

## Data Mining for Decision Making System Determining Web-Based Class Promotion (Case Study: SMK Tarunatama Getasan)

Yonsi Rendra Hanggawan<sup>1</sup>, Suprihadi<sup>2</sup>

<sup>1,2</sup>Universitas Kristen Satya Wacana, Indonesia

672017144@student.uksw.edu, suprihadi@uksw.edu

### Abstract

*In this study we want to make an application to predict student graduation, where with the existence of We hope that this application will make it easier and there will be transparency in the graduation system. The method used is to use the C4.5 algorithm which is implemented in the form of WEB, where the algorithm will be processed in the form of PHP. The result of the application made is whether the student passed or not.*

### Keywords

decision making system;  
data mining; web



### I. Introduction

Data mining is a process of retrieving data from a large data set which is then grouped to obtain a definite data. In this era of technology, we need to improve human resources. We use data mining to improve students' passing scores and to solve the problem of taking grades. In this case study, we took data from report cards, remedial, absenteeism, and student activity scores. After grouping the data then we process the data.

A common problem for students in Vocational High Schools/SMA is low test scores that affect graduation results or grade promotions so that these students stay in class.

Prediction of increasing test scores is very influential on improving the quality of students who are educated by Vocational High Schools. The results of the classification of improving the learning system for vocational students are in the form of a decision tree which is expected to assist teachers in making decisions to improve the quality of education. To help make decisions for teachers, it is necessary to design an application for a web-based decision-making system for grade promotion.

Utilization of existing data in the information system to support decision-making activities, it is not enough to just rely on operational data, a data analysis is needed to explore the potential of existing information. The government's current system of raising students causes uncertainty about the increase in students which causes anxiety for both students and teachers.

With the new system implemented by the government, it is difficult for the community, especially the school, to predict the increase in their students.

In this study, several problems were formulated as follows: 1. How to design a data mining application to predict the increase in students at SMK Tarunatama Getasan?

2. How to make a data mining application that can present data quickly?

3. How can data mining applications be used to support accurate decisions?

4. How data mining applications are able to produce new knowledge in the form of useful information?

The goal to be achieved from the implementation and writing of this research is to produce an application to obtain useful information to predict the increase in students at SMK Tarunatama Getasan with data mining techniques and the c.45 algorithm.

In this research, I created a website that will be used for the implementation of the C4.5 algorithm. I use the web because it makes it easier for teachers to check the results of the C4.5 algorithm calculations.

## II. Review of Literature

In a study it is necessary to support the results of previous research related to the study. According to Mandala and Rahmanto, using data mining to predict the graduation of MAN students can be a good solution because it can avoid errors in data collection from students due to filling out data can be retrieved from the .xls file (Mandala, 2013)

Meanwhile, according to Kurniawan and Information, to predict a student's graduation rate, the student's support and confidence must first be sought. The source of the data is taken from the main value and graduation data which is then processed using the Apriori algorithm. The data is connected so that the results or graduation rates of students can be known (Kurniawan, 2016).

The rapid advancement of technology in the world of information in various fields has had a major impact on mankind. According to Pramusinto (2020) the power of technology including digitalization and automation continues to grow and change the pattern of production, distribution, and consumption. As with other areas of life, technology is used to make changes, so also with the legal system as technology in making changes (Hartanto, 2020). Meanwhile, the use of information technology is the benefit expected by users of information systems in carrying out their duties where the measurement is based on the intensity of utilization, the frequency of use and the number of applications or software used (Marlizar, 2021). To address the challenges and harness the opportunities offered by digital technologies during this crisis, participants shared a concern to recognize and protect digital rights in particular around the areas of privacy and inclusion (Hariati, 2021). This can be interpreted as internet users in Indonesia belongs to the category of digital natives group (Gunawan, 2020). In the field of education, for example, technology has helped facilitate the teaching and learning process for teachers or students by making the work process more efficient (Sulisto, 2017).

Predictions are needed for students who are indicated to not graduate on time so that further action can be given using data mining (Prambudi, 2019).

In his research, what is sought is the value of support and confidence from the relationship between graduation rates and data from students using the decision tree algorithm (Rohmawan, 2013).

In this study, I took the Tarunatama Vocational School because the school in determining grade promotion also used a point system not only using study results. Therefore I want to make a graduation system using the C4.5 algorithm so that the score is more accurate.

## III. Research Method

The research was conducted at a Vocational High School (SMK Tarunatama Getasan). The document used for this research is data on student graduation reports in grade 12.

Each school has a different scoring system, but still refers to the national education regulations that have been set by the government. Where the assessment can be done with tests, daily tests, mid-semester tests, end-of-semester tests and grade-up tests.

The method used for research with quantitative methods, using the C4.5 algorithm. The decision-making process in the classification of student graduation is divided into several assessment criteria, namely: cognitive scores, psychomotor scores, affective scores, attendance and number of HER subjects. These criteria are used in determining the attributes in the C4.5 algorithm.

How to calculate the C4.5 algorithm. To select an attribute, it is based on the Gain value of the existing attribute.

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

(1)

Description: S : Set of cases

n : Number of partitions attribute A

|S<sub>i</sub>| : Number of cases on partition i

|S| : Number of cases in S Before getting the Gain value is to find the Entropy value.

Entropy is used to determine how informative an attribute input is to produce an attribute.

The basic formula for Entropy (S) =  $-\sum_{i=1}^n p_i \log_2 p_i$

S: Set of Cases

n : Number of partitions S

p<sub>i</sub> : Proportion of S<sub>i</sub> to S The data to be processed are grouped in a special format which is then grouped into Typical data (K) using the formula

$$\text{Number of Classes (K)} = 1 + 3.3 \log n$$

Value n = number of data

$$K = 1 + 3.3 \log (100) = 8$$

There are 8 classes.

**Table 1.** Attributes of assessment

Variable	Values
Cognitive	1,2,3,4,5,6,7,8
Psychomotor Values	1,2,3,4,5,6,7,8
Active Values	Very Good, Good, Less
Value Absence	Very Good, Good, Lack
Attendance	High, medium, low
Number of subjects	0, 1-3, >3

### Calculation of Entropy and Gain

In the decision tree stage and based on research data, the results of the Entropy calculation are obtained as follows

**Table 2.** Entropy Calculation

Attribute	Entropy	Information
Cognitive	0.590422095	0 ,266726343
Psychomotor	0.440851994	0.416296443
Affective	0.75610123	0.101047208
Presence	0.608787484	0.248360953
Total HER	0.574881575	0.282266863

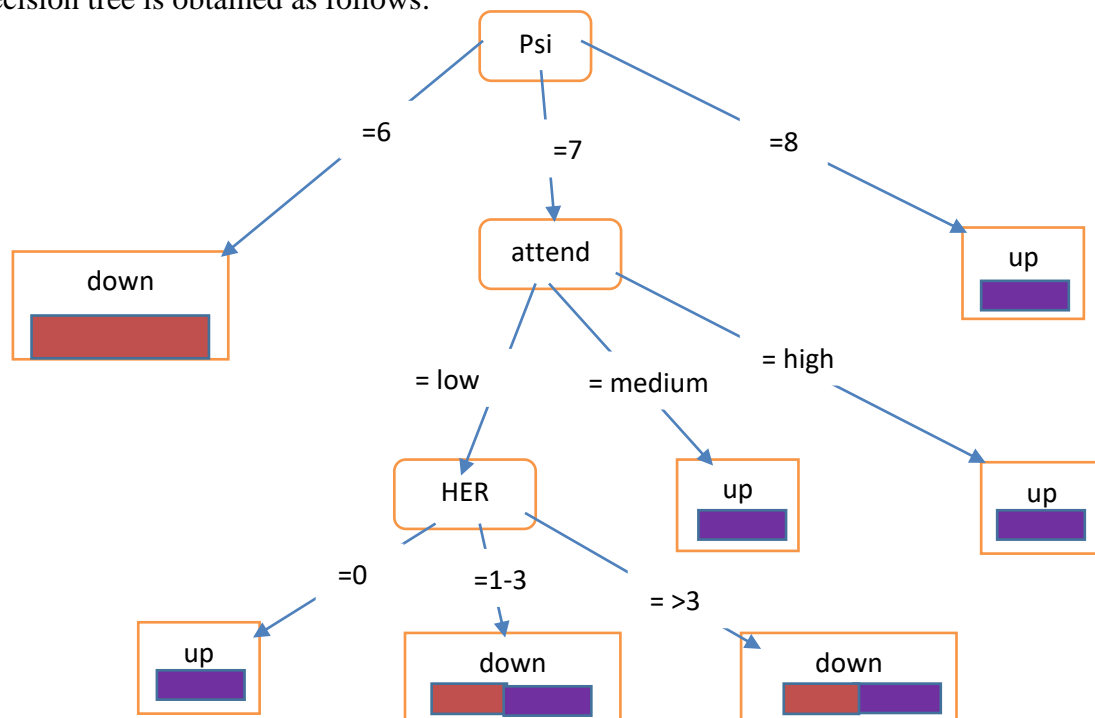
Based on the table above, it is known to occupy the top root in the decision tree attribute used which has the smallest entropy namely the Psychomotor attribute with a value of 0.440851994. For the description of the next branch under the psychomotor root recalculation is carried out to determine the entropy. The results of the data calculation are as follows:

**Table 3.** Data Calculation

Node		Gain	Information	
Psychomotor				
1.1.1	Frequency	6	0	Not Increase
6				
	Psychomotor		0	
	7			
	Affective	Good	0	
	Low			
	Attendance		0	
	Medium			
	1-3			
	HER		0	
>3				
1.1.2	Frequency	7		
6				
	Psychomotor	7	0.30802	
	8			
	Affective	VeryGood	0.06527	
		Good		
		low		
	attendance	Moderate	0.27772	
		high		
	0			
	Total HER	1-3	0.39341	
>3				
1.1.3	Frequency	8		Increase
7				
			0.12809	

		8		
	Very Good			
	Affective		0.07601	
	Good			
	low			
	attendance	Moderate	0	
	high			
	0			
	Total HER	1 -3	0	
	>3			

From the above calculation, then it is entered into the modeling decision tree and the decision tree is obtained as follows:



From the picture above, the test conditions for the class increase classification are obtained, namely:

IF  $\text{psi} \leq 74$  or  $\text{kog} \leq 74$  then ket = not going to class"

IF  $\text{psi} > 74$  and  $\text{psi} < 86.5$  If  $\text{hdr} = \text{low}$

If  $\text{her} = 0$ " then ket =

If  $\text{her} = 1-3$ " then ket = 50% chance of going up grade" If  $\text{her} = >3$ " then ket = probability of going to class 25%"

IF  $\text{psi} > 74$  and  $\text{psi} < 86.5$

If  $\text{hdr} = \text{medium}$ " then ket="upgrading IF  $\text{psi} > 74$  and  $\text{psi} < 86.5$

If  $\text{hdr} = \text{height}$ " th en ket="go up to grade" IF  $\text{psi} \geq 86.5$  then ket =" go to grade

To support the decision-making system for vocational school teachers, a web-based decision-making system application for grade promotion was built. Testing the decision tree in the next picture will be used as decision logic in designing decision-making information systems.

## IV. Results and Discussion

### 4.1 Results of the C45 Algorithm Calculation Process

The C45 algorithm procedure in performing the mining calculation process to build a decision tree. The following are the steps for the calculation procedure:

1. Data is grouped based on the attributes and the values in them.
2. Counting the amount of data on each existing attribute value
3. Classifying the calculated data into two groups based on the target goal, namely graduating on time and not on time
4. Calculating the total entropy of 55 student data
5. Calculating the entropy of each value attribute
6. Calculating the Gain of each attribute
7. Finding the attribute with the highest Gain to be used as Root
8. Determining the attribute value to be used as a branch
9. Determining the next node from the selected attribute based on the highest Gain.

### 4.2 Results of Tables and Figures of Decision Tree Calculation Results with the C45 Algorithm

#### 1) Table I Determination of Root

Data is grouped based on attributes and attribute values and then the total number is calculated, the number of students who graduate on time and incorrectly, then the Entropy and Gain values for each attribute are calculated.

The TOTAL row of the ENTROPY column in the table above is calculated by the following formula:

$$Entropy (Total) = \left(-\frac{20}{55} * \left(\frac{20}{55}\right)\right) + \left(-\frac{35}{55} * \left(\frac{35}{55}\right)\right)$$

$$Entropy (Total) = 0.945660305$$

Entropy calculation for each attribute value is calculated in the same way as total Entropy:

$$= (Gender Entropy, L) \left(-\frac{15}{32} * \left(\frac{15}{32}\right)\right) + \left(-\frac{17}{32} * \left(\frac{17}{32}\right)\right)$$

$$Entropy (Total) = 0.997180399$$

Meanwhile, the Gain value in the Gender row is calculated using the Gain formula as follows:

$$Gain (Total, Gender) =$$

$$Entropy (Total) - \sum_{i=1}^n \frac{Sex}{Total} * Entropy * (Gender)$$

$$Gain (Total, Gender) = 0.945 - \left(\frac{22}{55} * 0.997\right) + \left(\frac{23}{55} * 0.775\right)$$

$$Gain (Total, Gender) = 0.049598355$$

**Table 4. Root Determination**

Node	Attribute	Value	Number of Cases Total	Number of Cases Not on Time	Number of Cases on Time	Entropy	Gain
1	Total	Total	55	20	35	0.94560305	
	Gender						0.049598355
		L	32	15	17	0.997180399	
		P	23	5	18	0.755375413	
	Parents' Salary						0.005866231
		Low	12	5	7	0.979868757	
		Medium	28	9	19	0.905928216	
		High	15	6	9	0.970950594	
	Year of Entry						0.329081463
		2011	26	16	10	0.961236605	
		2012	9	4	5	0.99107606	
		2013	5	0	5	0	
		2014	6	0	6	0	
		2015	9	0	9	0	
	GPA						0.343883318
		Praise	15	1	14	0.353359335	
		Very Satisfactory	24	7	17	0.870864469	
		Satisfactory	7	3	4	0,985228136	
		Enough	9	9	0	0	
	Social Studies 1						0.2171293
		Praise	16	4	12	0.811278124	
		Very Satisfactory	9	0	9	0	
		Satisfactory	8	2	6	0.811278124	

		Enough	18	11	7	0.964078 765	
		Less	4	3	1	0.811278 124	
	Social Studies 2						0.261331924
		Praise	12	1	11	0.413816 85	
		Very Satisfact ory	16	4	12	0.811278 124	
		Satisfact ory	8	1	7	0.543564 443	
		Enough	18	13	5	0.852405 179	
		Less	1	1	0	0	
	Social Studies 3						0.137892731
		Praise	17	3	14	0.672294 817	
		Very Satisfact ory	13	4	9	0 1	
		—	—	—	—	—	
		—	—	—	—	—	
		—	—	—	—	—	
	—						—
		—	—	—	—	—	
		—	—	—	—	—	
		—	—	—	—	—	
		—	—	—	—	0.503258 335	
		Less	1	1	0	0	

The result data in the table above can be seen that the attribute with the highest gain is IPS 4, which is 0.370768444. Thus IPS 4 becomes the root node. There are 5 attribute values from IPS 4, namely Praise, Very Satisfactory, Satisfactory, Enough, and Less. The shape of the tree that is formed is shown in Figure 16.

When the analysis of the program that has been made is complete, the next step is the implementation stage. This stage is where the steps in the program that have been made, this stage is carried out to find out whether the system in the program is running well.

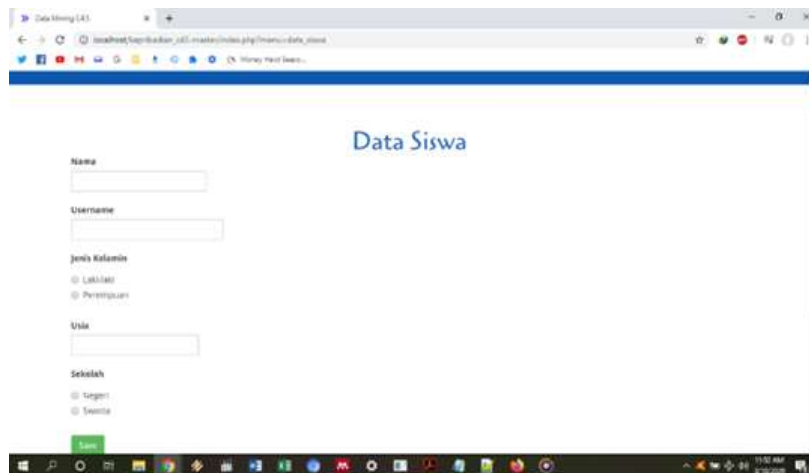
In this stage it is also to find out whether the user is satisfied with the program that has been made. In this stage will be given the steps in running the program from the beginning until the results of the program appear.





**Figure 1. Login Page**

Figure 1 is the login page, where we will enter our username and password. On this login page, students enter their username and password in the form of the school's main number. And for the teacher, enter the username and password according to the one made by the teacher.



**Figure 2. Student Data**

Figure 2 is a login page where we have to enter the username and password of the student which we will process later with the C4.5 algorithm. in Figure 2 students fill in data to log into the website where the data will be processed so that students can find out the results of learning.

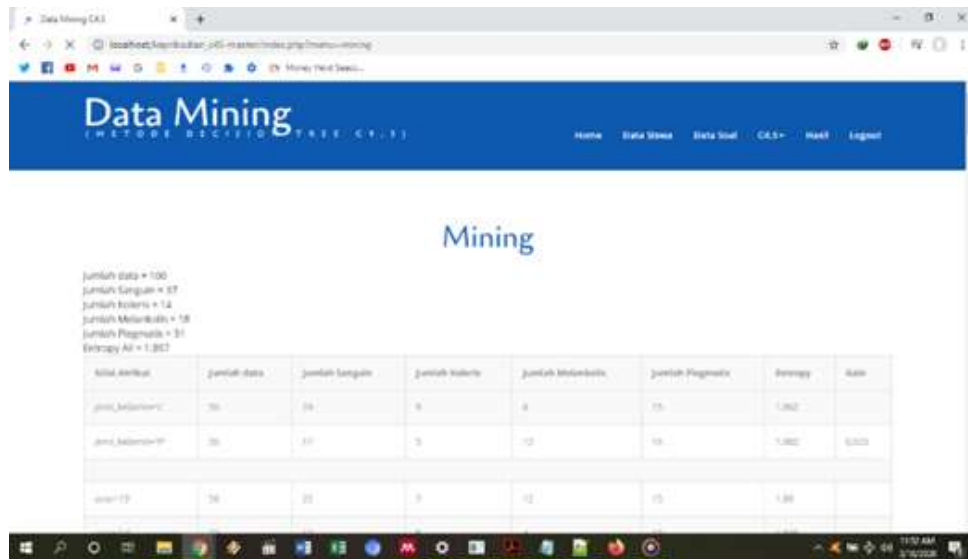


Figure 3 is a data checking process page with the C4.5 algorithm. After the calculation is done, the results will come out. In Figure 3 is the process of mining algorithm C4.5. This process uses data from students who have logged into the website. Where in this process data such as study scores, attendance and student activity are processed which will later be processed by the algorithm and will bring up the results of the data. Whether the student goes to class or stays class.

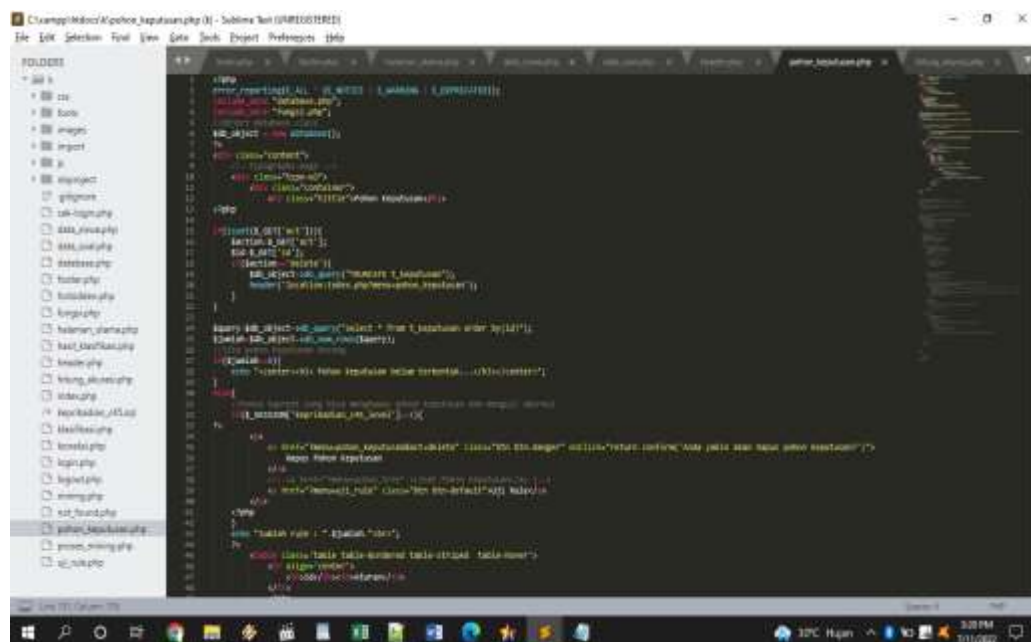


Figure 4 is a mining process and then the process is carried out using a Decision Tree or often called a decision tree. The process in Figure 4 is where after the student data is mined it will then be reprocessed using a decision tree in the decision there are several attributes in each branch that contains the student data being tested. The picture above is the source code for calculating data entropy using the decision tree method.

## V. Conclusion

Based on the results of the study, it was concluded that using the C4.5 Algorithm could help teachers in making decisions to predict the increase in grades for vocational students to be faster. The level of accuracy of the study using the C4.5 Algorithm and the selection of Information Gain features obtained an accuracy value of 83.33%. For the results of the decision tree, it can be seen that the attribute of psychomotor value is the highest determining factor for student grade promotion. In improving the learning system, special attention is needed for psychomotor subjects so that the level of student grades will increase. With the help of a web-based application, it is hoped that it will facilitate the performance of teachers in predicting the increase in the class of SMK students in large numbers.

## References

- E. P. Rohmawan, "Menggunakan Metode Decision Tree," J. Ilm. Matrik Vol.20 No.1, April 2018 21-30, pp. 21–30, 2013.
- E. P. W. Mandala and Rahmatino, "Aplikasi Data Mining Untuk Memprediksi Kelulusan Siswa Jurusan IPA MAN 3 Padang," Maj. Ilm. UPI YPTK, Vol. 20, No. 2, Oktober 2013, vol. 20, no. 2, pp. 22–28, 2013.
- Gunawan, G.G., and Sulaeman, M. (2020). Determining Factors in the Use of Digital Marketing and Its Effect on Marketing Performance in the Creative Industries in Tasikmalaya. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3 (3): 2543-2550.
- Hariati, P. (2021). Implementation of Digital Literacy toward Pandemic Situation. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 4 (2): 2920-2926.
- Hartanto, D. (2020). Sociology Review of Social Phenomenon, Social Rules and Social Technology. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3, (2): 1175-1184.
- H. Kurniawan and J. S. Informasi, "Aplikasi Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori Di Ibi Darmajaya Bandar Lampung," J. TIM Darmajaya Vol. 02 No. 01 Mei 2016 ISSN 2442-5567 | E-ISSN 2443-289X Page, vol. 2, no. 01, pp. 79–93, 2016.
- Marlizar, et.al. (2021). Effect of Service Quality and Use of E-Service Technology on Customer Loyalty: A Case Study of Maxim in Aceh. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 4, (4): 8002-8016.
- Pramusinto, N.D., Daerobi, A., and Hartanto, D. (2020). Labor Absorption of the Manufacturing Industry Sector in Indonesia. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3 (1): 549-561.
- R. D. Pambudi, A. A. Supianto, and N. Y. Setiawan, "Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya," J. Pengemb. Teknol. Inf. dan Ilmu Komputer. 2196, vol. 3, no. 3, pp. 2194–2200, 2019.
- S. Sulistio, "Aplikasi Prediksi Kelulusan Mahasiswa Dengan Metode Linear Discriminant Analysis," Comput. J. Comput. Sci. Inf. Syst., vol. 1, no. 1, p. 58, 2017, doi: 10.24912/computatio.v1i1.241.