivariate curve resolution-alternating least squares' (Talanta (2020) 208, (S0039914019310197), (10.1016/j.talanta.2019.120386))," *Talanta*, vol. 236, no. September 2021, p. 122682, 2022, doi: 10.1016/j.talanta.2021.122682.
- [29] M. M. Islam, F. H. Sojib, M. F. H. Mihad, M. Hasan, and M. Rahman, "The integration of explainable AI in Educational Data Mining for student academic performance prediction and support system," *Telemat. Informatics Reports*, vol. 18, no. May, p. 100203, 2025, doi: 10.1016/j.teler.2025.100203.
- [30] S. Mutrofin, M. M. Machfud, D. H. Satyareni, R. V. H. Ginardi, and C. Fatichah, "Komparasi Kinerja Algoritma C4.5, Gradient Boosting Trees, Random Forests, dan Deep Learning pada Kasus Educational Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 4, p. 807, 2020, doi: 10.25126/jtiik.2020742665.