

## **DECISION TREE C4.5 ALGORITHM FOR TUITION AID GRANT PROGRAM CLASSIFICATION (CASE STUDY: DEPARTMENT OF INFORMATION SYSTEM, UNIVERSITAS TEKNOKRAT INDONESIA)**

**Ahmad Ari Aldino<sup>1</sup>, Heni Sulistiani<sup>2</sup>**

<sup>1,2</sup> Universitas Teknokrat Indonesia  
Lampung, Indonesia  
[aldino@teknokrat.ac.id](mailto:aldino@teknokrat.ac.id)

### **Abstract**

In pandemic era, almost everyone struggles for their life. College students are such example. They have difficulty in paying tuition fee to continue their study. Based on this problematic situation, Universitas Teknokrat Indonesia grants the students who have good academic performance with tuition fee aid program. Many features used for determining the grant made it hard to make a decision in a short time or even takes very long time. To make it easier for management to decide who is the right student to get grant, it needs classification model. The purpose of this study is to classify the grant recipients by using decision tree C4.5 algorithm. The C4.5 algorithm is used because it can directly show the pattern of data in the decision tree form that can determine whether a potential student can be accepted as an awardee or not. The result of the study is classification model with 87 % accuracy, precision and recall for all part. It means the model perform quite well to be implemented into system.

***Keyword: Data Mining, Classification, Decision Tree, C4.5 Algorithm***

## **INTRODUCTION**

The high cost of tuition for some students is an obstacle in continuing to study in college especially during Covid-19 even for on-going students. This can result in students frequently postponing their study or even stopping midway. Granting scholarships has several criteria that are used as consideration in making a decision on whether to award scholarships or not. Universitas Teknokrat Indonesia provides tuition aid grant to its students who have good academic performance, so that the students can continue their studies and relieve the burden of parents. The number of criteria selected in determining scholarship award decisions results in management having difficulty in making a decision and the time it takes to be longer (Chermiti, 2019).

Classification is a process for finding a model or function that describes and distinguishes a data class or concept that has the purpose of using that classification result model to make predictions of object classes where the label class is unknown (Nowak, 2017). One of the popular classification techniques used is Decision Tree (Alsagheer, et al, 2017). The number of criteria used for the determination of tuition aid grant makes it difficult to make decisions on awarding scholarships. In addition, it takes a long time. The purpose of this research is to make classification of awardee using Decision Tree using C4.5 Algorithm.

Classification results are evaluated and validated with Confusion Matrix and Ten-fold Cross Validation to determine the accuracy, precision, and recall of Decision Tree in making scholarship classifications (Nowak, 2017).

## **LITERATURE REVIEW**

### **2.1 Decision Tree**

Decision tree is a classification method that uses a tree structure, where each node represents the attribute and its branch represents the value of the attribute, while the leaf is used to represent the class. The top node of this decision tree is called root. Rismayanti (2018) states that this method is a very popular method to use because the results of the model formed are easy to understand. It is named after the decision tree because the rules formed are similar to the shape of the tree. Trees are formed from binary recursive sorting processes in data groups, so the value of response variables in each data group makes sorting results more homogeneous (Sharma and Kumar, 2016). The concept of the decision tree is to turn the data into a decision tree and the rules of the decision. The main benefit of using the decision tree is the ability to simplify complex decision-making processes so that decision makers can interpret solutions to problems (Mesarić and Šebalj, 2016).

### **2.2 C4.5 Algorithm**

The C4.5 algorithm is one of the algorithms used to classify or group

datasets. The basis of the C4.5 algorithm is decision tree formation (Petropoulos et al, 2018). The branches of the decision tree are a question of classification while for the leaves are classes or groups. Because the purpose of C4.5 algorithm is to perform classification, so the result of processing datasets is in the form of grouping data into certain classes (Jabeur et al, 2020). The C4.5 algorithm is a development of the ID3 algorithm, where the flaws that the ID3 algorithm has can be covered by the C4.5 algorithm (Bekesiene and Hoskova-Mayerova, 2018).

Some things that distinguish the C4.5 algorithm from ID3 are:

1. Robust to noise data
2. Able to handle variables with both discrete and continuous types
3. Able to handle variables that have missing value
4. Can trim branches from the decision tree

In general, the C4.5 algorithm for building decision tree model are as follow:

1. Select a variable as the root
2. Create branches for each value
3. Divide cases into branches
4. Repeat the process for each branch until all cases on the branch have the same class.

The first thing to do to form a decision tree is to determine which attributes/variables are at the root of the decision tree. The way to

determine which variable is rooted is to use entropy, gain, split info, and gain ratio (Mittal et al, 2017).

Entropy is a parameter for measuring the level of diversity (heterogeneity) of the data set. If the value of the entropy gets greater, then the level of diversity of a data set is greater (Al-Barrak and Al-Razgan, 2016). The formula for calculating entropy is as follows:

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

Where:

S : case set

n : number of classification classes

$p_i$  : number of sample proportions for class i

Gain is a measure of the effectiveness of a variable in classifying data. The gain of a variable is the difference between the total entropy value and the entropy of that variable (Hamoud et al, 2018). Gain can be formulated by:

$$\text{Gain (S,A)} = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy (S}_i) \dots \dots \dots (2)$$

Where:

A : Variable

$|S_i|$  : Number of samples for i

$|S|$  : Number of samples for the entire data

In the C4.5 algorithm, gain values are used to determine which variables are nodes of a decision tree. A variable that has the highest gain

will be a node in the decision tree (Patel and Prajapati, 2018).

## RESEARCH METHODS

The Research Method used in the application of the C4.5 algorithm for Tuition Assistance Program Classification, using the research design addressed in Figure 1 follows:



**Figure 1.** Research flow diagram

### 3.1 Data Collection

Data collection is collecting data that will be used in the C4.5 classification algorithm process.

### 3.2 Data Preprocessing

Data Preprocessing is the process of transforming, merging, or converting data into the appropriate form, in order to be processed with the calculation of the C4.5 algorithm.

### 3.3 Feature Selection

Feature Selection is selecting the data to be used in the C4.5 classification algorithm process. The purpose of data selection is to create a target data set, select a data set, or focus on a subset of variables or sample data, where discovery will be made.

### 3.4 Model Calculation

Calculation of all attributes/variables, entropy uses formula (1) and information gain using formula (2) to find out the

highest information gain to be used as a root node in decision tree creation.

### 3.5 Decision Tree

Tree decisions are the result of the process of calculating entropy and information gain, after repeated calculations until all tree attributes have a class and can no longer be done the calculation process.

### 3.6 Rule Model

Rules model is an explanatory description that represents a decision tree.

### 3.7 Validation and Testing

Validation and testing are the test done to know whether all functions work well or not. Validation is done using confusion matrix and Ten-fold Cross Validation done by dividing a data set into ten segments that are equally large by randomize data. Validation and testing are performed to determine the accuracy, precision, and recall of classification prediction results. Accuracy is the percentage of records that are properly classified in dataset testing. Precision is the percentage of data classified as a good model which is actually also good. Recall is a measurement of the actual level of positive recognition (Tsami et al, 2018).

## RESULT AND DISCUSSION

The results and stages of the discussion process in the C4.5 classification algorithm are as follows:

4.1 Data Collection

The data used is secondary data obtained from the Information Systems Study Program of Universitas Teknokrat Indonesia.

Preprocessing data is done by removing redundant name, changing format in GPA column, and transforming competition and organization column into categorical data.

4.2 Preprocessing Data

**Table 1.** Raw data of Information System Department Student who apply scholarship

NPM	Name	Study Program	Semester	GPA	Competition	Organization	
18311006	Ade Rahman	S1 Sistem Informasi		5.3.80	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	UKMI Ar Rahman, UKM Protek	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
18311243	Adelia Rahmawati	S1 Sistem Informasi		5.3.51	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
19311062	ADELLA ESIKA DEWI	S1 Sistem Informasi		3.3.30	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	UKM KARATEHIMA SI	
17311001	ADITYA FAJAR RAMADHAN	S1 Sistem Informasi		7.3.58	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
17311001	ADITYA FAJAR RAMADHAN	S1 Sistem Informasi		7.3.58	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	-	
19311113	Adzqi Aulia	S1 Sistem Informasi		3.3.78		UKM Panahan	
18311282	Ahmad Ridho	S1 Sistem Informasi		5.3.52	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	Ukmi Ar-Rohman	
18311282	Ahmad Ridho	S1 Sistem Informasi		5.3.52	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	UKM Ukmi Ar-Rohman	
19311171	Aisyah vannie Bachtiar	S1 Sistem Informasi		3.3.64		-	
19311204	Ajid Mahendra	S1 Sistem Informasi		3.3.50	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	HIMA	
18311250	Aman swasono	S1 Sistem Informasi		5.3.46	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	Protek Programing Teknokrat	
19311274	Andharu Nur Trio Fajri	S1 Sistem Informasi		3.3.81	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	UKM Basket Teknokrat	
17311268	Andhita Sopiyan	S1 Sistem Informasi		7.3.61	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	Hima	
18311371	Anisa Septiawati	S1 Sistem Informasi		5.3.45	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	1. BEM FTIK Teknokrat2. Ukm Duta Teknokrat3. Ukm Teknokrat Tv	
18311021	Anjellah Berliana Sherly Ayuni	S1 Sistem Informasi		5.3.75	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	Robotik	
17311341	Aryan Manda Saputra	S1 Sistem Informasi		7	3.5	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	Himpunan Mahasiswa Sistem Informasi (HIMASI)
19311010	BAY HAQI MUHAMMAD	S1 Sistem Informasi		3.3.05	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	1. TEKNOKRAT TV 2. KARATE 3. HIMA SISTEM INFORMASI	
18311203	CINDY MAYCA	S1 Sistem Informasi		5.3.00	<a href="https://drive.google.com/open?id=">https://drive.google.com/open?id=</a>	MENWA	

**Table 2.** Processed data of Information System Department Student who apply scholarship

NPM	Name	Study Program	Semeste	GP	Competitio	Organizatio	Sholarshij
18311006	Ade Rahman	S1 Sistem Informasi	5	3,8	Yes	Yes	Accepted
18311243	Adelia Rahmawati	S1 Sistem Informasi	5	3,51	Yes	No	Accepted
19311062	ADELLA ESIKA DEWI	S1 Sistem Informasi	3	3,3	Yes	Yes	Accepted
17311001	ADITYA FAJAR RAMADHAN	S1 Sistem Informasi	7	3,58	Yes	No	Rejected
19311113	Adzqi Aulia	S1 Sistem Informasi	3	3,78	No	Yes	Accepted
18311282	Ahmad Ridho	S1 Sistem Informasi	5	3,52	Yes	Yes	Accepted
19311171	Aisyah vannie Bachtiar	S1 Sistem Informasi	3	3,64	No	No	Accepted
19311204	Ajid Mahendra	S1 Sistem Informasi	3	3,5	Yes	Yes	Accepted
18311250	Aman swasono	S1 Sistem Informasi	5	3,46	Yes	Yes	Accepted
19311274	Andharu Nur Trio Fajri	S1 Sistem Informasi	3	3,81	Yes	Yes	Accepted
17311268	Andhita Sopiyan	S1 Sistem Informasi	7	3,61	Yes	Yes	Accepted
18311371	Anisa Septiawati	S1 Sistem Informasi	5	3,45	Yes	Yes	Accepted
18311021	Anjellah Berliana Sherly Ayuni	S1 Sistem Informasi	5	3,75	Yes	Yes	Accepted
17311341	Aryan Manda Saputra	S1 Sistem Informasi	7	3,5	Yes	Yes	Accepted
19311010	BAY HAQI MUHAMMAD	S1 Sistem Informasi	3	3,05	Yes	Yes	Accepted
18311203	CINDY MAYCA	S1 Sistem Informasi	5	3	Yes	Yes	Rejected
18311170	Dani Hidayatullah	S1 Sistem Informasi	5	3,63	Yes	Yes	Accepted
17311332	DEBY SINTIA AMELIA	S1 Sistem Informasi	7	3,84	Yes	Yes	Accepted
17311199	Dedy Prasetyo	S1 Sistem Informasi	7	3,38	Yes	Yes	Rejected
18311023	Dena mellynia	S1 Sistem Informasi	5	3,33	Yes	Yes	Accepted
17311028	Desi Handayani	S1 Sistem Informasi	7	3,43	No	Yes	Rejected
18311158	Desi Maryana Ulpa	S1 Sistem Informasi	5	3,27	Yes	Yes	Rejected
19311144	Devi Juita	S1 Sistem Informasi	3	3,24	Yes	Yes	Accepted
18311378	Dita Aldina Rosmini	S1 Sistem Informasi	5	3,23	No	Yes	Rejected

4.3 Feature Selection

The variables selected for calculating decision tree model are GPA,

Competition, Organization, and Scholarship.

**Table 3.** Features used for calculating decision tree model

GPA	Competition	Organization	Scholarship
Excellent	Yes	Yes	Accepted
Excellent	Yes	No	Accepted
Very Good	Yes	Yes	Accepted
Excellent	Yes	No	Rejected
Excellent	No	Yes	Accepted
Excellent	Yes	Yes	Accepted
Excellent	No	No	Accepted
Very Good	Yes	Yes	Accepted
Very Good	Yes	Yes	Accepted
Excellent	Yes	Yes	Accepted
Excellent	Yes	Yes	Accepted
Very Good	Yes	Yes	Accepted
Excellent	Yes	Yes	Accepted
Very Good	Yes	Yes	Accepted
Very Good	Yes	Yes	Accepted
Very Good	Yes	Yes	Accepted
Good	Yes	Yes	Rejected
Excellent	Yes	Yes	Accepted
Excellent	Yes	Yes	Accepted
Very Good	Yes	Yes	Rejected
Very Good	Yes	Yes	Accepted
Very Good	No	Yes	Rejected
Very Good	Yes	Yes	Rejected
Very Good	Yes	Yes	Accepted
Very Good	No	Yes	Rejected

4.4 Model Calculation

Using the formula (1), we can calculate the Entropy value. Calculating Total Entropy is done by calculating the number of decisions "ACCEPTED" and "REJECTED" from all existing cases.

$$\begin{aligned}
 Entropy (Total) &= \left( -\frac{86}{124} \times \log_2 \left( \frac{86}{124} \right) \right) \\
 &+ \left( -\frac{38}{124} \times \log_2 \left( \frac{38}{124} \right) \right) \\
 &= 0,88903
 \end{aligned}$$

Entropy Total is calculating the total value of accepted decisions (86) and rejected decisions (38), while 124 is the total number of cases. Then calculate each Entropy of all existing variable values. And by using formula (2) we can calculate the value of Information Gain in each Attribute.

$$\begin{aligned}
 Gain (Total, GPA) &= 0,88903 \\
 &- \left( \left( \frac{7}{124} \times 0 \right) \right) \\
 &+ \left( \frac{67}{124} \times 0,97267 \right) \\
 &+ \left( \frac{50}{124} \times 0,40218 \right) \\
 &= 0,2013
 \end{aligned}$$

$$\begin{aligned}
 Gain (Total, Competition) &= 0,88903 \\
 &- \left( \left( \frac{89}{124} \times 0,82436 \right) \right) \\
 &+ \left( \frac{35}{124} \times 0,98522 \right) \\
 &= 0,01927
 \end{aligned}$$

$$\begin{aligned}
 Gain (Total, Organization) &= 0,88903 \\
 &- \left( \left( \frac{108}{124} \times 0,85241 \right) \right) \\
 &+ \left( \frac{16}{124} \times 1 \right) \\
 &= 0,01759
 \end{aligned}$$

After all the Values of Entropy and Information Gain are calculated, then the result of that calculation is

entered into table 3. From the calculation of entropy value and information gain above, it can be known that the largest information gain value is GPA with a value of 0,2013 and the smallest is Organization, with a value of 0,01759. Then, grab the largest Information Gain value make it the root node. Remove the previously

selected attribute and repeat the calculation of the Entropy value, Information Gain, by selecting the largest Information Gain and being used as the internal node of the tree. Repeat the calculation until all tree attributes have a class. The table below shows the entire set of calculation result.

**Table 4.** Entropy and Information Gain Calculation Table of all attributes in Node 1

		Total	Accepted	Rejected	Entropy	Information Gain
<b>Scholarship</b>		124	86	38	0,889034904	
<b>GPA</b>						0,201310316
	Good	7	0	7	0	
	Very Good	67	40	27	0,972669991	
	Excellent	50	46	4	0,40217919	
<b>Competition</b>						0,019266168
	Yes	89	66	23	0,824363354	
	No	35	20	15	0,985228136	
<b>Organization</b>						0,017585232
	Yes	108	78	30	0,852405179	
	No	16	8	8	1	

From the table above we can see that GPA has the number of information gain so that is will be the root of the decision tree. Then we continue the

calculation to find the node. In the following shows the result of calculation.

**Table 5.** Entropy and Information Gain Calculation Table of all attributes in Node 1.1

1.1

		Total	Accepted	Rejected	Entropy	Information Gain
<b>GPA Very Good</b>		67	40	27	0,972669991	
<b>Competition</b>						0,397794334
	Yes	46	32	14	0,886540893	
	No	21	8	13	0,958711883	
<b>Organization</b>						0,38823407
	Yes	57	37	20	0,934849024	
	No	10	3	7	0,881290899	

**Table 6.** Entropy and Information Gain Calculation Table of all attributes in Node 1.2

		Total	Accepted	Rejected	Entropy	Information Gain
<b>GPA</b>	<b>Excellent</b>	50	40	4	0,549050971	
<b>Competition</b>	Yes	38	34	4	0,485460761	0,740264671
	No	12	12	0	0	
<b>Organization</b>	Yes	44	41	3	0,359101626	0,730159049
	No	6	5	1	0,650022422	

4.5 Decision Tree

From the results of the calculation of entropy and information gain then we can

process into the Decision Tree. Here are the results of the Decision Tree:

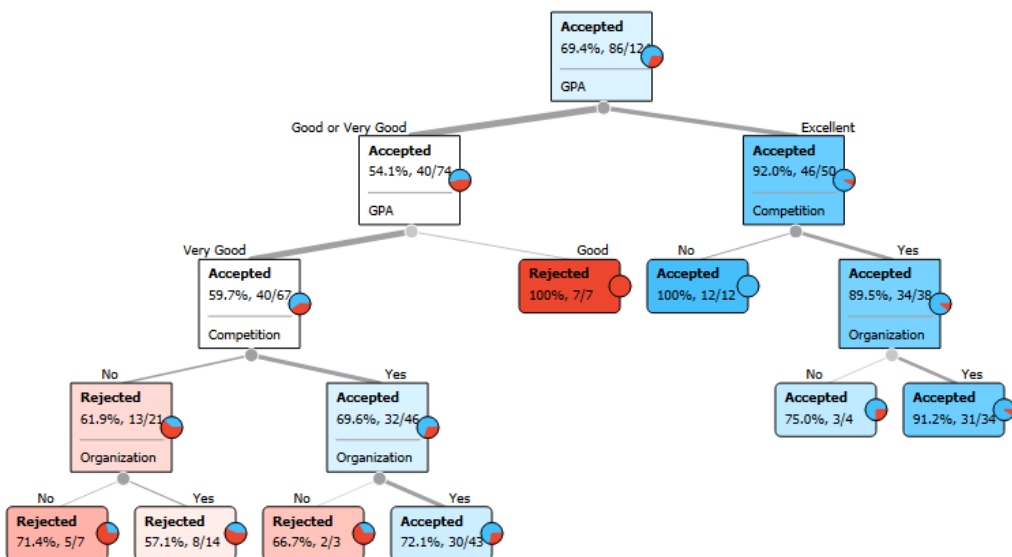


Figure 1. Decision tree model

4.6 Model Rules

From the decision tree formed in Figure 1, there is a rule model in determining the recommendation of sales partner acceptance. There are 14 rules that are formed, can be seen as follows:

*If* GPA = Good *then* Scholarship = Rejected  
*If* GPA = Very Good and Competition = No, Organization = No *then* Scholarship = Rejected



If GPA = Very Good and Competition = Yes, Organization = No then Scholarship = Rejected  
 If GPA = Very Good and Competition = No, Organization = Yes then Scholarship = Rejected  
 If GPA = Very Good and Competition = Yes, Organization = Yes then Scholarship = Accepted  
 If GPA = Excellent and Competition = No then Scholarship = Accepted  
 If GPA = Excellent and Competition = Yes, Organization = No then Scholarship = Accepted  
 If GPA = Excellent and Competition = Yes, Organization = Yes then Scholarship = Accepted

decisions and rules resulting from this research relate to the following:  
 Students with good GPA will be rejected from getting scholarship.  
 Students with good GPA are considered to get scholarship if they join competition and organization.  
 Students with excellent GPA will be accepted as awardee without considering competition or organization.

#### 4.7 Validation and Testing

Testing is done with cross-validation. One type of cross validation is ten-fold cross validation. Here are the results of confusion matrix and ten-fold cross validation using Python:

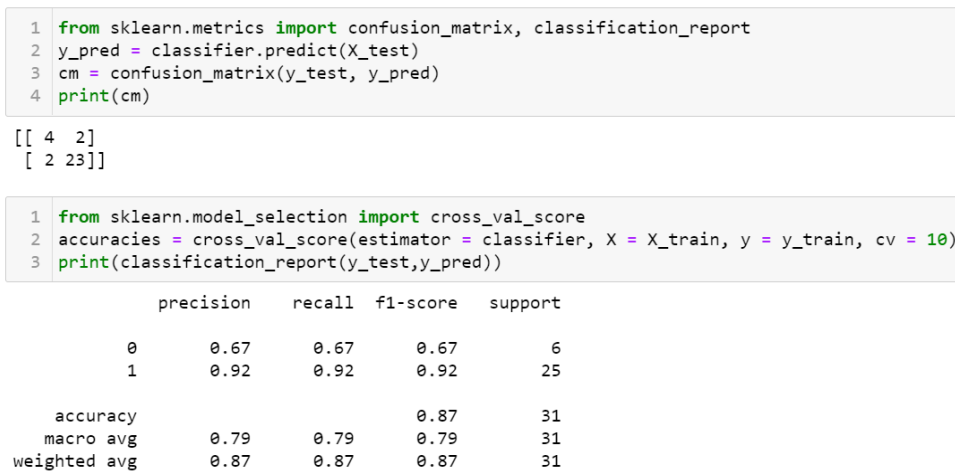


Figure 2. Confusion matrix and ten-fold cross validation results

Based on testing using the Ten-Fold Cross Validation method, the accuracy value is 87 %, the precision is 87 % and the recall is 87 %. This indicates that from the classification process carried out will be able to be applied to the recommendation of

acceptance of prospective sales partners.

#### CONCLUSION

The conclusions of the research are as follows:

1. C4.5 classification algorithm will be implemented on the decision of tuition aid grant program at Universitas Teknokrat Indonesia, judging by the accuracy, recall, and precision level of 87 % simultaneously, the calculations carried out will be able to predict and recommend the model well.
2. 8 rules model can be used as a reference in the design and creation of GUI applications.
3. This algorithm can be used to determine which students will get scholarship.

## REFERENCES

- Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, pp. 528–533, 2016, doi: 10.7763/ijiet.2016.v6.745.
- Alsagheer, A. F. H. Alharan, and A. S. A. Al-Haboobi, "Popular Decision Tree Algorithms of Data Mining Techniques: A Review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 6, no. 6, pp. 133–142, 2017.
- Ben Jabeur, A. Sadaoui, A. Sghaier, and R. Aloui, "Machine learning models and cost-sensitive decision trees for bond rating prediction," *J. Oper. Res. Soc.*, vol. 71, no. 8, pp. 1161–1179, 2020, doi: 10.1080/01605682.2019.1581405.
- Bekesiene and S. Hoskova-Mayerova, "Decision tree-Based classification model for identification of effective leadership indicators," *J. Math. Fundam. Sci.*, vol. 50, no. 2, pp. 121–141, 2018, doi: 10.5614/j.math.fund.sci.2018.50.2.2.
- Chermiti, "Establishing risk and targeting profiles using data mining: Decision trees," *World Cust. J.*, vol. 13, no. 2, pp. 39–58, 2019.
- Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 5, no. 2, p. 26, 2018, doi: 10.9781/ijimai.2018.02.004.
- Mesarić and D. Šebalj, "Decision trees for predicting the academic success of students," *Croat. Oper. Res. Rev.*, vol. 7, no. 2, pp. 367–388, 2016, doi: 10.17535/crorr.2016.0025.
- Mittal, D. Khanduja, and P. Tewari, "An Insight into 'Decision Tree

- Analysis,” *Int. J. Peer Rev. J. Ref. J. Index. J. UGC Approv. J. Impact Factor*, vol. 3, no. 12, pp. 111–115, 2017, [Online]. Available: [www.wvjmr.com](http://www.wvjmr.com).
- Nowak, “Defining Project Approach using Decision Tree and Quasi-hierarchical Multiple Criteria Method,” *Procedia Eng.*, vol. 172, pp. 791–799, 2017, doi: 10.1016/j.proeng.2017.02.125.
- Patel and P. Prajapati, “Study and Analysis of Decision Tree Based Classification Algorithms,” *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74–78, 2018, doi: 10.26438/ijcse/v6i10.7478.
- Petropoulos, V. Siakoulis, E. Stavroulakis, and A. Klamargias, “A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting,” *use big data Anal. Artif. Intell. Cent. Bank.*, vol. 50, no. August, pp. 30–31, 2018, [Online]. Available: [https://www.bis.org/ifc/publ/ifcb/49\\_49.pdf](https://www.bis.org/ifc/publ/ifcb/49_49.pdf).
- Rismayanti, “Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan ),” *J. Sist. Inf.*, vol. 02, no. 01, pp. 16–24, 2018.
- Sharma and S. Kumar, “A Survey on Decision Tree Algorithms of Classification in Data Mining,” *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.
- Tsami, G. Adamos, E. Nathanail, E. Budilovich, I. Yatskiv, and V. Magginas, “A decision tree approach for achieving high customer satisfaction at urban interchanges,” *Transp. Telecommun.*, vol. 19, no. 3, pp. 194–202, 2018, doi: 10.2478/ttj-2018-0016