

# Application of the C45 Algorithm to Predict Student Academic Scores

Zulham Sitorus<sup>1</sup>, Andi Ernawati<sup>2\*</sup>, Ananda Aulia<sup>3</sup>, Ayu oftasari<sup>4</sup>

Master of Information Technology, University Pembangunan Pancabudi, Medan, Indonesia

Email: <sup>1</sup>zulhamsitorus@dosen.pancabudi.ac.id, <sup>2</sup>aernawati296@gmail.com, <sup>3</sup>anandaaulia30@gmail.com, <sup>4</sup>ayuofta2849@gmail.com

Correspondence Author Email: [aernawati296@gmail.com](mailto:aernawati296@gmail.com)

**Abstract** – Student grades are the results of teaching and learning activities on a campus. So you can know your target for completing your studies. This research uses the C4.5 Algorithm which can help predict the results of student assessments. The dataset consists of student achievement index, place of residence, discipline, lecturer's role in lectures. From 40 datasets we have obtained a decision on student academic achievement and obtained performance results from accuracy results of 86.36% with class precision predicate Yes=84.62%, No=88.89% and class recall Yes=91.67%, No=80.00%.

**Keywords:** Mark; Academic; C4.5; Algorithm

## 1. INTRODUCTION

The application of the C45 algorithm in predicting student academic grades is an approach that is very relevant in the educational context. The C45 algorithm, also known as the decision tree algorithm, is a machine learning method used to build predictive models based on a set of decision rules that describe the relationship between input data features and the desired output[1]–[3]

This approach can specifically be applied in the context of predicting student academic grades by using historical data such as academic history, extracurricular activities, personal characteristics, and so on as input features. By using the C45 algorithm, we can build a prediction model that can determine student academic grades based on a combination of these features.

The use of the C45 algorithm in predicting student academic grades has become an interesting research topic in the field of information technology and education. The C45 algorithm is a powerful and well-known classification method in data mining applications. In this research, we will discuss the implementation of the C45 algorithm for a student academic prediction system and its benefits in improving the quality of education.[4]–[7]

The use of the C45 algorithm in predicting student academic grades has several advantages. One of the main advantages is the C45 algorithm's ability to handle continuous and nominal data. Because almost all of the scholarship recipient criteria attributes used are of continuous data type, the C45 algorithm is very effective in building scholarship recipient decision trees.

The aim of research on implementing big data analysis using the C45 algorithm for a student academic prediction system is to create a decision tree model that can predict student academic grades with high accuracy. By using the C45 algorithm, this research can help improve the quality of education by predicting student academic scores more accurately and on time. This can help the study program to pay special attention to students who are predicted not to graduate on time so that these students can improve their achievement index each semester so they can graduate on time.

This research uses data mining methods to collect and analyze student data. The data used in this research is student academic data, including GPA and National Examination scores. The C45 algorithm is used to build a decision tree that can predict student academic grades based on the data collected. The results of this research show that the C45 algorithm can be used well to predict student academic grades with high accuracy.

In conclusion, the application of the C45 algorithm to predict student academic grades can help improve the quality of education by predicting student academic grades more accurately and on time. The C45 algorithm has several advantages, including the ability to handle both continuous and nominal data, which makes it very effective in building scholarship recipient decision trees. Thus, this research can be used as a reference for study programs to improve the quality of education by predicting student academic scores more accurately and on time.[8]–[10]

Research conducted by Supayundi Pratama et al. Year 2023 with the title Application of Data Mining to Predict Student Academic Achievement Using the C4.5 Algorithm with CRISP-DM with results of 67.06% data accuracy.[11].

Research carried out by Embun Pajar Wati, in 2022 with the title Application of the KNN, Naive Bayes and C4.5 Algorithms in Predicting Student Graduation with 100% accuracy results.[12]. Research conducted by Anita Desiani, 2020 with the title Prediction of Students' Cumulative Academic Achievement Index Level Using Data Mining Techniques. in this class it is 28.6% and Precision is 40%.[13]. Research conducted by Safitri Linawati et al. In 2020 with the title Prediction of Student Academic Achievement Using Random Forest and C4.5 Algorithms with results having an accuracy value of 92.4%, precision of 91.4% and recall of 92.4% higher than Decision Tree C4.5.[14].

Research carried out by Abdul Rohman et al. 2019 with the title Implementation of Data Mining Using the C4.5 Decision Tree Algorithm to Predict Student Graduation at Pandanaran University with Results. Producing Student Graduation Patterns/Models that can be used for decisions in higher education. The results of this research produced 10 rules with an accuracy value of 65.98% with an AUC value of 0.874 [1]

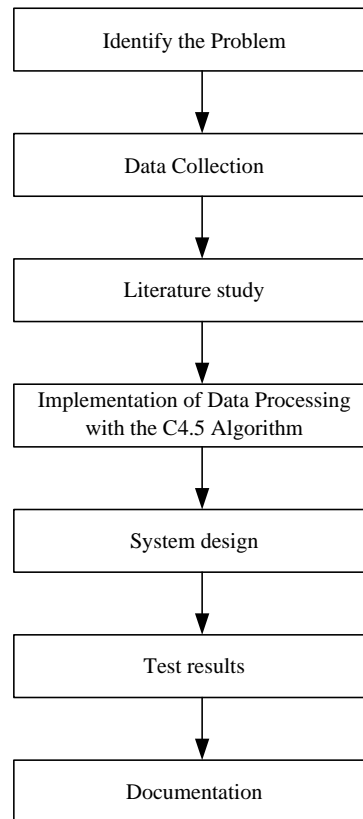
However, from previous research the author has a problem that needs to be re-examined using the C4.5 algorithm with a data set of 40 samples.

## 2. RESEARCH METHODOLOGY

### 2.1 Research Methods

The algorithm used in this research is C4.5, this method emphasizes the student's passing score. This research was conducted based on the problem formulation described in the previous chapter, namely to determine the classification of student graduation. This research was conducted at XYZ Campus. The research design used can be seen in the picture below.

### 2.2 Research Stages:



**Figure 1.** Research Stages

1. Identify the Problem  
Problem identification is carried out to explain and define the problems faced. At this stage, a solution to the problem is also sought.
2. Data Collection  
Data collection was carried out by looking for previous research that had been carried out to support solving the problems faced.
3. Literature study  
The author looks for theoretical references that are relevant to the cases or problems found
4. 4. Implementation of Data Processing with the C4.5 Algorithm  
This data processing is carried out when the data has been collected and tested using the C4.5 Algorithm
5. System design  
System design is a stage and process for defining an existing system or a new system.
6. Test results  
The test result is a result that has been implemented and has been approved by the program created. This stage is carried out to find out whether the system being built is running or not.
7. Documentation  
Documentation is evidence by displaying accurate files

### 2.3 Algorithm

The C4.5 algorithm was introduced by Quinlan in 1996 as an improved version of ID3. In ID3, Decision tree induction is only possible carried out on categorical type features (nominal or ordinal), while the numeric type (interval or ratio) cannot be used. The fix is: can not only handle type features categorical, but can also handle features with numeric type, and can also do cutting (pruning) decision trees, and decreasing (deriving) rule sets. Algorithm C4.5 [3], [15]–[17]

### 2.3 Data Mining

Data mining is the process of discovering useful new correlations, patterns and trends by mining large amounts of data repositories, using pattern recognition technologies such as statistics and mathematical techniques. [18]–[20]

## 3. RESULT AND DISCUSSION

### 3.1. Analysis of Method Application

The sample data that researchers used in this research was questionnaire data. Data was obtained from distributing questionnaires to 40 respondents from XYZ Campus. The application of data mining with the C4.5 Algorithm in this research is directly related to student data to explore student achievement, discipline and the role of lecturers in lectures. Like table 1.

**Table 1.** Sample Data Set.

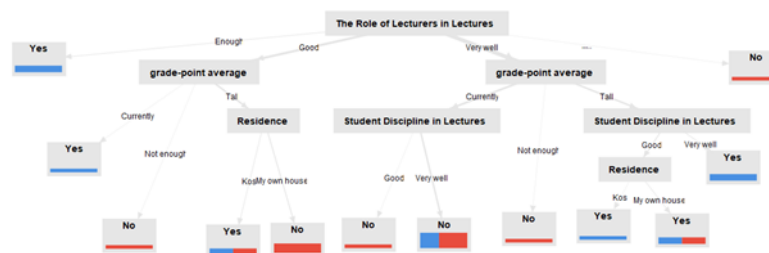
No	Study Program	Semester 1 Achievement Index	Semester 2	Residence	The Role of Lecturers in Coursework	Student Discipline in Lectures
1	Information Systems	3.41	3.48	Kos	2	3
2	Information Systems	3.64	3.71	Kos	3	3
3	Information Systems	3.64	3.6	Kos	3	3
4	Information Systems	3.61	3.75	Kos	3	3
5	Informatics Management	3.27	3.36	My own house	3	3
6	Informatics Management	3.43	3.3	My own house	4	4
7	Informatics Management	3.44	3.26	My own house	2	2
	Informatics Management	3.52	3.32	Kos	4	4
9	Informatics Management	3.43	3.35	Kos	4	3
10	Informatics Management	3.59	3.34	My own house	4	3
11	Informatics Management	3.29	3.22	My own house	4	4
12	Informatics Management	3.63	3.65	Kos	3	3
13	Informatics Management	3.9	3.75	My own house	3	2
14	Informatics Management	3.48	3.57	Kos	4	3
15	Informatics Management	3.40	3.47	Kos	4	4
16	Informatics Management	3.48	3.57	My own house	3	3
17	Informatics Management	3.40	3.47	Kos	4	4
18	Informatics Management	3.72	3.71	My own house	4	3
19	Informatics	3.30	3.48	Kos	4	4
20	Informatics Management	3.43	3.5	Kos	4	4
...	...	...	...	...	...	...
40	Informatics	3.33	3.32	Kos	4	4

In the Data Cleaning Stage, the dataset consists of 40 data that will be analyzed using Rapid Miner Decision Tree using the C4.5 Algorithm. The following is Table 2, the data that has been prepared.

**Table 2.** Preparation

grade-point average	Residence	The Role of Lecturers in Lectures	Student Discipline in Lectures	Nilai Mahasiswa
Currently	Kos	Enough	Good	Yes
Tall	Kos	Good	Good	Yes
Tall	Kos	Good	Good	No
Not enough	Kos	Good	Good	No
Tall	My own house	Good	Good	No
Tall	My own house	Very well	Very well	Yes
Tall	My own house	Enough	Enough	Yes
Currently	Kos	Very well	Very well	Yes
Tall	Kos	Very well	Good	Yes
Tall	My own house	Very well	Good	Yes
Tall	My own house	Very well	Very well	Yes
Currently	Kos	Good	Good	Yes
Tall	My own house	Good	Enough	No
Currently	Kos	Very well	Good	No
Currently	Kos	Very well	Very well	No
Tall	My own house	Good	Good	No
Not enough	Kos	Very well	Very well	No
Tall	My own house	Very well	Good	No
Currently	Kos	Very well	Very well	No
Currently	Kos	Very well	Very well	Yes
...	...	.....	.....	No
Currently	Kos	Very well	Very well	No

In the Data Cleaning Stage, the dataset consists of 40 data that will be analyzed using Rapid Miner Decision Tree using the C4.5 Algorithm. The following is Table 2, the data that has been prepared.



**Figure 1.** Decision Tree

From the results of the decision tree above, the following description is obtained

**Tree**

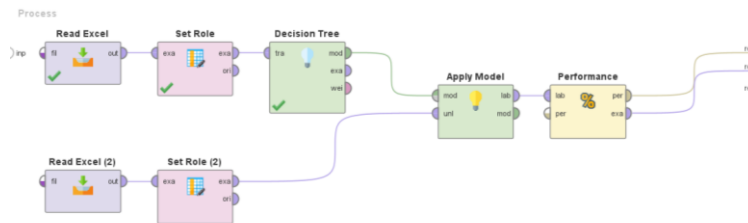
```

The Role of Lecturers in Lectures = Enough: Yes (Yes=2, No=0)
The Role of Lecturers in Lectures = Good
| grade-point average = Currently: Yes (Yes=1, No=0)
| grade-point average = Not enough: No (Yes=0, No=1)
| grade-point average = Tall
| | Residence = Kos: Yes (Yes=1, No=1)
| | Residence = My own house: No (Yes=0, No=3)
The Role of Lecturers in Lectures = Very well
| grade-point average = Currently
| | Student Discipline in Lectures = Good: No (Yes=0, No=1)
| | Student Discipline in Lectures = Very well: No (Yes=2, No=3)
| grade-point average = Not enough: No (Yes=0, No=1)
| grade-point average = Tall
| | Student Discipline in Lectures = Good
| | | Residence = Kos: Yes (Yes=1, No=0)
| | | Residence = My own house: Yes (Yes=1, No=1)
| | | Student Discipline in Lectures = Very well: Yes (Yes=2, No=0)
The Role of Lecturers in Lectures = .....: No (Yes=0, No=1)

```

**Figure 2.** Decision Tree Description

After we have created the Decision Tree, the classification model is tested using the C4.5 algorithm. This test is carried out by analyzing the performance level. Following are the test results on the data set



**Figure 3.** Testing with Rapid Miner

accuracy: 81.82%

	true Yes	true No	class precision
pred. Yes	8	2	80.00%
pred. No	2	10	83.33%
class recall	80.00%	83.33%	

**Figure 4.** Data Set Test Results

From the test results above, an accuracy of 81.82% was obtained with class precision predicate Yes=80.00%, No=83.33% and class recall Yes=80.00%, No=83.33%

**4. CONCLUSION**

In this research, XYZ Campus student data has been classified using the C4.5 decision tree algorithm method. The dataset consists of student achievement index, place of residence, discipline, lecturer's role in lectures. From 40 datasets we have obtained a decision on student academic achievement and obtained performance results from accuracy results of 81.82% with class precision predicate Yes=80.00%, No=83.33% and class recall Yes=80.00%, No=83.33%. This becomes a reference for lecturers and students to improve the learning process in lectures and student discipline in studying.

**REFERENCES**

- [1] A. Rohman and A. Rufiyanto, "Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran," *Proceeding SINTAK 2019*, pp. 134–139, 2019.
- [2] R. Kurniasari and A. Fatmawati, "Penerapan Data Mining dengan Algoritma C4.5 untuk Penentuan Jurusan Sekolah Menengah Atas," *J. INSYPRO (Information Syst. Process.*, vol. 6, no. 2, pp. 3–9, 2021, doi: 10.24252/insypro.v6i2.7912.
- [3] M. Chair, Y. N. Nasution, and N. A. Rizki, "Aplikasi Klasifikasi Algoritma C4.5 (Studi Kasus Masa Studi Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2008)," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 12, no. 1, p. 50, 2017, doi: 10.30872/jim.v12i1.223.

- [4] F. R. Wiratama, "Implementasi Algoritma C4 . 5 untuk Analisa Performa Pelayanan," *Explor. Inform.*, vol. 6, pp. 127–135, 2017.
- [5] A. Waqar, "Intelligent decision support systems in construction engineering: An artificial intelligence and machine learning approaches," *Expert Syst. Appl.*, vol. 249, p. 123503, Sep. 2024, doi: 10.1016/J.ESWA.2024.123503.
- [6] M. Auf der Landwehr, M. Trott, K. von Bülow, and C. von Viebahn, "Pyrrhic victory? Towards a decision support system for simulation projects in industrial practice based on Fuzzy cognitive mapping," *Comput. Ind. Eng.*, vol. 186, p. 109706, 2023, doi: <https://doi.org/10.1016/j.cie.2023.109706>.
- [7] Y. Kwon and Z. Lee, "A hybrid decision support system for adaptive trading strategies: Combining a rule-based expert system with a deep reinforcement learning strategy," *Decis. Support Syst.*, vol. 177, p. 114100, 2024, doi: <https://doi.org/10.1016/j.dss.2023.114100>.
- [8] Abdul Karim, "Implementasi Metode Multi-Objective Optimization On The Basis Of Ratio Analysis dalam Seleksi Mahasiswa Program Indonesia Pintar," *Bull. Comput. Sci. Res.*, vol. 3, no. 5, pp. 351–356, 2023, doi: 10.47065/bulletincsr.v3i5.283.
- [9] A. Karim, S. Esabella, M. A. Hanafiah, and ..., "Sistem Pakar Deteksi Penyakit Otitis dengan Perbandingan Metode Certainty Factor, Teorema Bayes, dan Dempster Shafer," *J. Media ...*, vol. 7, pp. 2003–2013, 2023, doi: 10.30865/mib.v7i4.6595.
- [10] A. Karim, S. Esabella, K. Kusmanto, S. Suryadi, and E. Purba, "Penerapan Metode Teorema Bayes Dalam Mendiagnosa Penyakit Autoimun," *Buuld. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 254–263, 2023, doi: 10.47065/bits.v5i1.3407.
- [11] S. Pratama, I. Iswandi, A. Sevtian, and T. P. Anjani, "Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma C4.5 dengan CRISP-DM," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 20–14, 2023, doi: 10.30871/jaic.v7i1.4998.
- [12] E. F. Wati and B. Rudianto, "Universitas Bina Sarana Informatika 1 Teknik Informatika, Universitas Nusa Mandiri ,2 Jl. Kramat Raya No.98, Senen, Jakarta Pusat 10450 1 Jl," *Raya Jatiwaringin*, vol. 11, no. 2, pp. 168–175, 2022.
- [13] A. Desiani, S. Yahdin, and D. Rodiah, "Prediksi Tingkat Indeks Prestasi Kumulatif Akademik Mahasiswa dengan Menggunakan Teknik Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1237–1244, 2020, doi: 10.25126/jtiik.2020722493.
- [14] S. Linawati, S. Nurdiani, K. Handayani, and L. Latifah, "Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest Dan C4.5," *J. Khatulistiwa Inform.*, vol. 8, no. 1, pp. 47–52, 2020, doi: 10.31294/jki.v8i1.7827.
- [15] M. Bari, S. H. Sitorus, and U. Ristian, "IMPLEMENTASI METODE NAÏVE BAYES PADA APLIKASI PREDIKSI PENYEBARAN WABAH PENYAKIT ISPA ( Studi Kasus : Wilayah Kota Pontianak )," vol. 06, no. 03, pp. 205–214, 2018.
- [16] L. Septiana, "Perancangan Sistem Pakar Diagnosa Penyakit Ispa Dengan Metode Certainty Factor Berbasis Android," *J. TECHNO Nusa Mandiri*, vol. XIII, no. 2, p. 89, 2016.
- [17] R. C. Menezes *et al.*, "Severe viral lower respiratory tract infections in Brazilian children: Clinical features of a national cohort," *J. Infect. Public Health*, vol. 17, no. 1, pp. 1–9, 2024, doi: <https://doi.org/10.1016/j.jiph.2023.09.015>.
- [18] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *J. Inform.*, vol. 10, no. 1, pp. 28–36, 2023, doi: 10.31294/inf.v10i1.15115.
- [19] A. Karim, "Implementation of the Multi-Objective Optimization Method on the Basic of Ratio Analysis ( MOORA ) and Entropy Weighting in New Employee Recruitment," vol. 5, no. 2, pp. 704–712, 2024, doi: 10.47065/josh.v5i2.4859.
- [20] N. Nurliyadi and A. Karim, "Sistem Pendukung Keputusan Penerimaan Analis Di Pusat Penelitian Kelapa Sawit Menggunakan Metode Complex Proportional Assessment (Copras)," *Bul. Ilm. Inform. Teknol.*, vol. 2, no. 1, pp. 32–42, 2023.