Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2024. Vol. 4. No. 1; e-ISSN: 2808-4519

Classification of Factors Causing the Decline in Student Learning Achievement in The Post-Pandemic Period Using the C4.5 Algorithm (Case Study: STMIK Kaputama)

Nazwa Nurul Hafizah Rangkuti[1*], A M H Pardede[2], Tioria Pasaribu[3]

[1, 2, 3]STMIK KAPUTAMA

nazwastabat2019@gmail.com[1], akimmhp@gmail.com[2], pasaributioria@gmail.com[3]*

Abstract

The COVID-19 pandemic that has hit Indonesia since 2020 has brought significant changes to various aspects of life, including the learning system in universities. Universities, which originally implemented face-to-face learning processes, were forced to adapt to online learning. However, this change causes various obstacles, especially for students who experience learning loss, namely a decrease in interest and motivation to learn which has an impact on academic achievement. This research aims to classify the factors that cause the decline in student learning achievement in the post-pandemic period at STMIK Kaputama using the C4.5 algorithm. Using data from STMIK Kaputama students as a sample, this research analyzes various factors such as access to technology, involvement in online learning, participation in face-to-face learning, social support, learning motivation, economic conditions, family, college, and academic stress levels. It is hoped that the results of this research will provide a deeper understanding of the dominant factors that cause a decline in learning achievement, as well as become a reference for educational institutions in developing strategies to overcome the negative impacts of online learning during the pandemic

Keywords: *C4.5 Algorithm, Decrease in Achievement, Data Classification*

1. Introduction

In 2020, almost all countries were affected by the COVID-19 outbreak, Indonesia was one of the countries exposed to this virus. Thepandemic period of the COVID-19 virus has a considerable impact on people's lives in Indonesia. The revocation of the Enforcement of Restrictions on Community Activities at the end of 2022 by President Joko Widodo indicates that Indonesia is at the end of the pandemic. This seems to be a breath of fresh air for various sectors, including the education sector. Since mid-2022, educational institutions at various levels have begun to carry out face-to-face academic and non-academic activities, including universities.

The return of students to schools and universities does not necessarily solve the problems in the education sector during the pandemic. The prolonged online learning system has caused many students to experience potential learning loss in the post-pandemic period. Learning loss is the loss of interest in learning due to the lack of interaction with educators during the learning process. The symptoms seen by children who experience learning loss are intellectual and skill decline, disrupted growth and development, experiencing psychological and psychosocial stress and gaps in access to learning, and one of them is decreased learning achievement [1].

Based on the description of the problem, this study aims to classify the factors that cause decreased student learning achievement in the post-pandemic period using the c4.5 algorithm, namely to find out what factors contribute to decreased student learning.

2. Literature Review

2.1. Data Mining

Data mining is one of the main parts or processes of Knowledge Discovery in Database (KDD), which is a form of activity that collects and uses past data to find regularities, patterns or relationships in a larger data set.

Data mining is also a logical process to find useful information. Once found information and patterns can be used for supporting tools in decision-making in developing businesses. Data mining tools can provide answers to a variety of business-related questions that are too difficult to solve. Data mining can also be used to forecast future trends that allow business people to make effective, proactive, and dynamic decisions.

2.2. Algorithm C4.5

The C4.5 algorithm is an algorithm used to generate a decision tree developed by Ross qiunlan. The basic idea of this algorithm is to create a decision tree based on the selection of attributes that have the highest priority or can be called having the highest gain value based on the entropy value of these attributes as the axis of classification attributes. At its stage the C4.5 algorithm has 2 working principles, namely: Creating a decision tree, and creating rules (rule model). The rules formed from the decision tree will form a condition in the form of if then [2].

2.3. The Covid-19 Pandemic

The COVID-19 pandemic has had a very significant impact on the education sector around the world. Schools, universities and other educational institutions have been forced to close their physical doors and switch to online learning in response to the spread of the virus. This sudden change requires students, teachers and parents to adapt quickly to new technologies and teaching methods. Limited access to technology is one of the main challenges, especially for students from economically disadvantaged backgrounds who do not have adequate computer equipment or internet connection. In addition, online learning also shows gaps in engagement and effectiveness, with some students showing increased stress and decreased motivation to learn due to social isolation and lack of face-to-face interaction with teachers and classmates [3].

2.4. Factors Decreasing Learning Achievement

Learning interest is a characteristic of a person in the form of a very high specific tendency to acquire knowledge and skills by learning. Student learning interest in participating in learning is something that is important in the smooth teaching and learning process. Students who have a high interest in learning in the learning process can support the teaching and learning process to get better, and vice versa, low student interest in learning will reduce the quality of learning and will affect learning outcomes [4], [5].

3. Research Methods

3.1. Research Methods

research methods are basically is a scientific way to get data with specific purposes and certain uses [6], [7].

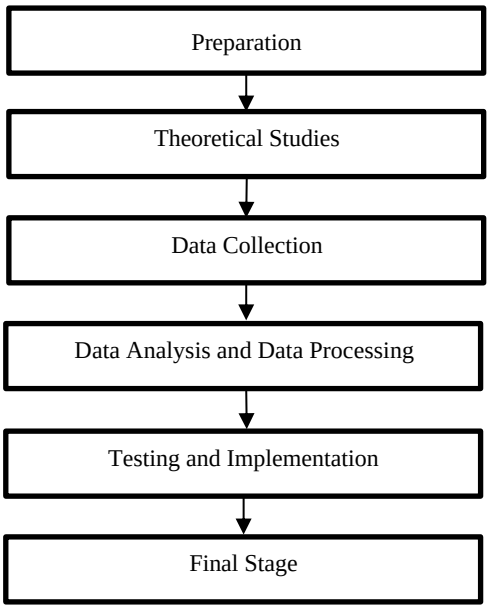

Fig. 1: Research Workflow

1. Preparation; This stage is the initial stage of the research process which will be carried out by determining the background of the problem, by looking for a problem and then formulating what problems are and how the process of solving them [4], [8].
2. Theoretical Studies; The At this stage a theoretical study of the problem will be carried out. existing problem. The study is carried out to determine the concepts that will be used in in research, especially about data mining, C4.5 algorithm, questionnaire data, rapidminer software that will be used in the process. questionnaire, rapidminer software that will be used in the analysis process carried out in this research. analysis process carried out in this study.
3. Data Collection
4. At This stage is intended to collect supporting data in order to can predict the Factors Causing Declining Learning Achievement Students in the Post-Pandemic Period.
5. Data Analysis and Data Processing
6. At this stage, supporting data will be analyzed, namely data-data on the Classification of Factors Causing Decreased Student Learning Achievement In the Post-Pandemic Period which has been obtained in the previous stage, grouping data on the Factors Causing Decreased Learning Achievement of Students in the Post-Pandemic Period using data mining techniques technique with the C4.5 algorithm as a problem solving method
7. Testing and Implementation

8. At this stage, data variable testing and data implementation will be carried out. data and the preparation of the system program
9. Final Stage
10. This stage is the stage of drawing conclusions and suggestions that can be done in the preparation of the thesis. With the conclusion, the
11. will be known the results of the entire thesis and it is hoped that with suggestions there will be improvements and benefits for others.
12. there will be improvements and benefits for others.

4. Results And Discussion

4.1. Results

In analyzing data in a study, supporting data is needed so that a study can run as expected. The data collection process is taken from a questionnaire distributed with the number of questions related to the factors causing the decline in learning achievement which has 9 questions and the data to be classified is the number of results of the questionnaire answers. The supporting data for research in this study are:

Table 1: Data table of factors decreasing learning achievement

Maha siswa	Prestasi Menurun	Akses Teknologi	Keterlibatan Daring	Partisipasi Tatap Muka	Dukungan Sosial	Motivasi Belajar	Kondisi Ekonomi	Tingkat Stres Akademik
1	Tidak	Baik	Tinggi	Rendah	Baik	Tinggi	Cukup	Rendah
2	Iya	Baik	Tinggi	Tinggi	Baik	Tinggi	Kurang	Rendah
3	Iya	Kurang	Tinggi	Tinggi	Kurang	Rendah	Kurang	Sedang
4	Tidak	Baik	Tinggi	Tinggi	Baik	Tinggi	Baik	Rendah
5	Iya	Cukup	Sedang	Sedang	Sedang	Sedang	Cukup	Sedang
6	Iya	Kurang	Sedang	Rendah	Sedang	Rendah	Kurang	Sedang
7	Tidak	Baik	Tinggi	Rendah	Sedang	Tinggi	Baik	Tinggi
8	Tidak	Kurang	Sedang	Tinggi	Baik	Sedang	Kurang	Rendah
9	Iya	Kurang	Rendah	Sedang	Sedang	Rendah	Kurang	Sedang
10	Tidak	Baik	Tinggi	Sedang	Kurang	Tinggi	Cukup	Rendah
11	Tidak	Baik	Sedang	Tinggi	Baik	Tinggi	Baik	Sedang
12	Iya	Cukup	Sedang	Sedang	Sedang	Sedang	Kurang	Sedang
13	Iya	Kurang	Rendah	Rendah	Kurang	Rendah	Cukup	Tinggi
14	Tidak	Baik	Tinggi	Tinggi	Baik	Tinggi	Kurang	Rendah
15	Iya	Cukup	Sedang	Rendah	Sedang	Sedang	Cukup	Sedang
16	Iya	Kurang	Rendah	Sedang	Sedang	Rendah	Cukup	Tinggi
17	Tidak	Cukup	Tinggi	Sedang	Sedang	Tinggi	Baik	Sedang
18	Tidak	Cukup	Tinggi	Tinggi	Sedang	Sedang	Cukup	Rendah
19	Iya	Cukup	Sedang	Rendah	Sedang	Rendah	Cukup	Tinggi
20	Tidak	Kurang	Tinggi	Tinggi	Kurang	Sedang	Baik	Rendah

next is to calculate the enthropy and gain values, with the following formula:

$$\text{Entropy (S)} = \sum_{i=1}^{n} - \pi * \log_2 \pi$$

gain formula:

$$\text{Gain (S, A)} = Entropy(S) \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy$$

Calculation of all attributes is done to get entropy and gain values, especially the highest gain value. The results of the calculation of all attributes are in table 4.2. In the table is the process for calculating the decision tree using the C4.5 algorithm method.

Table 2: Entropy and gain calculation results

Atribut		Jlh Kasus	naik	menurun	Enthropy	Gain
Total		**20**	**10**	**10**		
Faktor Akses Teknologi	Baik	7	6	1	0.591	0,2147
	Cukup	6	2	4	0.198	
	Kurang	7	2	5	0.863	
Faktor keterlibatan dalam pembelajaran daring	Tinggi	10	8	2	0.722	0.3636
	Sedang	6	2	4	0.918	
	Rendah	4	0	4	0	
Faktor Partisipasi Kembali ke pembelajaran tatap muka	Tinggi	8	6	2	0.722	0.18595
	Sedang	7	3	4	0.985	
	Rendah	5	1	4	0.722	
Dukungan Sosial						0.1645

	Baik	6	5	1	0.65	
	Sedang	10	3	7	0.881	
	Kurang	4	2	2	1	
Motivasi Belajar	Tinggi	8	7	1	0.543	0.4828
	Sedang	6	3	3	1	
	Rendan	6	0	6	0	
Kondisi Ekonomi	Baik	5	5	0	0	0.31635
	Cukup	8	3	5	0.954	
	Kurang	7	2	5	0.863	
Tingkat Stress Akademik	Rendah	7	6	1	0.591	0.21685
	Sedang	9	3	6	0.918	
	Tinggi	4	1	3	0.811	

Based on the gain calculation, the variable that has the highest gain is Learning Motivation with a gain of 0.4828. By using the C4.5 algorithm and the calculation of information gain, we can determine the main factors that affect student learning achievement after the pandemic. In this example, Learning Motivation is the variable with the greatest influence, followed by other variables such as Engagement in Online Learning and Economic Condition. The resulting decision tree helps in understanding how these factors interact to influence student learning outcomes.

4.2. Discussion

Here are the results of the decision tree using the rapid miner application

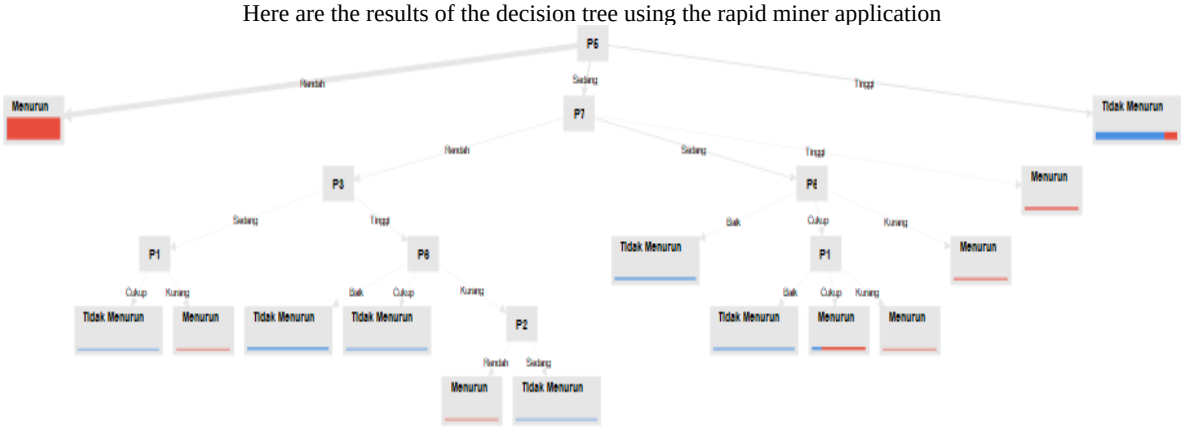

Fig. 2: Cluster Results from Data on Factors Affecting Student Learning Achievment

Description:
Student learning achievement is a low P5 (Learning Motivation) factor that affects the decline in learning achievement with the largest number of decreases as much as 121 data and those that do not decrease 1 data and if the P5 (Learning Motivation) factor is high students do not experience a decrease in learning achievement as much as 32 data.

5. Conclusions And Suggestions

5.1. Conclusions

1. The C4.5 algorithm is able to effectively identify and classify the factors that cause a decrease in student achievement in the post-pandemic period. By utilizing existing data, this algorithm builds a decision tree that maps the relationship between various student attributes. The C4.5 algorithm works by dividing data into smaller subsets based on the most significant attributes, resulting in a model that can be used to predict the main causal factors of decreased learning achievement.
2. From the results of the implementation of the C4.5 Algorithm, it is found that there is a main factor in the decline in student achievement is the low P5 (Learning Motivation) factor which affects the decline in learning achievement with the largest number of decreases of 121 data and those that do not decrease 1 data and if the P5 (Learning Motivation) factor is high students do not experience a decrease in learning achievement as much as 32 data.

5.2. Suggestions

The suggestions that will be a requirement to be better in writing this thesis and useful in the future are as follows:
1. It is hoped that the tests produced will be more developed with better results.
2. It is hoped that it can add more criteria or variables, in order to get even better results and information.
3. It is hoped that it can be developed again or combined with other methods in order to produce even better analysis or results in the future.

References

[1] Terayanti, Y. A. (2020). Pengaruh Pembelajaran Pada Anak Berkebutuhan Khusus Di Masa Pandemi Covid-19.

[2] Sukma, Halfis, dan Hermawan. (2019). Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5, Volume V No. 1 Februari 2019 P-ISSN 2442-2436, E-ISSN: 2550-0120
[3] Korompot, S., Rahim, M., & Pakaya, R. (2020). Persepsi siswa tentang faktor yang mempengaruhi minat belajar. JAMBURA Guidance and Counseling Journal, 1(1), 40–48
[4] Sugiyono, S. (2015). Metode Penelitian Kuantitatif Kualitatif Dan R & D Cetakan 17. Bandung: CV Alfabeta
[5] W. Rahayu, "Synthetic Minority Oversampling Technique (SMOTE) for Boosting the Accuracy of C4.5 Algorithm Model", *j. of artif. intell. and eng. appl.*, vol. 3, no. 3, pp. 624–630, Jun. 2024.
[6] M. Rudi Fanani and E. B. Agustina, "Implementation Of the C4.5 Algorithm With The Backward Elimination Feature Selection For MSME Product Sales Strategy", *j. of artif. intell. and eng. appl.*, vol. 3, no. 3, pp. 677–681, Jun. 2024.
[7] Rika Nur Adiha, "APPLICATION OF DATA MINING IN DETERMINING SOCIAL ASSISTANCE RECIPIENTS WITH C4.5 ALGORITHM ", *j. of artif. intell. and eng. appl.*, vol. 1, no. 2, pp. 118–127, Feb. 2022.
[8] Adeita A. Ndraha, J. T. Hardinata, and Y. P. Purba, "Application Of The C4.5 Algorithm To Determining Student's Level Of Understanding", *j. of artif. intell. and eng. appl.*, vol. 1, no. 2, pp. 162–167, Feb. 2022.