

# Algoritma C4.5 in mapping the admission patterns of new Students in Engineering Computer

Yunita Sari Siregar<sup>1)</sup>, Boni Oktaviana Sembiring<sup>2)</sup>, Hasdiana<sup>3)</sup>, Arie Rafika Dewi<sup>4)</sup>,  
Herlina Harahap<sup>5)</sup>

<sup>1)2)3)4)5)</sup> Universitas Harapan Medan, Indonesia

<sup>1)</sup>[yunitasarisiregar1990@gmail.com](mailto:yunitasarisiregar1990@gmail.com), <sup>2)</sup>[bonioktaviana@yahoo.co.id](mailto:bonioktaviana@yahoo.co.id), <sup>3)</sup>[hasdiana.stth@gmail.com](mailto:hasdiana.stth@gmail.com),

<sup>4)</sup>[arie.juny@gmail.com](mailto:arie.juny@gmail.com), <sup>5)</sup>[Herlina\\_Hrp@yahoo.com](mailto:Herlina_Hrp@yahoo.com)

**Submitted** : Sep 26, 2021 | **Accepted** : Oct 10, 2021 | **Published** : Oct 10, 2021

**Abstract:** University of Harapan Medan is one of the private universities in North Sumatra which has computer-based study programs such as Informatics Engineering and Information Systems. Every year this college receives many registrations from students who have completed their education at the school stage. The large number of incoming student data makes it difficult for the admin to select new students who will register. In this study using the C4.5 algorithm data mining method to map the pattern of student admissions selection in the field of Engineering and computers. The attributes used are the average value of report cards (high, enough, low), basic academic ability tests (very high, high, medium, low, very low), basic computer knowledge tests (very high, high, enough, low, very low) and interviews (good, bad). Data mining is a mathematical calculation process that uses algorithms and requires large data. While the C4.5 algorithm is an algorithm that processes data by calculating entropy and information gain, where after the calculation process is carried out, those who get the largest information gain value will become nodes and branches. This C4.5 algorithm will describe a decision tree that will form a pattern in student selection. The results of this study indicate that in mapping the selection pattern of interview attributes into level 1 nodes, the attributes of the basic computer knowledge test become the level 1 branch, the attributes of the basic academic ability test become the level 2 branch and the attribute average value of report cards becomes the level 3 branch.

**Keywords:** University of Harapan Medan, Engineering and Computers, Students, Data Mining, Algorithm C4.5

## INTRODUCTION

Harapan University Medan is one of the private universities in North Sumatra which has computer-based study programs such as Informatics Engineering and Information Systems. Every year, many students register to continue their education to the tertiary level. The students who registered came from High School (SMA), Vocational High School (SMK), and Madrasah Aliyah (MA). The large number of incoming student data makes it difficult for the admin to do the selection of new student admissions. One of the methods used in mapping student selection patterns is data mining algorithm C4.5. The C4.5 algorithm is an algorithm that performs data processing by calculating entropy and information gain. Where after the process of calculating the information gain value, the values obtained for each attribute will be compared. And the ones that get the greatest value will (Handayani, 2019) Data mining is a series of processes to explore the added value of a set of data in the form of knowledge that has not been known manually. Data mining actually has long roots in fields of science such as artificial intelligence, machine learning, statistics and databases. Some of the techniques that are often mentioned in the Data Mining literature include clustering, classification, association rule mining, neural network, and genetic algorithms (Handoko, 2016).

\*Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The C4.5 algorithm is one of the algorithms contained in data mining classification and is a model or function that explains or distinguishes concepts or data classes with the aim of estimating the unknown class of an object. In general, the C4.5 algorithm to build a decision tree is to choose an attribute as the root, create a branch for each value, divide the cases into branches, then repeat the process for each branch until all cases in the branch have the same class (Handayani, 2019)

Previous research conducted by (Lukhayu Pritalia, 2018), concluded that data mining with the C4.5 algorithm can be implemented to determine the availability of e-commerce goods with two categories, namely the goods are not in a condition or not yet available for ordering goods for customers/backorders and goods are in available condition for ordering goods for customers and the results of the analysis using the C4.5 algorithm to determine the time of determining the availability of goods have an accuracy rate of 98.9%.

Based on the above background, the problem in this study is to map the selection pattern of new student admissions in the field of Engineering and Computers using the C4.5 algorithm. And the purpose of this research is to get a pattern mapping in the form of a decision tree, making it easier for admins to select new student admissions in accordance with the fields of engineering and computers.

### LITERATURE REVIEW

Knowledge Discovery In Database (KDD) is a method to obtain knowledge from existing databases. In the database there are tables that are interconnected / related. The results of the knowledge obtained in the process can be used as a knowledge base for decision-making purposes (Yuli Mardi, 2019). Knowledge Discovery In Database (KDD) is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data and relationships with integration techniques and scientific discoveries, interpretation and visualization of patterns in a number of data sets (Elisa, 2017). Knowledge discovery as a process consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation). Data mining refers to the process of mining knowledge from a very large set of data (Firdaus, 2017).

Data mining is defined as the process of finding patterns in data. This process is automated or often semi-automatic. The pattern found must be meaningful and the pattern provides benefits, usually economic benefits. Large amount of data required (Siregar & Harliana, 2018b). Data mining is an iterative and interactive process to find new patterns or models that are perfect, useful and understandable in a massive database. Data mining contains a search for trends or patterns desired in a large database to help decision makers in the future, these patterns are recognized by certain devices that can provide a useful and insightful analysis of data that can then be studied more thoroughly, which may use other decision support tools (Sikumbang, 2018).

Data mining is between the branches of computer science and statistics, utilizing advances from both disciplines to extract information from large databases. Data mining is a series of processes to explore the added value of knowledge that has not been known from a data set. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to identify useful information and related knowledge from large databases (Lukhayu Pritalia, 2018).

The characteristics of data mining are as follows (Siregar & Harliana, 2018a):

- a. Data mining deals with the discovery of something hidden and certain data patterns that were not previously known.
- b. Data mining usually uses very large data. Usually big data is used to make the results more reliable.
- c. Data mining is useful for making critical decisions, especially in strategy.

Data mining techniques are divided into several techniques. Some of the techniques and properties of data mining are as follows (Merdekawati, 2018) :

- a. Classification Determines a new data record of one of several pre-defined categories (or classes).
- b. Regression predicts the value of a given continuous variable based on the value of another variable, assuming a linear or non-linear dependency model. This technique is widely studied in statistics, the field of artificial neural networks.
- c. Clustering partitions data sets into sub-sets or groups in such a way that elements of a particular group have a set of shared properties, with a high degree of similarity in one group and a low level of similarity between groups. Also called unsupervised learning.
- d. Association Rules detect a set of attributes that appear together (co-occur) in frequent frequencies and form a number of rules of those sets.
- e. Sequential (Sequence Mining) Looking for a number of events that generally occur together

Classification is the process of finding a set of models/functions that describe and distinguish data into specific classes, with the aim of using that model in determining the class of an object of unknown class (Handayani, 2019). The purpose of classification is to find a model of a training set that distinguishes attributes into corresponding categories or classes, the model is then used to classify attributes whose classes have not been previously known (Azwanti, 2018).

\*Corresponding author



Data mining classification is the placement of objects into one of several predefined categories. Classification is widely used to predict classes on a certain label, namely by classifying data (building models) based on training sets and values (class labels) in classifying certain attributes and using them in classifying new data (Susanto & Sudiyatno, 2014). There are two classification processes consisting of: the learning/training process is to build a model using training data, where each record in the training data is analyzed based on its attribute values using the C4.5 algorithm to obtain the model. Then the testing process is to test the testing data using the model that has been obtained from the training process (Handayani, 2019).

Decision tree is the most popular classification method used. Apart from being relatively fast in construction, the results of the built model are easy to understand. The decision tree works by forming a decision tree that can be concluded with certain classification rules, one of which is C4.5 (Azwanti, 2018). Decision trees have several advantages, namely easy to develop a model, easy to understand by users, and able to handle noisy data and unknown data. Decision tree consists of several parts, namely inside nodes, branches, and leaf nodes (Pratama & Hanum, 2016).

The C4.5 algorithm is one method for making a decision tree based on the training data that has been provided. Some of the developments carried out on the C4.5 are, among others, being able to overcome missing values, being able to handle continuous data, and pruning. The C4.5 algorithm was introduced by J. Ross Quinlan which is the development of the ID3 algorithm, the algorithm is used to form a decision tree. Decision tree is considered as one of the most popular approaches, in the decision tree classification consists of a node that forms a root, the root node has no input (Elisa & Harman, 2019). C4.5 is part of the algorithm for classification in machine learning and data mining. C4.5 is an algorithm that is suitable for classification problems in machine learning and data mining (Fitriana, Ripanti, & Tursina, 2018).

There are several stages in making a decision tree with the C4.5 algorithm, namely (Merdekawati, 2018):

- a. Prepare training data. Training data is usually taken from historical data that has happened before and has been grouped into certain classes.
- b. Determine the root of the tree. The root will be taken from the selected attribute, by calculating the gain value of each attribute, the highest gain value will be the first root. Before calculating the gain value from the attribute, first calculate the entropy value. To calculate the entropy value used the formula :

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

Information :

S = Case Set

n = number of partitions S

Pi = Proportion of Si to S

- c. Then calculate the gain value using the formula:

$$Gain (S,A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i) \tag{2}$$

Information :

S = Case Set

A = Features

n = Number of attribute partitions A

|Si| = Proportion of Si to S

|S| = Number of cases in S

- d. Repeat step 2 until all records are partitioned.
- e. The decision tree partitioning process will stop when:
  - 1) All records in node N have the same class.
  - 2) There are no attributes in the partitioned record anymore.
  - 3) There are no records in the empty branch.

## METHOD

At this stage will be explained about the methodology in determining the pattern of selection of new students using the data mining algorithm c4.5 as follows:

- a. Analysis and Identification of Problems. This stage is a step to identify the problem to be discussed, namely the use of data mining algorithm C4.5 in mapping the selection pattern of new students in the computer field.
- b. Studying Literature. At this stage regarding theoretical studies that support research

\*Corresponding author



- c. Data collection. In this study, the data that will be used are 70 prospective new students who register at Harapan University. Where the attributes used are: the average value of report cards, basic academic ability tests, basic computer knowledge tests and interviews.
- d. Analysis and testing methods In the data analysis stage, it will be done by solving the data mining algorithm C4.5. The resolution steps include:
  - 1) Perform entropy and gain calculations then select attributes as roots
  - 2) Perform entropy and gain calculations and create branches for each value
  - 3) Split cases in branches
  - 4) Repeat the process in each branch, until all cases in the branch have the same class.

**RESULT**

The attribute variables in the implementation of data mining algorithm C4.5 in mapping the selection pattern of new students in the field of Engineering and computers can be seen in table 1.

Table 1. Attribute Variables

Name	Attribute
A1	Average Report (Rapor)
A2	Basic Academic Ability Test (TKDA)
A3	Basic Computer Knowledge Test (TPDK)
A4	Interview

The process of pattern mapping in the selection of new students in the field of Engineering and computers uses 70 prospective student data. Can be seen in table 2.

Table 2. Prospective Student Data

Students	Attribute				Decision
	A1	A2	A3	A4	
1	High	Very High	Enough	Good	Accepted
2	High	High	Very High	Good	Accepted
3	Enough	Very High	Enough	Bad	Accepted
4	Low	High	Low	Bad	Rejected
5	High	Very High	Very High	Good	Accepted
6	Enough	Very High	Very Low	Bad	Rejected
7	High	High	Very Low	Bad	Rejected
8	Low	Very High	Very Low	Good	Accepted
9	Low	Medium	High	Bad	Rejected
10	Enough	High	High	Bad	Accepted

To get which attribute will be node 1, it is necessary to calculate the total entropy. The amount of data used as the research sample amounted to 70, of which 50 prospective students had accepted results, and 20 prospective customers were rejected. To calculate the total entropy value and the entropy value for each attribute with equation 1. after getting the entropy value for each attribute, you can calculate the gain value for each attribute with equation 2 so that the results are in table 3 below.

Table 3. Calculation of Node 1

Node 1	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
	Total		70	50	20	0.8631206	
	Evaluation						0.010738
	Average Report	High		21	16	5	0.7918584
Enough			24	18	6	0.8112781	
Low			25	16	9	0.9426832	

\*Corresponding author



	Evaluation						0.0772306
Basic Academic Ability Test	Very high	16	14	2	0.5435644		
	High	16	13	3	0.6962123		
	Medium	12	9	3	0.8112781		
	Low	13	6	7	0.9957275		
	Very low	13	8	5	0.9612366		
	Evaluation						0.3319133
Basic Computer Knowledge Test	Very high	14	14	0	0		
	Tall	15	14	1	0.3533593		
	Enough	16	13	3	0.6962123		
	Low	13	7	6	0.9957275		
	Very low	12	2	10	0.6500224		
	Evaluation						1.0116341
Interview	Good	22	5	27	0.1231913		
	Bad	28	15	43	0.4680771		

From table 3, it can be seen the entropy value and information gain value for each attribute. To get the decision tree node 1, can use the highest gain value. In the decision tree search, the largest gain value is in the interview attribute, which is 1.0116341. Thus the interview can be used as node 1. There are 2 classifications on the interview attributes, namely good and bad. Where a good classification becomes a level 1.1 branch and a bad classification becomes a level 1.2 branch.

The next stage is to calculate the value of entropy and information gain at branch level 1.1. The calculation results can be seen in table 4.

Table 4. Calculation of Branch Level 1.1

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
1.1	Interview	Good	27	22	5	0.691289869	
	Evaluation						0.4539588
	Average Report	High	10	7	3	0.881290899	
		Enough	12	10	2	0.650022422	
		Low	5	5	0	0	
	Evaluation						0.5268645
	Basic Academic Ability Test	Very High	8	8	0	0	
		High	1	1	0	0	
		Medium	6	6	0	0	
		Low	6	3	3	1	
		Very low	6	4	2	0.918295834	
	Evaluation						0.5825807
	Basic Computer Knowledge Test	Very High	7	7	0	0	
		High	4	4	0	0	
		Enough	8	8	0	0	
		Low	3	1	2	0.918295834	
		Very low	5	2	3	0.970950594	

From the results in table 4 it can be seen that the attribute with the highest gain is a basic computer knowledge test of 0.5825807. . Thus, the attributes of the basic computer knowledge test can be used as a level 1.1 branch. There are 2 classifications on the attributes of the basic computer knowledge test, namely low and very low. Where the low classification becomes a level 2.1 branch and the poor classification becomes a level 2.2 branch.

The next stage is to calculate the value of entropy and information gain at branch level 1.2. The calculation results can be seen in table 5.

Table 5. Branch Calculation Level 1.2

\*Corresponding author



Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
1.2	Interview	Bad	43	28	15	0.933025295	
	Evaluation						0.0400146
	Average Report	High	11	9	2	0.684038436	
		Enough	12	8	4	0.918295834	
		Low	20	11	9	0.992774454	
	Evaluation						0.0165268
	Basic Academic Ability Test	VeryHigh	8	6	2	0.811278124	
		High	14	8	3	0.937572617	
		Medium	6	3	3	1	
		Low	7	3	4	0.985228136	
		Very Low	7	4	3	0.985228136	
	Evaluation						0.4172243
	Basic Computer Knowledge Test	Very High	7	7	0	0	
		High	11	10	1	0.439496987	
		Enough	8	5	3	0.954434003	
Low		10	6	4	0.970950594		
VeryLow		7	0	7	0		

From the results in table 5, it can be seen that the attribute with the highest gain is a basic computer knowledge test of 0.4172243. Thus the attributes of the basic computer knowledge test can be used as a level 1.2 branch. There are 3 classifications on the attributes of the basic computer knowledge test, namely high, enough and low. Where the high classification becomes a level 2.3 branch, the classification simply becomes a level 2.4 branch and the low classification becomes a level 2.5 branch.

The next stage is to calculate the value of entropy and information gain at branch level 2.1. The calculation results can be seen in table 6.

Table 6. Calculation of Branch Level 1.2

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
2.1	TPDK	Low	3	1	2	0.918295834	
	Evaluation						0.2516292
	Average Report	High	2	1	1	1	
		Enough	1	0	1	0	
	Evaluation						0.9182958
	Basic Academic Ability Test	Medium	1	1	0	0	
		Low	1	0	1	0	
Very Low		1	0	1	0		

From the results of table 6 it can be seen that the attribute with the highest gain is the basic academic ability test of 0.9182958. Thus the attributes of the basic academic ability test can be used as a level 2.1 branch. There are 3 classifications on the attributes of the basic academic ability test, namely medium, low and very low. In the attribute of the basic academic ability test, each classification gets an entropy value of 0.

The next stage is to calculate the value of entropy and information gain at branch level 2.2. The calculation results can be seen in table 7.

Table 7. Calculation of Branch Level 2.2

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
2.2	TPDK	Very low	5	2	3	0.970950594	
	Evaluation						0.5709506
		High	2	0	2	0	

\*Corresponding author





	Average Report	Enough	2	1	1	1	
		Low	1	1	0	0	
	Evaluation						0.9709506
	Basic Academic Ability Test	Very High	1	1	0	0	
		Medium	1	1	0	0	
		Low	2	0	2	0	
		Very low	1	0	1	0	

From the results of table 7 it can be seen that the attribute with the highest gain is the basic academic ability test of 0.9709506. Thus the attributes of the basic academic ability test can be used as a level 2.2 branch. There are 4 classifications on the attributes of the basic academic ability test, namely very high, medium, low, and very low. In the attribute of the basic academic ability test, each classification gets an entropy value of 0.

The next stage is to calculate the value of entropy and information gain at branch level 2.3. The calculation results can be seen in table 8.

Table 8. Calculation of Branch Level 2.3

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
2.3	TPDK	High	11	10	1	0.439496987	
	Evaluation						0.1113479
	Average Report	High	4	4	0	0	
		Enough	2	2	0	0	
		Low	5	4	1	0.721928095	
	Evaluation						0.1890527
	Basic Academic Ability Test	Very High	1	1	0	0	
		High	4	4	0	0	
		Medium	3	2	1	0.918295834	
		Low	1	1	0	0	
		Very Low	2	2	0	0	

From the results in table 8 it can be seen that the attribute with the highest gain is the basic academic ability test of 0.1890527. Thus the attributes of the basic academic ability test can be used as a level 2.3 branch. There are 5 classifications on the attributes of the basic academic ability test, namely very high, high, medium, low and very low. Where the classification is being branch level 3.1

The next stage is to calculate the value of entropy and information gain at branch level 2.4. The calculation results can be seen in table 9

Table 9. Calculation of Branch Level 2.4

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
2.4	TPDK	Enough	8	5	3	0.954434003	
	Evaluation						0.0157121
	Average Report	High	2	1	1	1	
		Enough	3	2	1	0.918295834	
		Low	3	2	1	0.918295834	
	Evaluation						0.954434
	Basic Academic Ability Test	Very High	2	2	0	0	
		High	3	3	0	0	
		Low	2	0	2	0	
		Very Low	1	0	1	0	

\*Corresponding author



From the results of table 9 it can be seen that the attribute with the highest gain is the basic academic ability test of 0.954434. Thus the attributes of the basic academic ability test can be used as a level 2.4 branch. There are 4 classifications on the attributes of the basic academic ability test, namely very high, high, low, and very low. In the attribute of the basic academic ability test, each classification gets an entropy value of 0.

The next stage is to calculate the value of entropy and information gain at branch level 2.5. The calculation results can be seen in table 10.

Table 10. Calculation of Branch Level 2.5

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
2.5	TPDK	Low	10	6	4	0.970950594	
	Evaluation						0.3219281
	Average Report	High	2	2	0	0	
		Enough	4	3	1	0.811278124	
		Low	4	1	3	0.811278124	
	Evaluation						0.6464393
	Basic Academic Ability Test	Very High	3	3	0	0	
		High	4	3	1	0.811278124	
		Medium	1	0	1	0	
		Low	1	0	1	0	
Very Low		1	0	1	0		

From the results in table 10, it can be seen that the attribute with the highest gain is the basic academic ability test of 0.6464393. Thus the attributes of the basic academic ability test can be used as a level 2.5 branch. There are 5 classifications on the attributes of the basic academic ability test, namely very high, high, medium, low and very low. Where the high classification becomes a level 3.2 branch.

The next stage is to calculate the value of entropy and information gain at branch level 3.1. The calculation results can be seen in table 11.

Table 11. Calculation of Branch Level 3.1

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
3.1	TKDA	Medium	3	2	1	0.918295834	
	Evaluation						0.9182958
	Average Report	High	1	1	0	0	
		Enough	1	1	0	0	
		Low	1	0	1	0	

From the results in table 11, it can be seen that the average report card attribute is the last level branch with a gain value of 0.9182958. There are 3 classifications on the average report card attribute, namely high, enough, and low.

The next stage is to calculate the value of entropy and information gain at branch level 3.2. The calculation results can be seen in table 12..

Table 12. Calculation of Branch Level 3.2

Node	Attribute	Class	Number of Cases (S)	Accepted	Rejected	Entropy	Gain
3.2	TKDA	Low	4	3	1	0.811278124	
	Evaluation						0.8112781
	Average Report	High	1	1	0	0	
		Enough	2	2	0	0	
		Low	1	0	1	0	

\*Corresponding author





From the results in table 12, it can be seen that the average report card attribute is the last level branch with a gain value of 0.8112781. There are 3 classifications on the average report card attributes, namely high, enough, and low.

### DISCUSSIONS

Based on the results of calculations that have been carried out using the C4.5 algorithm data mining method in mapping the selection pattern of new students in the field of Engineering and computers to get an overview of the decision tree. It can be seen in Figure 1.

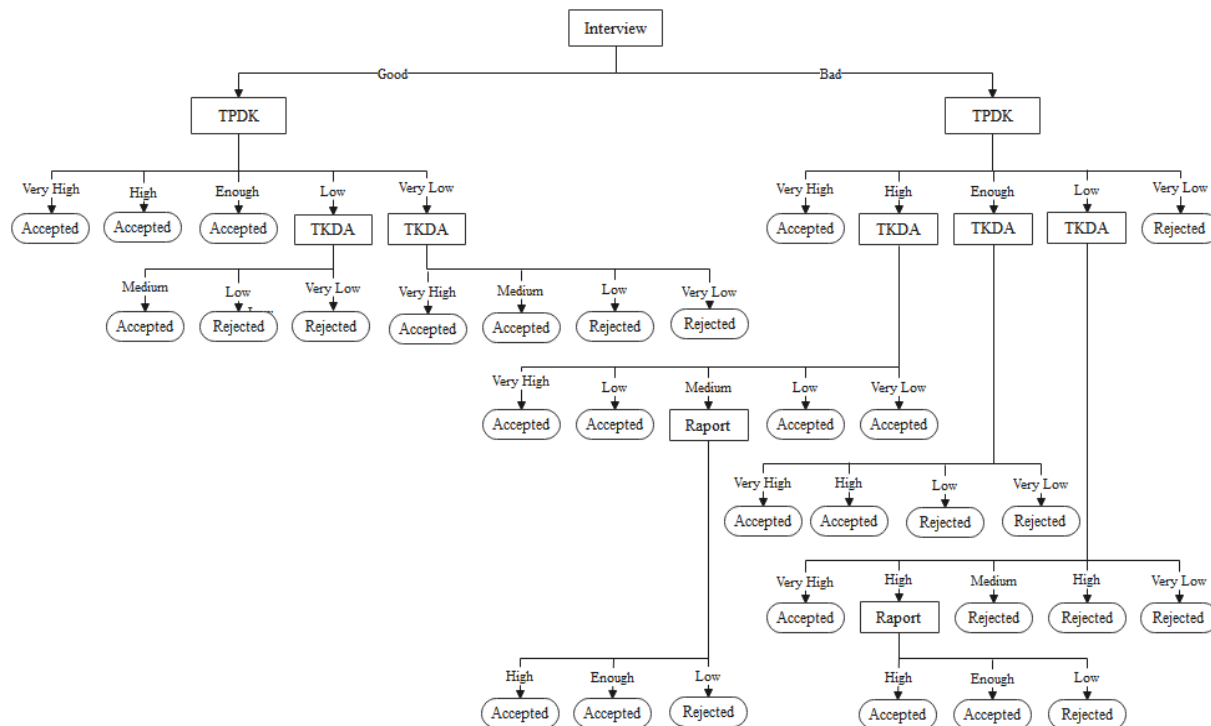


Figure 1. New Student Pattern Mapping Decision Tree

Based on Figure 1, the results of the decision tree calculation in mapping the selection pattern for new student admissions in engineering and computer fields are as follows:

1. If Interview = Good, AND TPKD = Very High THEN Decision = Accepted
2. If Interview = Good, AND TPKD = High THEN Decision = Accepted
3. If Interview = Good, AND TPKD = Enough THEN Decision = Accepted
4. If Interview = Good, AND TPKD = Low, AND TKDA = Medium THEN Decision = Accepted
5. If Interview = Good, AND TPKD = Low, AND TKDA = Low THEN Decision = Rejected
6. If Interview = Good, AND TPKD = Low, AND TKDA = Very Low THEN Decision = Rejected
7. If Interview = Good, AND TPKD = Very Low, AND TKDA = Very High THEN Decision = Accepted
8. If Interview = Good, AND TPKD = Very Low, AND TKDA = Medium THEN Decision = Accepted
9. If Interview = Good, AND TPKD = Very Low, AND TKDA = Low THEN Decision = Rejected
10. If Interview = Good, AND TPKD = Very Low, AND TKDA = Very Low THEN Decision = Rejected
11. If Interview = Bad, AND TPKD = Very High THEN Decision = Accepted
12. If Interview = Bad, AND TPKD = High, AND TKDA = Very High THEN Decision = Accepted
13. If Interview = Bad, AND TPKD = High, AND TKDA = High THEN Decision = Accepted
14. If Interview = Bad, AND TPKD = High, AND TKDA = Medium, AND Raport = High THEN Decision = Accepted
15. If Interview = Bad, AND TPKD = High, AND TKDA = Medium, AND Raport = Enough THEN Decision = Accepted
16. If Interview = Bad, AND TPKD = High, AND TKDA = Medium, AND Raport = Low THEN Decision = Rejected
17. If Interview = Bad, AND TPKD = High, AND TKDA = Low, THEN Decision = Accepted

\*Corresponding author



18. If Interview = Bad, AND TPDK = High, AND TKDA = Very Low, THEN Decision = Accepted
19. If Interview = Bad, AND TPDK = Enough, AND TKDA = Very High, THEN Decision = Accepted
20. If Interview = Bad, AND TPDK = Enough, AND TKDA = High, THEN Decision = Accepted
21. If Interview = Bad, AND TPDK = Enough, AND TKDA = Low, THEN Decision = Rejected
22. If Interview = Bad, AND TPDK = Enough, AND TKDA = Very Low, THEN Decision = Rejected
23. If Interview = Bad, AND TPDK = Low, AND TKDA = High, AND Raport = High THEN Decision = Accepted
24. If Interview = Bad, AND TPDK = Low, AND TKDA = High, AND Raport = Enough THEN Decision = Accepted
25. If Interview = Bad, AND TPDK = Low, AND TKDA = High, AND Raport = Low THEN Decision = Accepted
26. If Interview = Bad, AND TPDK = Low, AND TKDA = High, AND Raport = Enough THEN Decision = Accepted
27. If Interview = Bad, AND TPDK = Low, AND TKDA = Medium, THEN Decision = Rejected
28. If Interview = Bad, AND TPDK = Low, AND TKDA = Low, THEN Decision = Rejected
29. If Interview = Bad, AND TPDK = Low, AND TKDA = Very Low, THEN Decision = Rejected
30. If Interview = Bad, AND TPDK = Very Low, THEN Decision = Rejected

### CONCLUSION

Based on the results of research conducted using data mining methods in determining the nutmeg selection of new student admissions in the field of Engineering and computers, it is concluded that the highest factor affecting 70 new student candidate data is the interview attribute. Then the attribute test of basic computer knowledge becomes the level 1 branch, the attribute of the basic academic ability test becomes the level 2 branch and the attribute of the average value of the report card becomes the last level branch. And in this study resulted in a decision tree of 30 rules

### SUGGESTION

Based on the analysis of data testing in research, as a material for further research development, it can be done by adding a larger amount of data and testing with data mining methods such as the C5.0 algorithm.

### REFERENCES

- Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 13(1), 33–38. <https://doi.org/10.30872/jim.v13i1.629>
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- Elisa, E., & Harman, R. (2019). *ALGORITMA C4.5 UNTUK KLASIFIKASI PEMILIHAN CALON KETUA RW*. 7(2), 71–80.
- Firdaus, D. (2017). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer. *Jurnal Format*, 6(2), 91–97.
- Fitriana, J., Ripanti, E. F., & Tursina, T. (2018). Sistem Pendukung Keputusan Pemilihan Mahasiswa Berprestasi dengan Metode Profile Matching. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 6(4), 157–164. <https://doi.org/10.26418/justin.v6i4.27113>
- Handayani, I. (2019). *Penyakit Disk Hernia Dan Spondylolisthesis Dalam Kolumna Vertebralis*. 1(2), 83–88. <https://doi.org/10.12928/JASIEK.v13i2.xxxx>
- Handoko, K. (2016). Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan). *Jurnal Teknologi Dan Sistem Informasi*, 02(03), 31–40. Retrieved from <http://teknosi.fti.unand.id/index.php/teknosi/article/view/70>
- Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 1(1), 47–56. <https://doi.org/10.24002/ijis.v1i1.1727>
- Merdekawati, A. (2018). Sistem Pendukung Keputusan Penerimaan Beasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Baitul Maal). *Jurnal Pendidikan Teknologi Dan Kejuruan*, 15(1), 113–123.

\*Corresponding author



<https://doi.org/10.23887/jptk-undiksha.v15i1.13067>

- Pratama, A. Y., & Hanum, Y. (2016). *Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 99 Jakarta Untuk Siswa/Siswi SMPN 9 Jakarta Menggunakan Decision Tree*. 21(1), 49–54. <https://doi.org/10.31227/osf.io/vedu7>
- Sikumbang, E. D. (2018). Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Jurnal Teknik Komputer AMIK BSI (JTK)*, Vol 4, No.(September), 1–4.
- Siregar, Y. S., & Harliana, P. (2018a). Algoritma Fuzzy C-Means Pada Aplikasi Matlab Dalam Menentukan Dosen Pembimbing Tugas Akhir. *Seminar Nasional Unisla*, 213–217.
- Siregar, Y. S., & Harliana, P. (2018b). Analisis perancangan algoritma fuzzy c-means dalam menentukan dosen pembimbing tugas akhir. *Jurnal & Penelitian Teknik Informatika*, 3(1), 181–185.
- Susanto, H., & Sudiyatno, S. (2014). Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Pendidikan Vokasi*, 4(2), 222–231. <https://doi.org/10.21831/jpv.v4i2.2547>
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2.

\*Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.