Journal of Artificial Intelligence and Engineering Applications
Websites:https://ioinformatic.org/

15th October 2021. Vol. 1. No. 1

Application of the C4.5 Algorithm in Teaching Teachers' Skills on Learning Effectiveness

Feby Widya Sari[1]*, M. Safii[2]*

[1] STIKOM Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia*
[2] AMIK Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia*
[3]Jln. Sudirman Block A No. 1-3 Pematangsiantar, North Sumatra
** febywidya0200@gmail.com, *m.safii@amiktunasbangsa.ac.id*

Abstract

Teachers as educators and education personnel have a very important role in improving the quality of education in schools. In an effective teaching and learning process, teacher skills in teaching are very important. This study aims to classify the skills of teachers in SMA Yayasan Pendidikan Keluarga using the Decision Tree method with the application of the C4.5 Algorithm in order to improve the teaching system in an effort to increase students' understanding of the learning process. In determining the teaching skills of teachers, classification is carried out into the labels "Relevant" and "Not Relevant" which has 5 variables, namely Age, Length of Work, Number of Teaching Hours, Students, and Learning Media. Sources of data used in this study obtained by conducting observations and interviews.

Keywords: *data mining, C.45 algorithm, teacher skills, learning effectiveness.*

1. Introduction

Education is one of the factors of progress and independence of a nation and state. The more advanced the education of a country, the bigger the country will be[1]. A teacher is a person who imparts knowledge to students[2]. As educators and education personnel, teachers have a very important role in improving the quality of education in schools. In order to improve the development of students, it is very necessary for teachers to improve a comparative understanding of their students[3]. Teachers' skills in teaching must have the competencies needed to carry out the duties and functions of teachers by teaching, educating, guiding, training, and evaluating their students. Teachers are required not only to be able to teach, but also to be able to adapt to technological developments and developments in the world of work[4].

The level of understanding of students in the learning process is one of the main things besides being supported by a high willingness to learn. The inability of students to understand the lessons given by teachers at school is because some teachers in teaching prefer to apply methods and teaching methods that tend to be monotonous, causing students to become bored and less motivated, so that teachers receive less attention from students in class which leads to the impact of learning outcomes getting worse. low[5]. The existence of several factors that support students are able to understand in the teaching and learning process takes place. The main factor comes from the way the teacher communicates to the students. Another factor is the number of students, the more students in the learning process, the less students focus on understanding the material provided and the number of hours of teaching and teaching experience from teachers who support the delivery of good and effective material to students.

The teacher's task is not only limited to teaching, explaining teaching materials, but also training, guiding and facilitating their students[6]. To determine the teaching skills of teachers, classification is carried out into the labels "Relevant" and "Not Relevant" which has 5 variables, namely Age, Length of Work, Number of Teaching Hours, Students, and Learning Media. To overcome this problem, a research was conducted using the C4.5 Algorithm. The C4.5 algorithm is a development of the ID3 algorithm which has advantages in overcoming missing values, continuing data and pruning[7]. A decision tree is a flowchart like tree structure, where each internal node represents a test on an attribute, each branch shows the results of the test, and the leaf node shows the classes or class distribution.[8].Decision tree or decision tree is the process of changing the shape of the data (table) into a tree model, converting the tree model into rules, and simplifying the rules. The decision tree is the result of the entropy and information gain calculation process[9].

Ade Yuliana conveyed the results of his research on student satisfaction using the C4.5 algorithm to predict Lecturer Performance at Tedc Polytechnic Bandung. From this research, it can be seen the factors that influence student satisfaction so that lecturers can improve their performance[10]. Oscario conveyed the results of his research on the prediction of Learning Style Suitability for Elementary School Students using the C4.5 algorithm. From this research, it can be seen things that affect the suitability of students' learning styles so that the teaching and learning process becomes effective[11].

2. Results and Discussion

In this chapter the author describes the results of research and data analysis. By implementing the final results of the C4.5 Algorithm using 2 stages, namely by calculating manually and testing using the RapidMiner 5.3 application.

2.1. Data Processing Using C4.5 Algorithm

To get the results of the research conducted, the following is a description of manual calculations using the C4.5 Algorithm in determining the skills of teaching teachers on learning effectiveness. There are 5 criteria used, namely: Age, Length of Work, Number of Teaching Hours, Students, Learning Media. The following sub-criteria of each criterion are as follows:

Table 1. Research data

Name	Age	Length of work	Number of Teaching Hours	Student	Learning Media	Information
A1	Old	5 years	24>35 hours	Minimum	Print	Relevant
A2	Old	15 years	24>35 hours	Minimum	Print	Relevant
A3	Old	15 years	<24 hours	Maximum	Non Print	Relevant
A4	Old	15 years	24>35 hours	Maximum	Print	Relevant
A5	Young	5 years	<24 hours	Minimum	Print	Relevant
A6	Young	5 years	<24 hours	Minimum	Print	Relevant
A7	Young	5 years	>35 hours	Maximum	Print	Relevant
A8	Old	15 years	24>35 hours	Minimum	Print	Relevant
A9	Young	5 years	<24 hours	Minimum	Non Print	Relevant
A10	Young	5 years	24>35 hours	Minimum	Non Print	Irrelevant
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
A30	Old	10 years	24>35 hours	Minimum	Non Print	Relevant
A31	Young	10 years	24>35 hours	Minimum	Print	Relevant
A32	Old	15 years	24>35 hours	Minimum	Print	Relevant
A33	Old	5 years	24>35 hours	Minimum	Print	Relevant
A34	Young	5 years	>35 hours	Maximum	Non Print	Irrelevant

Table 2. Data Explanation

No.	Attribute	Explanation
1.	Age	Old (over 36 years old), Young (under 35 years old)
2.	Length of work	5 Years, 10 Years, 15 Years
3.	Number of Teaching Hours	<24 hours (less than 24 teaching hours), 24>35 hours (24-35 teaching hours), >35 hours (more than 35 teaching hours)
4.	Student	Minimum (1-250 students), Maximum (more than 250 students)
5.	Learning Media	Print and Non Print

After the data is determined, the calculation of the C4.5 algorithm begins by selecting the root attribute first by looking for the total number of cases, the number of relevant and irrelevant cases. After that, the gain is calculated for each attribute.

Calculating the total entropy:

Entropy [Total] $= \left(-\frac{27}{34}\ x\ log_2\left(\frac{27}{34}\right)\right) + \left(-\frac{7}{34} x log_2\left(\frac{7}{34}\right)\right)$

Entropy [Total] $= 0.733538$

Calculating the entropy and gain Age:

Entropy [Age – Old] $= \left(-\frac{17}{20}\ x\ log_2\left(\frac{17}{20}\right)\right) + \left(-\frac{3}{20} x log_2\left(\frac{3}{20}\right)\right)$

$= 0.60984$

Entropy [Young age] $= \left(-\frac{10}{14}\ x\ log_2\left(\frac{10}{14}\right)\right) + \left(-\frac{4}{14} x log_2\left(\frac{4}{14}\right)\right)$

$= 0.863121$

Gain [Total, Age] $= 0.733538 - \left(\left(\frac{20}{34} x\ 0.60984\right) + \left(\frac{14}{34} x\ 0.863121\right)\right)$

$= 0.019405751$

Calculating the entropy and gain Working Time:

Entropy [Length of Work – 5 Years] $= \left(-\frac{11}{18}\ x\ log_2\left(\frac{11}{18}\right)\right) + \left(-\frac{7}{18} x log_2\left(\frac{7}{18}\right)\right)$

$= 0.964079$

Entropy [Length of Work – 10 Years] = 0

Entropy [Length of Work – 15 Years] = 0

Gain [Total, Length of work] $= 0.733538 - \left(\left(\frac{18}{34} x\ 0{,}964079\right) + \left(\frac{7}{34} x\ 0\right) + \left(\frac{9}{34} x\ 0\right)\right)$
$= 0.223143289$

Calculating entropy and gain Number of Teaching Hours (JJM):

Entropy [JJM - <24 hours] $= \left(-\frac{7}{10}\ x\ log_2\left(\frac{7}{10}\right)\right) + \left(-\frac{3}{10} x log_2\left(\frac{3}{10}\right)\right)$
$= 0$

Entropy [JJM - 24 hours-35 hours] $= \left(-\frac{17}{20}\ x\ log_2\left(\frac{17}{20}\right)\right) + \left(-\frac{3}{20} x log_2\left(\frac{3}{20}\right)\right)$
$= 0.60984$

Entropy [JJM - >35 hours] $= \left(-\frac{3}{4}\ x\ log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} x log_2\left(\frac{1}{4}\right)\right)$
$= 0$

Gain [Total, Number of Teaching Hours] $= 0.733538 - \left(\left(\frac{10}{34} x\ 0\right) + \left(\frac{20}{34} x\ 0.60984\right) + \left(\frac{4}{34} x\ 0\right)\right)$
$= 0.374808338$

Calculating student entropy and gain:

Entropy [minimum] $= \left(-\frac{22}{28}\ x\ log_2\left(\frac{22}{28}\right)\right) + \left(-\frac{6}{28} x log_2\left(\frac{6}{28}\right)\right)$
$= 0.749595$

Entropy [max] $= \left(-\frac{5}{6}\ x\ log_2\left(\frac{5}{6}\right)\right) + \left(-\frac{1}{6} x log_2\left(\frac{1}{6}\right)\right)$
$= 0.650022$

Gain [Total, Student] $= 0.733538 - \left(\left(\frac{28}{34} x\ 0.749595\right) + \left(\frac{6}{34} x\ 0.650022\right)\right)$
$= 0.001514349$

Calculating entropy and gain Learning Media :

Entropy [print] $= \left(-\frac{23}{23}\ x\ log_2\left(\frac{23}{23}\right)\right) + \left(-\frac{0}{23} x log_2\left(\frac{0}{23}\right)\right)$
$= 0$

Entropy [non print] $= \left(-\frac{4}{11}\ x\ log_2\left(\frac{4}{11}\right)\right) + \left(-\frac{7}{11} x log_2\left(\frac{7}{11}\right)\right)$
$= 0.94566$

Gain [Total, Student] $= 0.733538 - \left(\left(\frac{23}{34} x\ 0\right) + \left(\frac{11}{34} x\ 0.94566\right)\right)$
$= 0.427589007$

Table 3. Node Calculation Results 1.1

Node 1		CASE NUMBER	Relevant	Irrelevant	ENTROPY	GAIN INFORMATION
TOTAL		34	27	7	0.733538	
Age						0.019405751
	Old	20	17	3	0.60984	
	Young	14	10	4	0.863121	
Length of work						0.223143289
	5 years	18	11	7	0.964079	
	10 years	7	7	0	0	
	15 years	9	9	0	0	
Number of Teaching Hours						0.374808338
	>24 hours	10	7	3	0	
	24>35 hours	20	17	3	0.60984	
	>35 hours	4	3	1	0	
Student						0.001514349
	minimum	28	22	6	0.749595	
	Max	6	5	1	0.650022	
Learning Media						**0.427589007**
	print	23	23	0	0	
	non print	11	4	7	0.94566	

From the calculation results in Table 3. The highest attribute value obtained is Learning Media with a gain of 0.427589007. Then the Learning Media attribute is selected as the root node. The value of the print attribute class is empty, so there is no need to perform calculations on the attribute class. For the non-print attribute class, the results between the relevant and irrelevant decisions have not been obtained, so further calculations need to be carried out.

2.2. Testing Process With RapidMiner

In the final stage of implementing the C4.5 Algorithm, adjustments are made to the results of manual calculations through testing using RapidMiner 5.3 software. Testing of the results of manual calculations using RapidMiner software is carried out through several stages of the process as follows:

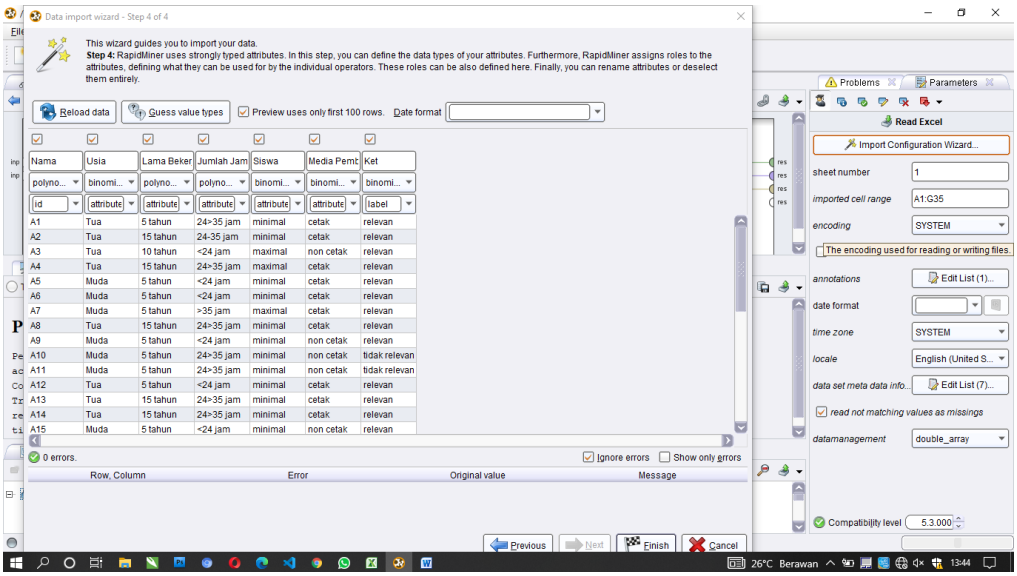

Figure 1. Import Data Display

Importing data, the data that has been successfully imported is then processed into a decision tree with a data processing model with the initial process of modeling data processing by adding a Split Validation operator as shown in Figure 2.

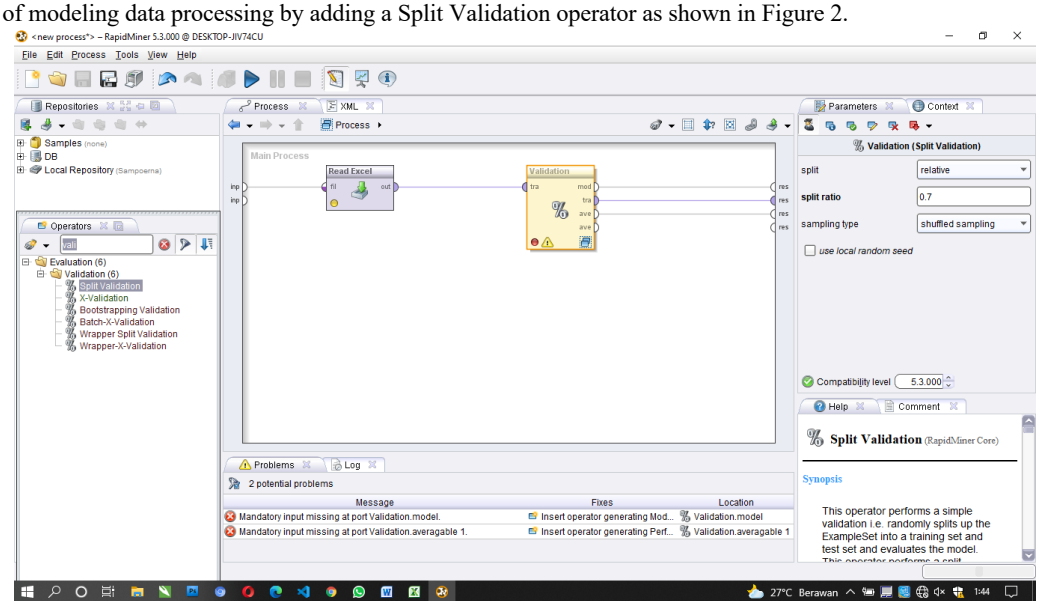

Figure 2. Connectivity between Data and Decision Tree Model (1)

In the second stage of the process, namely making the decision tree model rules in the split validation operator by double clicking on the split validation operator, then the operator validation page will appear which is divided into training and testing areas as shown in Figure 3. following :

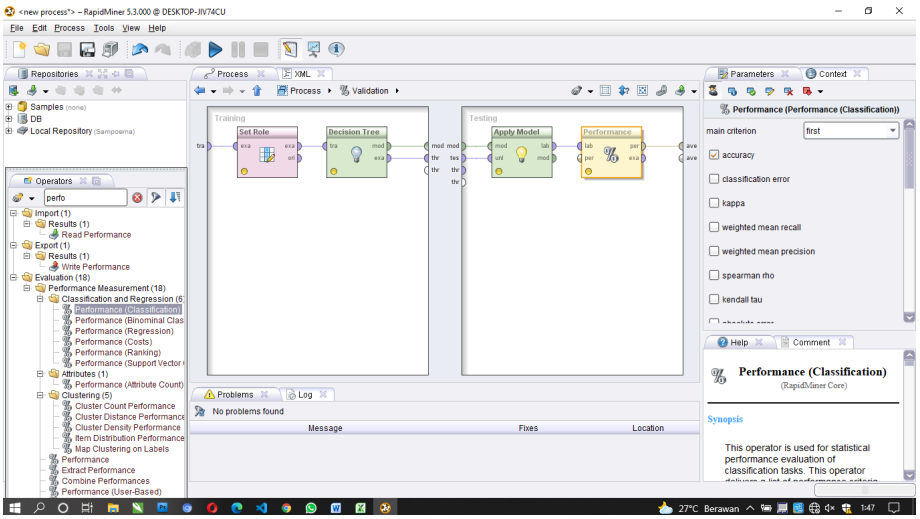

Figure 3. Connectivity between Data and Decision Tree Models (2)

Next, drag and drop the Set Role operator and decision tree in the training area, the Apply Model operator and the performance (classification) operator in the testing area, then connect each operator. The results of data processing with a decision tree model in accordance with the RapidMiner software, can be seen in Figure 4. as follows :

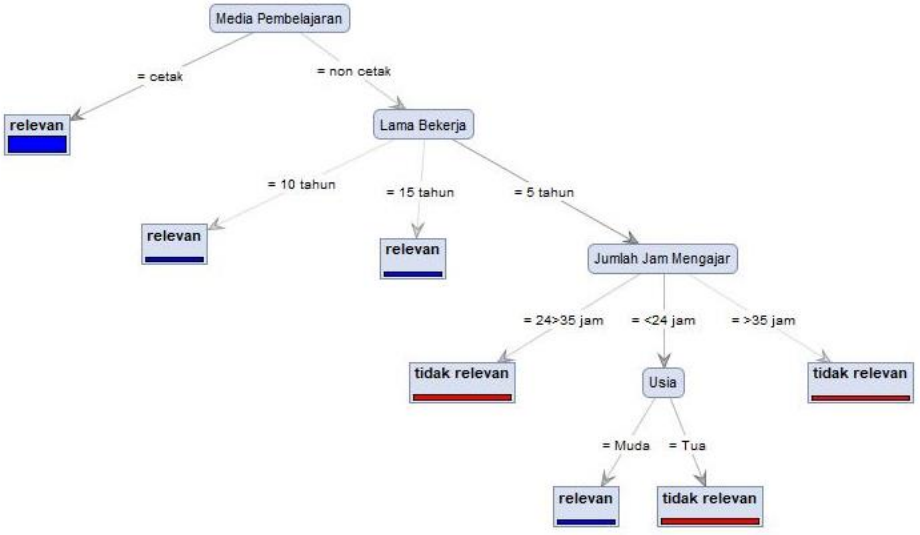

Figure 4. Decision Tree on Rapidminer

Figure 4. above is a decision tree generated in Rapidminer with rules or rules that can be seen in the text view in Figure 5. following :

Tree

```
Media Pembelajaran = cetak: relevan {relevan=23, tidak relevan=0}
Media Pembelajaran = non cetak
|   Lama Bekerja = 10 tahun: relevan {relevan=1, tidak relevan=0}
|   Lama Bekerja = 15 tahun: relevan {relevan=1, tidak relevan=0}
|   Lama Bekerja = 5 tahun
|   |   Jumlah Jam Mengajar = 24>35 jam: tidak relevan {relevan=0, tidak relevan=3}
|   |   Jumlah Jam Mengajar = <24 jam
|   |   |   Usia = Muda: relevan {relevan=2, tidak relevan=0}
|   |   |   Usia = Tua: tidak relevan {relevan=0, tidak relevan=3}
|   |   Jumlah Jam Mengajar = >35 jam: tidak relevan {relevan=0, tidak relevan=1}
```

Figure 5. Rule Decision Tree on Rapidminer

The picture above shows the results of a complete description of the decision tree that has been formed using the C4.5 algorithm. From the description results also show that the use of data mining algorithm C4.5 is good to use in the process of extracting data (data mining process) to draw some conclusions which are visualized with a decision tree.

2.3. Data Validation

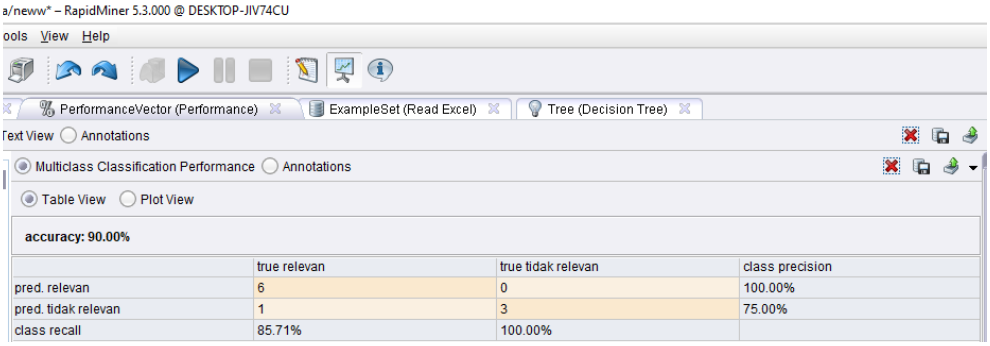

Figure 6. Algorithm Accuracy Value C4.5

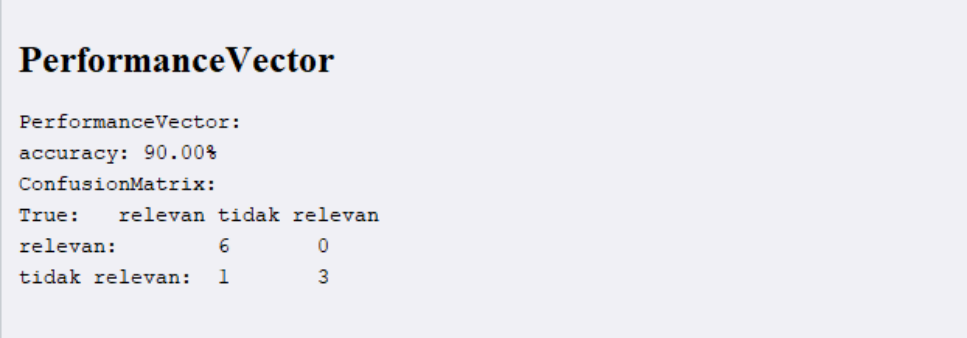

Figure 7. Value of Performance Vector Algorithm C4.5

Based on data processing using RapidMiner software, the system accuracy value is 90.00%. Where the model that has been formed is tested for accuracy by entering or testing derived from training data using split validation on the Rapidminer 5.3 application to test the level of accuracy.

3. Conclusion

Based on the results of research in determining teaching teacher skills on the effectiveness of learning using the C4.5 Algorithm Method, it can be concluded that the problem in determining teacher skills can be solved using data mining, namely the C4.5 Algorithm which produces 7 rules and the level of accuracy generated by the method is 90% and from calculations using the C4.5 Algorithm, the most dominant factor is Learning Media with a gain value of **0.427589007**

Reference

[1] MA Sembiring, "Application of the C45 Algorithm Decision Tree Method to Predict Student Learning Outcomes Based on Academic History," *J. Teknol. and Sis. inf.*, vol. 3, no. 1, pp. 60–65, 2016.

[2] N. Illahi, "The Role of Professional Teachers in Improving Student Achievement and Quality of Education in the Millennial Era," *J. Asy-Syukriyyah*, vol. 21, no. 1, pp. 1–20, 2020, doi:10.36769/asy.v21i1.94.

[3] D. Naibaho, "The role of the teacher as a facilitator in the development of students," *J. Christ. Hum.*, vol. 2, no. 1, pp. 77–86, 2018.

[4] MY Almursyid, "The Perception of Vocational Teachers of SMK Negeri 1 Bukittinggi on the Application of 21st Century Teacher Competencies," *CIVED (Journal Civ. Eng. ...*, pp. 2128–2134, 2018, [Online]. Available: http://ejournal.unp.ac.id/index.php/cived/article/view/9933.

[5] M. Monawati. and F. Fauzi., "The Relationship between Teacher's Teaching Creativity and Student Achievement,"*J. Basic Charm*, vol. 6, no. 2, pp. 33–43, 2018, doi:10.24815/pear.v6i2.12195.

[6] A. Prasetyo and R. Susanti, "Web-Based Sales Information System at PT. The Prosperous Light of Sentosa Blitar,"*J. Ilm. Technol. inf. Asia*, vol. 10, no. 2, pp. 1–16, 2016.

[7] SM Monalisa and F. Hadi, "C4.5 Algorithm in Determining New Student Majors," *Ultim. J. Tech. information.*, vol. 12, no. 2, pp. 108–113, 2020, doi:10.31937/ti.v12i2.1838.

[8] F. Riandari and A. Simangunsong, "The application of the C4.5 algorithm to measure the level of student satisfaction," *J. Mantik Penusa*, vol. 3, no. 2, pp. 1–7, 2019.

[9] P. Graduation, M. Using, C. Algorithm, S. Case, and DI University, "Predictions of student graduation using the c4.5 algorithm (a case study at the university of civilization)," vol. 1, no. 2, pp. 70–77, 2020.

[10] A. Yuliana and DB Pratomo, "Predicting Student Satisfaction on Lecturer Performance at TEDC Polytechnic Bandung," *2017*, pp. 377–384, 2017.

[11] O. OSCARIO, J. JASMIR, and Y. NOVIANTO, "APPLICATION OF THE C4.5 ALGORITHM TO PREDICT LEARNING STYLE FITNESS FOR ELEMENTARY SCHOOL STUDENTS (CASE STUDY: SD SARIPUTRA C4.5 Algorithm To Predict Learning Style Suitability for Elementary School Students (Case Study: SD Sariputra Jambi)," *J. Process.*, vol. 14, no. 2, p. 141, 2019, doi:10.33998/processor.2019.14.2.637.