Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of with C4.5 algorithm to measure the level of student satisfaction with student services

Bosker Sinaga[1)*], Jonson Manurung[2)], Nera Mayana Br Tarigan[3)], Siska Feronika Br Sitepu[4)], Nadela Barus[5)]
[1,2,3,4,5)] STMIK Pelita Nusantara, Medan, Indonesia
[1)] boskersinaga@gmail.com, [2)]jhonson.geo@gmail.com, [3)]neramayana658@gmail.com,
[4)]siskaferonika@gmail.com, [5)]nadelabarus@email.com

Abstract: In measuring the level of student satisfaction with student services, it is better to use a method/algorithm to find out how much in certain criteria/services the level of student satisfaction and which services need to be improved. In providing student services, it has been going well, but the difficulty in measuring the level of student satisfaction with student services has not used an algorithm. Data mining as a data mining technique is very important to use in extracting data from measurements that have been carried out so far. Data mining in analyzing data uses several algorithms, one of which is the C4.5 algorithm. The research method is the survey research method, which is a survey research method. This is a research method conducted using surveys or collecting data through research respondents. The purpose of this study is how to apply data mining with the C4.5 algorithm in measuring the level of student satisfaction with student services. This study targets the measurement results of academic criteria/services, guidance and counseling services, interest and talent services, scholarship services and health services. The results of this study were to determine the role of the decision tree on measuring the level of student satisfaction with student services.

Keywords: data mining, algoritma C4.5, satisfaction.

INTRODUCTION

The development of technology in all sectors is very fast, and all fields of business and work must follow the development of technology. Both in helping in terms of systems and terms analysis. Universities in analyzing data should use technology both from the method or algorithm used.

STMIK Pelita Nusantara is one of the universities in North Sumatra that provides student services to students. Student development is the responsibility of the entire academic community. Therefore, the condition of educational interaction between students and their environment consisting of lecturers, education staff, and infrastructure on campus must be realized in a conducive atmosphere full of intimacy. A conducive atmosphere to produce quality graduates requires the integration of the learning process which includes curricular and extra-curricular activities. These curricular and extracurricular activities will lead students to have good intellectual, emotional, and spiritual maturity.

In providing student services, STMIK Pelita Nusantara provides questionnaires to students to get feedback and measure the level of student satisfaction. The problem of the level of student service that runs but in taking measurements does not go well and has an impact on the lack of consistency in student services. In measuring the level of student satisfaction, STMIK Pelita Nusantara processes data only in data processing using an excel application by obtaining the average of each measurement criterion. In analyzing the data, STMIK Pelita Nusantara should use several methods/algorithms to see the comparison of the analysis results.

Data mining or Knowledge Discovery in Database (KDD) is a collection activity, using data to find regularities and a series of processes to explore added value from a set of data with patterns and relationships in large data sets (Aidi Saputra et al., 2020). The application of data mining to measure the level of student satisfaction with student services uses the C4.5 algorithm (Rismayanti, 2017).

C.45 algorithm is an algorithm to build a decision tree (decision making). The C.45 algorithm is one of the decision tree induction algorithms, namely ID3 (Iterative Dichotomiser 3). ID3 was developed by J. Ross Quinlan. In the ID3 algorithm procedure, the inputs are training samples, training labels, and attributes. The C.45 algorithm is the development of ID3 (Muhammad Arif Rahman, 2015) (Sitompul, 2017).

*corresponding author : Bosker Sinaga

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Several other researchers have applied the c4.5 algorithm in measuring the level of satisfaction. Febriarini and Astuti's research entitled Application of the C4.5 Algorithm for Predicting Passenger Satisfaction on the Trans Semarang Bus Rapid Transit (BRT). The purpose of this study was to determine and analyze the satisfaction of BRT Trans Semarang passengers by using data mining techniques. Based on the results of the classification using the C4.5 algorithm, it shows an increase in each tester that has been carried out with a fairly good final accuracy result of 95% which indicates that the C4.5 algorithm is suitable to be used to measure the satisfaction level of BRT Trans Semarang passengers (Febriarini & Astuti, 2019).

Ainnur, et al's research with the title Application of Data Mining Classification of Student Satisfaction Levels with Information System Services in the Information Systems Study Program. The purpose of the study was to determine the effect of Academic Information System services on student satisfaction. Data validity is done with internal validity. The variables used include (1) System Quality, (2) Information Quality, (3) Service Quality, and (4) Performance Expectations. It is hoped that the results of this research can be applied to improve the service of the Academic Information System so that the existing system can run better. (Ainnur et al., 2021)

Based on the background and previous research, the application of data mining with the C4.5 algorithm is very suitable to measure the level of student satisfaction with student services.

LITERATURE REVIEW

Table 1. Literature Review

No.	Researcher Name (year)	Method	Data	Results
1.	(Aidi Saputra et al., 2020)	Algoritma C4.5	100 student data at SMA YPI Swasta Dharma Budi. Questionnaire data are interest in learning, learning methods, learning offerings, learning media	The results of this study are the C4.5 Algorithm. Produce 20 (twenty) rules and the accuracy level generated by this method is 86.67% and the most dominant factor is learning presentation (C3) with a gain value of 0.353960518.
2.	(Febriarini & Astuti, 2019)	Algoritma C4.5	Survey data for 200 Trans Semarang Bus passengers	The results of the classification using the C4.5 algorithm show an increase in each tester that has been carried out with a fairly good final accuracy result of 95% which indicates that the C4.5 algorithm is suitable for measuring the satisfaction level of Trans Semarang BRT passengers.
3.	(Ramadhan et al., 2020)	Algoritma C4.5	Aspects of Reliability (Reliability), Responsiveness (Responsiveness), Assurance (Insurance), Empathy (Empathy), Tangibles (Real Form)	Based on data processing using RapidMiner software, the accuracy value is 96.50%. In accordance with these provisions, the results of manual calculations with RapidMiner testing will get the same results.
4.	(Fadillah et al., 2020)	Algoritma C4.5	The results of the questionnaire given to students at random in the STIKOM Tunas Bangsa Pematangsiantar environment	The result of the research is that data mining with the C4.5 algorithm can be applied to classify the level of student satisfaction on the performance of computer laboratory assistants. The effect of computer laboratory assistant performance services at STIKOM Tunas Bangsa on student satisfaction is the reliability of a computer laboratory assistant.
5.	(Ainnur et al., 2021)	Algoritma C4.5	The results of the questionnaire given to students randomly	The level of satisfaction of STIKOM Tunas Bangsa students can be clearly measured. After analyzing, from several aspects that became the benchmark for assessing student satisfaction, it can be seen that the most dominant aspect of satisfaction from several aspects of the quality of the information system provided by the STIKOM Tunas Bangsa campus.
6.	(Hendri et al., n.d.)	Algoritma C4.5	Data were obtained from distributing questionnaires to 35 visitors	From the research above, it can be concluded that 83% of visitors are satisfied with the wildlife park facilities.
7.	(Azwanti, 2018)	Algoritma C4.5	Observation and interview data for AMIK Labuhan Batu students	The C4.5 algorithm is considered an algorithm that is very helpful in classifying data because the characteristics of the classified data can be obtained clearly, both in the form of a decision tree structure and if-then rules, making it easier for users to extract information on the relevant data.
8.	(Novianti et al., 2016)	Algoritma C4.5	TPA Results, Junior High School Report Cards and	The results of the study, obtained the results of the classification of students' majors that have been tested

*corresponding author : Bosker Sinaga

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

No.	Researcher Name (year)	Method	Data	Results
			Student Interests.	according to the level of accuracy of 89.74%.
9.	(Muhammad Arif Rahman, 2015)	Algoritma C4.5	40 prospective scholarship recipients at the Postgraduate Program of IAIN Raden Intan Lampung	The results are based on the calculation of the C.45 Algorithm as many as 18 (eighteen) students who are not eligible to become scholarship recipients because they have a GPA <3.00, then 8 (eight) students who are not eligible to become scholarship recipients because they have <5 years of service and work Non-PNS, so as many as 14 (fourteen) students are eligible to become scholarship recipients because they have met the criteria for scholarship recipients in terms of GPA, occupation and predetermined period of service.
10.	(Rufiyanto et al., 2021)	Algoritma C4.5	Pandanaran University student questionnaire data	Produce models, rules and applications for predicting student satisfaction with an accuracy value of 87.95% and an AUC value of 0.995 so that it includes very good data classification.

METHOD

The research method is important for a researcher to achieve a goal, and can find answers to the problems posed. The research stages start from identifying the problem to publishing scientific articles, as shown in the following fishbone diagram:

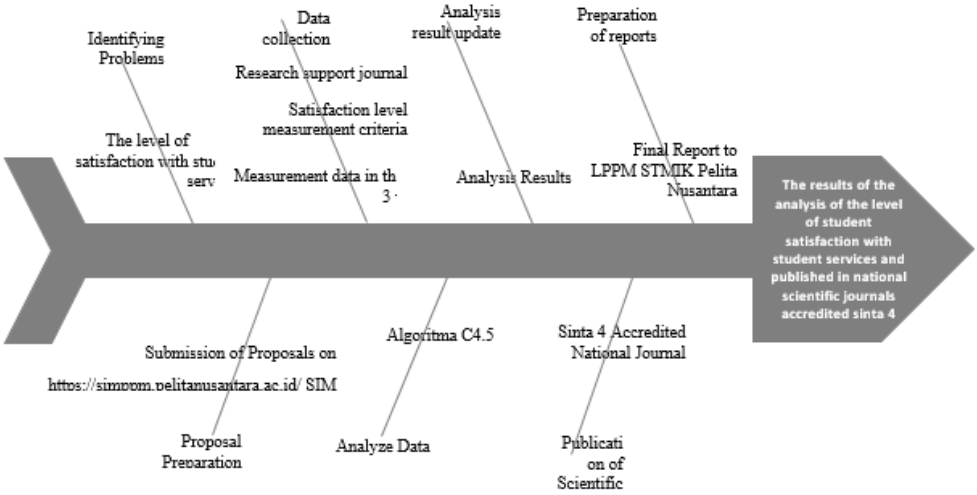

Figure 1. Research Stages

Table 2. Example Table

Parameter	Value
Maximum Repetition	10, 50, 100
Learning Ratio	0.1 , 0.5, 1
Minimum Error	0.1, 0.01, 0.001

Both the figure and the table in the order of writing start from 1, not by chapter. While writing formulas or equations can follow the following rules: (Arifin & Fitrianah, 2018)

$$C_j = \sqrt{\sum_{i=1}^{n}(x_i - w_{ij})^2} \quad \text{.....(1)}$$

RESULT

There are several processes that will be carried out by the author in conducting data mining, namely:

4.1 Data Selection

Selecting the data set that will be used in this study is the questionnaire data for STMIK Pelita Nusantara students. For the questionnaire data, 102 data were successfully collected.

.2 Cleaning

*corresponding author : Bosker Sinaga

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Cleaning aims to clean the questionnaire data for STMIK Pelita Nusantara students by deleting data that is not used or that does not have a complete attribute value (missing value). The questionnaire data which initially contained 102 data became 78 data that were ready to be processed.

4.3 Data Transformation

At this stage the questionnaire data containing the information very satisfied was changed to a value of 3, quite satisfied was changed to 2 and not satisfied was changed to 1. For the questionnaire data made by the author consisted of 10 questions, 2 questions represented Tangibles, reliability, Responsiveness, Assurance and Empathy. The data is converted into numbers and the average is sought in each statement.

4.3 C4.5. Algorithm Method Analysis

Table 3. 19 Rule data representing 78 data Cleaning

No	*Name*	*Tangibles*	*Reliabiility*	*Responsiveness*	*Assurance*	*Empathy*	Results
1	Deli Putra Damai Lase	very satisfied	very satisfied	very satisfied	very satisfied	very satisfied	satisfied
2	Wahid Zailani	very satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	satisfied
3	Putri Zulaika	very satisfied	quite satisfied	very satisfied	very satisfied	quite satisfied	satisfied
4	Yenny Sembiring	Quite satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	satisfied
5	Finna Handayani	very satisfied	quite satisfied	very satisfied	quite satisfied	quite satisfied	satisfied
6	Tri Hotmi Dear Nainggolan	very satisfied	quite satisfied	quite satisfied	quite satisfied	Not satisfied	Not satisfied
7	Muhammad Zaini	Quite satisfied	quite satisfied	very satisfied	quite satisfied	quite satisfied	satisfied
8	Mutia Sani	Quite satisfied	quite satisfied	very satisfied	very satisfied	very satisfied	satisfied
9	Sintia Anida	Quite satisfied	quite satisfied	quite satisfied	quite satisfied	quite satisfied	Not satisfied
10	Siska Adelina Br Sembiring	Quite satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	Not satisfied
11	Eka Pratiwi	very satisfied	quite satisfied	quite satisfied	very satisfied	very satisfied	satisfied
12	Putri Dimas Sari	Quite satisfied	quite satisfied	quite satisfied	quite satisfied	Not satisfied	Not satisfied
13	Efisutriani Sadawa	very satisfied	quite satisfied	quite satisfied	quite satisfied	very satisfied	satisfied
14	Mhd. Hafas Hanafia Maharaja	very satisfied	quite satisfied	very satisfied	very satisfied	very satisfied	satisfied
15	M. Rizky Syahputra S	very satisfied	quite satisfied	quite satisfied	very satisfied	quite satisfied	Not satisfied
16	M. Fathan Aqmar Nst	very satisfied	quite satisfied	quite satisfied	quite satisfied	quite satisfied	Not satisfied
17	Ayu Ningtiyas	very satisfied	very satisfied	quite satisfied	quite satisfied	very satisfied	satisfied
18	Putri Mutia Hakim	very satisfied	very satisfied	very satisfied	quite satisfied	very satisfied	satisfied
19	Perdi Gulo	very satisfied	very satisfied	quite satisfied	very satisfied	very satisfied	satisfied

Calculation manually algorithm C4.5

1. Calculate the Entropy and GAIN values for each criterion node 1

a. *Entrophy Tangible*

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-6/19 * $\log_2$ (6/19)) + - 13/19 * $\log_2$ (13/19)

= **0,899743759**

Entropy (*Tangible* Very satisfied)

= (-3/13 * $\log_2$ (3/13)) + - 10/13 * $\log_2$ (10/13)

= 0,779349837

Entropy (*Tangible* Quite satisfied)

= (-3/6 * log2 (3/6)) + - 3/6 * $\log_2$ (3/6)

= 1

Entropy (*Tangible* Not satisfied)

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)

= 0

Gain (Total, *Tangibles* Node 1)

$$GAIN\ (S, A) = Entropy\ (S) - \Sigma^{n}_{i=1} - \frac{|S_i|}{|S|} * Entropy(S_i)$$

= **0,899743759***((13/19*0,779349837) + (6/19*1) + (0/19*0)

= 0,050715

b. *Entropy Responsiveness* Node 1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-6/19 * $\log_2$ (6/19)) + - 13/19 * $\log_2$ (13/19)

= **0,899743759**

Entropy (*Responsiveness* Very satisfied)

= (-0/7 * $\log_2$ (0/7)) + - 7/7 * $\log_2$ (7/7)

= 0

Entropy (*Responsiveness* Quite satisfied)

= (-6/12 * log2 (6/12)) + - 6/12 * $\log_2$ (6/12)

= 1

Entropy (*Responsiveness* Not satisfied)

*corresponding author : Bosker Sinaga

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)
= 0

GAIN (S, A) = *Entropy* (S) - $\Sigma^{n}_{i=1}$ - $\frac{|S_i|}{|S|}$ * *Entropy*(S_i)

= **0,899743759**– ((7/19*0) + (12/19*1)+(0/19*0))
= 0,268165

c. *Entropy Reliablilty Node* 1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)
Entropy (*S*) = (-6/19 * $\log_2$ (6/19)) + - 13/19 * $\log_2$ (13/19)
= **0,899743759**
Entropy (*Reliability* Very satisfied)
= (-1/7 * $\log_2$ (1/7)) + - 6/7 * $\log_2$ (6/7)
= 0,591672779
Entropy (*Reliability* Quite satisfied)
= (-5/12 * $\log_2$ (5/12)) + - 7/12 * $\log_2$ (7/12)
= 0,979868757
Entropy (*Responsiveness* Not satisfied)
= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)
= 0
GAIN (S, A) = *Entropy* (S) - $\Sigma^{n}_{i=1}$ - $\frac{|S_i|}{|S|}$ * *Entropy*(S_i)
= **0,899743759**– ((7/19*0,591672779) + (12/19*0,979868757) + (0/19*0))
= 0,062895

d. Calculating the Gain value (Total, *Assurance* Node 1)

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)
Entropy (*S*) = (-6/19 * $\log_2$ (6/19)) + - 13/19 * $\log_2$ (13/19)
= **0,899743759**
Entropy (*Assurance* Very satisfied)
= (-1/7 * $\log_2$ (1/7)) + - 6/7 * $\log_2$ (6/7)
= 0,591672779
Entropy (*Assurance* Quite satisfied)
= (-5/12 * $\log_2$ (5/12)) + - 7/12 * $\log_2$ (7/12)
= 0,979868757
Entropy (*Assurance* Not satisfied)
= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)
= 0
GAIN (S, A) = *Entropy* (S) - $\Sigma^{n}_{i=1}$ - $\frac{|S_i|}{|S|}$ * *Entropy*(S_i)
= **0,899743759**– ((7/19*0,591672779) + (12/19*0,918295834) + (0/19*0))
= 0,062895

e. *Calculating the Gain value* (Total, *Empathy* Node 1)

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)
Entropy (*S*) = (-6/19 * $\log_2$ (6/19)) + - 13/19 * $\log_2$ (13/19)
= **0,899743759**
Entropy (*Empathy* Very satisfied)
= (-0/8 * $\log_2$ (0/8)) + - 8/8 * $\log_2$ (8/8)
= 0
Entropy (Empathy Quite satisfied)
= (-4/9 * $\log_2$ (4/9)) + - 5/9 * $\log_2$ (5/9)
= 0,99107606
Entropy (*Empathy* Not satisfied)
= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)
= 0
GAIN (S, A) = *Entropy* (S) - $\Sigma^{n}_{i=1}$ - $\frac{|S_i|}{|S|}$ * *Entropy* (S_i)
= **0,899743759**– ((8/19*0) + (10/19*1) + (1/19*0))
= 0,373428

Table 4. Calculation of Node 1.1

	Node	Number of Cases	not satisfied	Satisfied	Entropi	Gain
1,1	Total	9	5	4	**0,9911**	
	Tangibles					0,0072

*corresponding author : Bosker Sinaga

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Node	Number of Cases	not satisfied	Satisfied	Entropi	Gain
-Very satisfied	5	3	2	0,9710	
-Quite satisfied	4	2	2	1	
-Not satisfied	0	0	0	0	
Reliabiility	9	5	4		0,0183
-Very satisfied	3	2	1	0,9183	
-Quite satisfied	6	3	3	1	
-Not satisfied	0	0	0	0	
Respon siveness	9	5	4		0,3789
-Very satisfied	3	3	0	0	
-Quite satisfied	6	2	4	0,9183	
-Not satisfied	0	0	0	0	
Assurance	9	5	4		0,0026
-Very satisfied	2	1	1	**1**	
-Quite satisfied	7	4	3	**0,9852**	
-Not satisfied	0	0	0	**0**	

f. Pohon Keputusan dengan nilai GAIN *Responsiveness* yang tertinggi

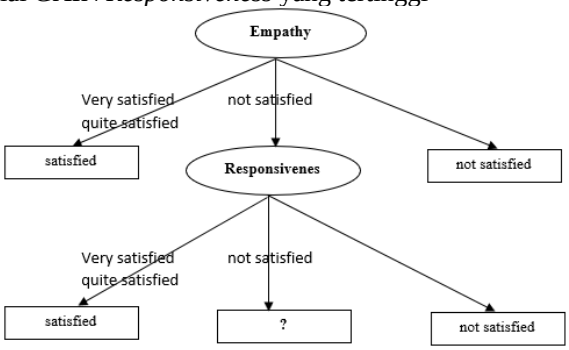

Figure 2. Node Decision Tree 1.1

For the next highest GAIN, Responsiveness is 0.378878837. Then the Responsiveness becomes node 1.1.1 which has three values, namely "Satisfied", “Quite satisfied” and “Not satisfied”.

For Very satisfied and not satisfied only has one value, namely satisfied and not satisfied. As for the quite satisfied rule, there are still two decisions "satisfied" and "not satisfied" for that to be calculated again.

Table 5. Questionnaire data with the highest GAIN Responsiveness

Name	*Tangibles*	*Reliabiility*	*Responsiveness*	*Assurance*	*Empathy*	*Results*
Wahid Zailani	very satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	satisfied
Yenny Sembiring	quite satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	satisfied
Sintia Anida	quite satisfied	quite satisfied	quite satisfied	quite satisfied	quite satisfied	not satisfied
Siska Adelina Br Sembiring	quite satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	not satisfied
M. Rizky Syahputra S	very satisfied	quite satisfied	quite satisfied	very satisfied	quite satisfied	not satisfied
Putri Mutia Hakim	very satisfied	quite satisfied	quite satisfied	quite satisfied	quite satisfied	not satisfied

2. Calculate the Entropy and GAIN values for each node criterion 1.1.1
   a. *Entropy Tangibles* Node 1.1.1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-2/6 * $\log_2$ (2/6)) + - 4/6 * $\log_2$ (4/6)

= 0,918295834

Entropy (*Tangibles* Very satisfied)

= (-1/3 * $\log_2$ (1/3)) + - 2/3 * $\log_2$ (2/3)

= 0,918295834

Entropy (*Tangibles* Quite satisfied)

= (-1/3 * $\log_2$ (1/3)) + - 2/3 * $\log_2$ (2/3))

= 0,918295834

Entropy (*Tangibles* Not satisfied)

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)

= 0

$$GAIN\ (S, A) = Entropy\ (S) - \Sigma^{n}_{i=1} - \frac{|S_i|}{|S|} * Entropy(S_i)$$

= 0,918295834 – ((3/6*0,918295834)+(3/6*0,918295834)+(0/6*0))

*corresponding author : Bosker Sinaga

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

= 0

b. *Entropy Reliability* Node 1.1.1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-2/6 * $\log_2$ (2/6)) + - 4/6 * $\log_2$ (4/6)

= 0,918295834

Entropy (*Reliability* Very satisfied)

= (-2/3 * $\log_2$ (2/3)) + - 1/3 * $\log_2$ (1/3)

= 0,918295834

Entropy (*Reliability* Quite satisfied)

= (-0/3 * $\log_2$ (0/3)) + - 3/3 * $\log_2$ (3/3))

= 0

Entropy (*Reliability* Not satisfied)

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)

= 0

$$GAIN\ (S, A) = Entropy\ (S) - \Sigma^{n}_{i=1} - \frac{|S_i|}{|S|} * Entropy(S_i)$$

= 0,918295834 – ((3/6*0,918295834)+(3/6*0)+(0/6*0))

= 0,4591479

c. *Entropy Assurance* Node 1.1.1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-2/6 * $\log_2$ (2/6)) + - 4/6 * $\log_2$ (4/6)

= 0,918295834

Entropy (*Assurance* Very satisfied)

= (-0/1 * $\log_2$ (0/1)) + - 1/1 * $\log_2$ (1/1)

= 0

Entropy (*Assurance* Quite satisfied)

= (-2/5 * $\log_2$ (2/5)) + - 3/5 * $\log_2$ (3/5))

= 0,970950594

Entropy (*Assurance* Not satisfied)

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)

= 0

$$GAIN\ (S, A) = Entropy\ (S) - \Sigma^{n}_{i=1} - \frac{|S_i|}{|S|} * Entropy(S_i)$$

= 0,918295834 – ((1/6*0)+(5/6*0,970950594)+(0/6*0))

= 0,1091703

Table 6. Node Calculation 1.1.1

	Node	Number of Cases	not satisfied	Satisfied	Entropi	Gain
1.1.1	Total	6	2	4	0,9183	
	Tangi bles					0
	Very satisfied	3	1	2	0,9183	
	Quite satisfied	3	1	2	0,9183	
	Not satisfied	0	0	0	0	
	Relia biility	6	2	4		0,4591
	Very satisfied	3	2	1	0,9183	
	Quite satisfied	3	0	3	0	
	Not satisfied	0	0	0	0	
	Assu rance	6	2	4		0,1092
	Very satisfied	1	0	1	0	
	Quite satisfied	5	2	3	0,9710	
	Not satisfied	0	0	0	0	

d. Decision Tree with GAIN highest responsiveness Value

*corresponding author : Bosker Sinaga

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

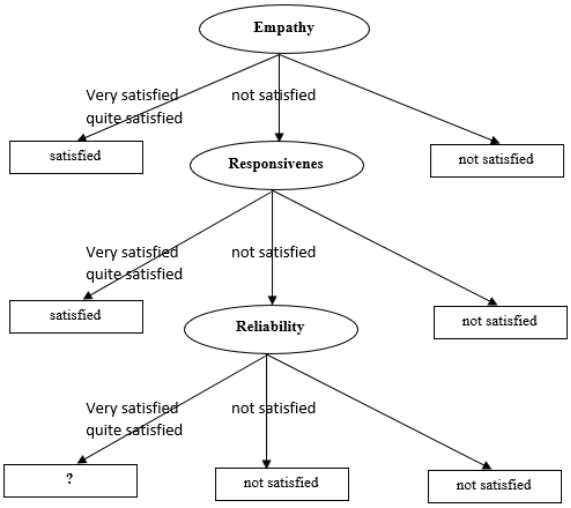

Figure 3. Node Decision Tree 1.1.1

For the next highest GAIN, Reliability is 0.4591479. Then Responsiveness becomes node 1.1.1.1 which has three values, namely "Very satisfied", "Quite satisfied" and "Not satisfied". For Quite satisfied and not satisfied only has one value, namely satisfied and not satisfied. As for the quite satisfied rule, there are still two decisions "satisfied" and "not satisfied" for that to be calculated again.

Table 7. Questionnaire data with the highest GAIN Responsiveness

Nama	*Tangibles*	*Reliabiility*	*Responsiveness*	*Assurance*	*Empathy*	Hasil
Wahid Zailani	very satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	satisfied
Yenny Sembiring	Quite satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	satisfied
Siska Adelina Br Sembiring	Quite satisfied	very satisfied	quite satisfied	quite satisfied	quite satisfied	Not satisfied

3. Calculating the Entropy and GAIN values for each node criterion 1.1.1

a. *Entropy Tangibles* Node 1.1.1.1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-2/3 * $\log_2$ (2/3)) + - 1/3 * $\log_2$ (1/3)

= 0,918296

Entropy (*Tangibles* Very satisfied)

= (-1/1 * $\log_2$ (1/1)) + - 0/1 * $\log_2$ (0/1)

= 0

Entropy (*Tangibles* Quite satisfied)

= (-1/2 * $\log_2$ (1/2)) + - 1/2 * $\log_2$ (1/2))

= 1

Entropy (*Tangibles* Not satisfied)

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)

= 0

GAIN (S, A) = *Entropy* (S) - $\Sigma^{n}_{i=1}$ - $\frac{|S_i|}{|S|}$ * *Entropy*(S_i)

= 0,918296 – ((1/3*0)+(2/3*1)+(0/3*0))

= 0,251629

b. *Entropy Assurance* Node 1.1.1.1

Entropy (S) = $\Sigma^{n}_{i=0}$ - pi * $\log_2$ (pi)

Entropy (*S*) = (-2/3 * $\log_2$ (2/3)) + - 1/3 * $\log_2$ (1/3)

= 0,918296

Entropy (*Assurance* Very satisfied)

= (-0/0 * $\log_2$ (0/0)) + - 0/0 * $\log_2$ (0/0)

= 0

Entropy (*Assurance* Quite satisfied)

= (-2/3 * $\log_2$ (2/3)) + - 1/3 * $\log_2$ (1/3)

= 0,918296

*corresponding author : Bosker Sinaga

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11774
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Entropy (*Assurance* Not satisfied)

$= (-0/0 * \log_2 (0/0)) + - 0/0 * \log_2 (0/0)$

$= 0$

$GAIN\ (S, A) = Entropy\ (S) - \Sigma^{n}_{i=1} - \frac{|S_i|}{|S|} * Entropy(S_i)$

$= 0{,}918296 – ((0/3*0)+(3/3*0{,}918296)+(0/3*0))$

$= 0$

Table 8. Node Calculation 1.1.1.1

Node		Number of Cases	not satisfied	Satisfied	Entropi	Gain
1.1.1.1	Total	3	2	1	0,9183	
	Tangibles					0,252
	Very satisfied	1	1	0	0	
	Quite satisfied	2	1	1	1	
	Not satisfied	0	0	0	0	
	Assurance	3	2	1		0
	Very satisfied	0	0	0	0	
	Quite satisfied	3	2	1	0,918	
	Not satisfied	0	0	0	0	

a. Final Decision Tree

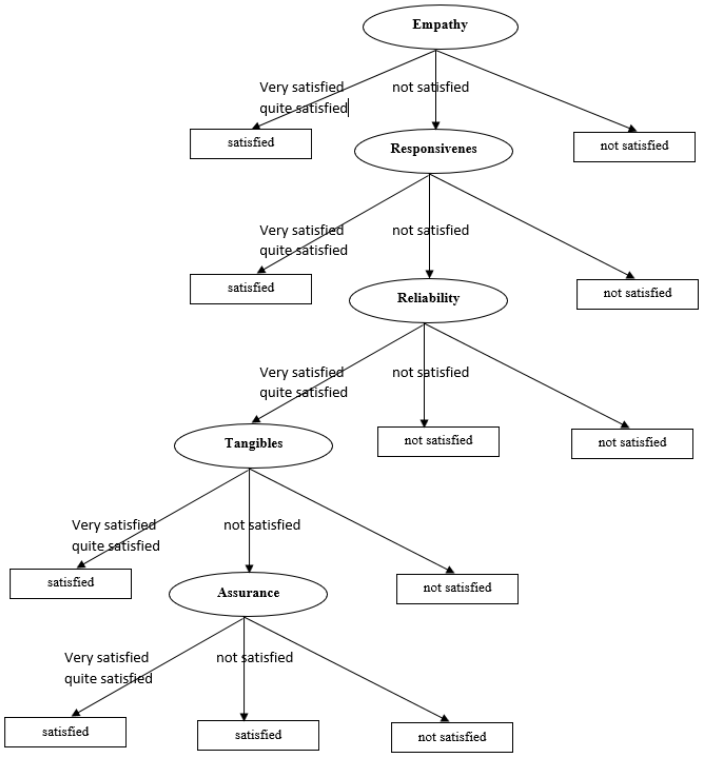

Figure 4. Node Decision Tree 1.1.1.1

CONCLUSION

The conclusions in this study are:

To apply data mining with the C4.5 algorithm to the level of student satisfaction and student services, namely choosing an attribute as the root, based on the highest gain value of the existing attributes. Create a branch for each value, meaning to make a branch according to the number of values of the highest gain variable. Divide each case into branches, based on the calculation of the highest gain value and the calculation is carried out after calculating the initial highest gain value and then the process of calculating the highest gain is carried out again without including the initial gain variable value. Repeating the process in each branch so that all cases in the

*corresponding author : Bosker Sinaga

e-ISSN : 2541-2019
p-ISSN : 2541-044X

branch have the same class, repeating all the highest gain calculation processes for each case branch until the calculation process can no longer be carried out.

The results of data analysis on the level of student satisfaction and student services with the application of data mining with the C4.5 algorithm, namely the level of satisfaction.

REFERENCES

Aidi Saputra, K., Tata Hardinata, J., Ridwan Lubis, M., Retno Andani, S., & Syahputra Saragih, I. (2020). KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online. *Media Online)*, *1*(3), 113–118. https://djournals.com/klik

Ainnur, D., Saragih, R., Safii, M., & Suhendro, D. (2021). Penerapan Data Mining Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi di Program Studi Sistem Informasi. In *Journal of Information System Research* (Vol. 2, Nomor 2). https://ejurnal.seminar-id.com/index.php/josh/

Arifin, M. F., & Fitrianah, D. (2018). Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada. *InComTech*, *8*(2), 87–102. https://doi.org/10.22441/incomtech.v8i1.2198

Azwanti, N. (2018). ALGORITMA C4.5 UNTUK MEMPREDIKSI MAHASISWA YANG MENGULANG MATA KULIAH (STUDI KASUS DI AMIK LABUHAN BATU). *Jurnal SIMETRIS*, *9*(1).

Fadillah, W. R., Hartama, D., Damanik, I. S., Safii, M., Suhendro, D., Tunas, S., Pematangsiantar, B., & Tunas, A. (2020). Implementasi Data Mining C4.5 Dalam Mengukur Tingkat Kepuasan Mahasiswa Terhadap Kinerja Asisten Laboratorium Komputer. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, *2*, 403–414.

Fasilitas, T., Margasatwa, T., Hendri, J., & Oscar, D. (n.d.). Penerapan Algoritma C4.5 Dalam Mengukur Kepuasan Pengunjung. In *Jurnal* (Vol. 3, Nomor 1). http://ejournal.bsi.ac.id/ejurnal/index.php/infortech73

Febriarini, A. S., & Astuti, E. Z. (2019). Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang. *Eksplora Informatika*, *8*(2), 95–103. https://doi.org/10.30864/eksplora.v8i2.156

Jurusan,], Komputer, S., Mipa, F., Tanjungpura, U., Prof, J., & Nawawi, H. H. (2016). IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK PENJURUSAN SISWA (STUDI KASUS: SMA NEGERI 1 PONTIANAK) [1] Beti Novianti, [2] Tedy Rismawan [3] Syamsul Bahri. In *Jurnal Coding, Sistem Komputer Untan* (Vol. 04, Nomor 3).

Muhammad Arif Rahman. (2015). Algoritma C45 Untuk Menentukan Mahasiswa Penerima Beasiswa (Studi Kasus : Pps Iain Raden Intan Bandar Lampung). *Jurnal Teknologi Informasi Magister Darmajaya*, *1*(02), 118–128.

Ramadhan, G., Perdana Windarto, A., Irawan, E., Saputra, W., Okprana, H., Infomasi, S., Tunas, S., Pematangsiantar, B., & Tunas, A. (2020). Penerapan Data Mining Menggunakan Algoritma C4.5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, *2*, 376–385.

Rismayanti, R. (2017). Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan. *Jurnal Media Infotama*, *12*(2), 116–120. https://doi.org/10.37676/jmi.v12i2.413

Rufiyanto, A., Rochcham, M., & Rohman, A. (2021). Prediksi Kepuasan Mahasiswa dengan Menggunakan Algoritma C4.5 terhadap Pelayanan Akademik. *TRANSFORMTIKA*, *18*(2), 210–214.

Sitompul, J. 17110443. (2017). Analisa Sentimen Masyarakat Terhadap Calon Legislatif Partai Politik Dengan Menerapkan Algoritma Text Mining. *Jurnal Teknik Mesin (JTM)*, *06*(2), 69–73.

*corresponding author : Bosker Sinaga