http://ejournal.seaninstitute.or.id/index.php/InfoSains
Jurnal Info Sains : Informatikan dan Sains, Volume 13, No 02, 2023
E-ISSN.2797-7889, P-ISSN.2089-3329

Application of the Classification Decision Tree Method to Determine Student Satisfaction Factors for Student Services

[1]Yuda Perwira, [2]Amran Sitohang, [3]Mutiara Pandjaitan, [4]Kuza Simamora
[1,2,3,4] Informatics Engineering, STMIK Pelita Nusantara
Email: yudaperwira25@gmail.com[1], amranryan89@gmail.com[2], mutiarapnjt20@gmail.com[3], simamora23@gmail.com[4]

Keywords

Data Mining
Clasfification Decision tree

Abstract. This study aims to apply the Classification Decision Tree method in knowing the factors that influence student satisfaction with student services in tertiary institutions. The Classification Decision Tree method is used to build a decision tree model that can identify the factors that most influence student satisfaction. The data used in this study is survey data on student satisfaction with student services in tertiary institutions, which consists of several variables such as service quality, facilities, information availability, and others. The data will be processed using the Classification Decision Tree algorithm to build a decision tree model that can predict student satisfaction based on the factors that influence it. The results of this study obtained an important root or root of student satisfaction with student services. The first is student welfare services and the second is organizational development services and the results of the test data show an accuracy of 87%.

1. INTRODUCTION

Good and satisfying student services are an important factor in increasing student satisfaction and the prestige of tertiary institutions. Therefore, many tertiary institutions are trying to improve the quality of student services by collecting student satisfaction survey data for the student services provided. However, in processing student satisfaction survey data, it is often difficult to identify which factors have the most influence on student satisfaction. In addition, it is also difficult to determine priorities in making improvements to aspects that are considered unsatisfactory by students. The Classification and Regression Tree (CART) method is a data mining method that can be used to overcome these problems. This method can build a decision tree model that can identify the factors that have the most influence on student satisfaction based on the survey data that has been collected. In the context of this study, the use of the CART method is expected to assist tertiary institutions in improving the quality of student services by identifying the factors that most influence student satisfaction and providing recommendations for improvements accordingly. Thus, universities can increase student satisfaction and the prestige of universities as a whole.

2. METHOD

Decision Tree

The structure of a flowchart is like a tree, where each internal node represents an attribute test, each branch represents a test result, and leaf nodes represent a class or class distribution. The decision tree flow follows from the root node to the leaf node which contains predictions [1].

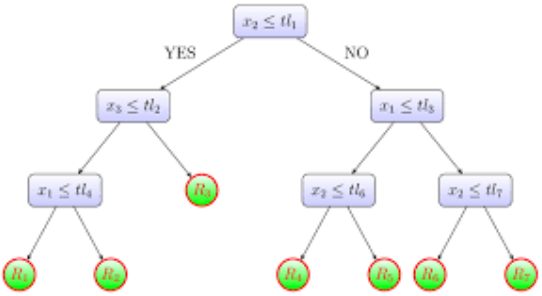

Figure 1. Decision Tree

Decision tree:

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

1. Choosing a Root Node is based on the highest Gini Index and Information Gain.
2. At each iteration this algorithm calculates the Gini.
3. Index and Information Gain consider which each node has never used before.
4. Choose a root node based on the lowest Gini Index or highest I.G.
5. Then separate the set S to produce a subset of data.
6. The algorithm continuously iterates over each subset and ensures that the attributes are fresh and makes up the Decision Tree.

Information Gain with Entrophy Values Entrophy is a formula for calculating attribute homogeneity (A) from a sample data (S). With formulas:

$$\boldsymbol{Enthropy(S)} = \sum pi\ log2pi\ n\ i{=}1 \quad (1)$$

S = Set of cases in sataset
A = Feature (attribute)
N = number of partitions attribute S
Pi = the proportion of Si to S then:

$$\boldsymbol{Enthropy(S)} = \sum p(i|s) log2p(i|s)\ n\ i{=}1$$

Then Gain(S,A) is the Information Gain of attribute A in the sample collection S.

$$\boldsymbol{Gain(S,A)}{=}Entropy(S) \sum |Si|\ |S|\ n\ i{=}1\ Entrop\ y(Si) \quad (2)$$

3. RESULTS AND DISCUSSION

Data

The data used in this study is data from the results of questionnaires distributed to students to determine satisfaction with Student Services. The following is a list of questions from the questionnaire

Table 1 List of Questions

No	Question	Answer			
1	Welfare Service	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)
2	Organization development	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)
3	Interests and Talents	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)
4	Softskill, Reasoning and Scientific Services	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)
5	Development of student creativity programs	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)
6	Guidance Counseling Services	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)
7	Satisfaction with Student Affairs Services	Dissatisfied (1)	Dissatisfied (2)	Not (3)	Very Satisfied (3)

Decision tree classification analysis

And the following is the data from the questionnaire results that have gone through the Selection, Repoceccing/Cleaning stage.

Table 2 Questionnaire Data Processing Results

Welfare Service	Organization development	Interests and Talents	Softskill, Reasoning and Scientific Services	Development of student creativity programs	Guidance Counseling Services	The thirdness of students towards student services
3	3	3	What?	3	3	3
3	2	3	1	3	1	1
4	4	4	4	4	4	3

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

1	3	3	2	3	2	1
3	3	2	1	3	3	3
4	3	3	1	3	3	3
1	3	3	3	3	3	1
1	2	3	2	3	3	1
3	2	3	2	3	3	1
2	1	3	2	3	2	1
3	2	1	3	3	3	3
4	3	1	3	3	3	3
3	3	3	2	3	4	3
4	2	3	3	4	3	3

Furthermore, Data Transformation to Find Entropy and Gain Values

The Decision Tree Algorithm formula is divided into 2 formulas. To calculate the gain, use the formula as shown in equation 1 below:

Gaint (S, A) = Entrophy (S) - $\sum_{i=1}^{n} \frac{|S1|}{|S|} * Entropy(Si) \sum_{i=1}^{n} \frac{|S1|}{|S|} * Entropy(Si)$

Information:

S : Set of cases

A : Attribute

n : Number of partitions attribute A

|Si| : The number of cases on the i-th partition

|S| : Number of cases in S

Meanwhile, the entropy value calculation can be seen in equation 2 below

Enrophy (s) = $\sum_{i=1}^{n} -pi * log2pi \sum_{i=1}^{n} -pi * log2pi$

1. Entrophy (Total) = $\left(-\frac{8}{14} * Log2\left(\frac{8}{14}\right) + -\frac{6}{14} * Log2\left(\frac{6}{14}\right)\right)$
   $\left(-\frac{8}{14} * Log2\left(\frac{8}{14}\right) + -\frac{6}{14} * Log2\left(\frac{6}{14}\right)\right)$

   Entrophy Total = 0,98

And so on to look for entropy from Welfare Services (Highly Dissatisfied, Dissatisfied, Dissatisfied, Satisfied, Very Satisfied) to Interaction Technical Assistance Services (Highly Dissatisfied, Dissatisfied, Dissatisfied, Satisfied, Very Satisfied)

2. Gain (Interaction Facility) = $0.98 - \left(\left(\frac{0}{14} * 0\right) + \left(\frac{2}{14} * 0\right) + \left(\frac{2}{14} * 0\right) + \left(\frac{6}{14} * 0.91\right) + \left(\frac{4}{14} * 0.5\right)\right)$
   $0.98 - \left(\left(\frac{0}{14} * 0\right) + \left(\frac{2}{14} * 0\right) + \left(\frac{2}{14} * 0\right) + \left(\frac{6}{14} * 0.91\right) + \left(\frac{4}{14} * 0.5\right)\right)$

   Gain (Interaction facility) = 0.44

And so on to look for Gain (Learning Object Facilities), Gain (Material According to SRP), Gain (Course Achievement Evaluation), Gain (Ease of Access) and gain (Technical Assistance Services)

Furthermore, complete entropy and gain data can be seen in the table below

Table 3 Data Transformation to Find ENtropy and Gain Values

	NUMBER OF CASES	DESTR OYED	NOT SATISFIED	ENTROPY	GAIN
TOTAL	14	8	6	0,985228136	
Welfare Services					**0,448816**
1	2	0	2	0	
2	2	0	2	0	
3	6	4	2	0,918295834	
4	4	4	2	0,5	

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

Organization Development					**0,2069**
1	1	0	1	0	
2	5	2	3	0,970950594	
3	7	5	2	0,863120569	
4	1	1	0	0	
Interests and Talents					**0,291692**
1	0	0	0	0	
2	3	3	0	0	
3	10	4	6	0,970950594	
4	1	1	0	0	
Scientific Reasoning Service					**0,3322**
1	1	0	1	0	
2	8	4	4	1	
3	4	3	1	0,811278124	
4	1	1	0	0	
Student Creativity Program Development					**-0,85714**
1	0	0	0	0	
2	0	0	0	0	
3	12	6	6	1	
4	2	2	0	0	
Guidance Counseling Services					**-0,59033**
1	0	0	0	0	
2	3	0	3	0	
3	9	6	3	0,918295834	
4	2	2	0	0	
Soft Skill Service					
1	0	0	0	0	
2	0	0	0	0	
3	12	6	6	1	
4	2	2	0	0	

Then look for the Max Gain Value, the highest Gain will be chosen to find the root (root) in the above calculation the highest gain is the Interaction Facility.

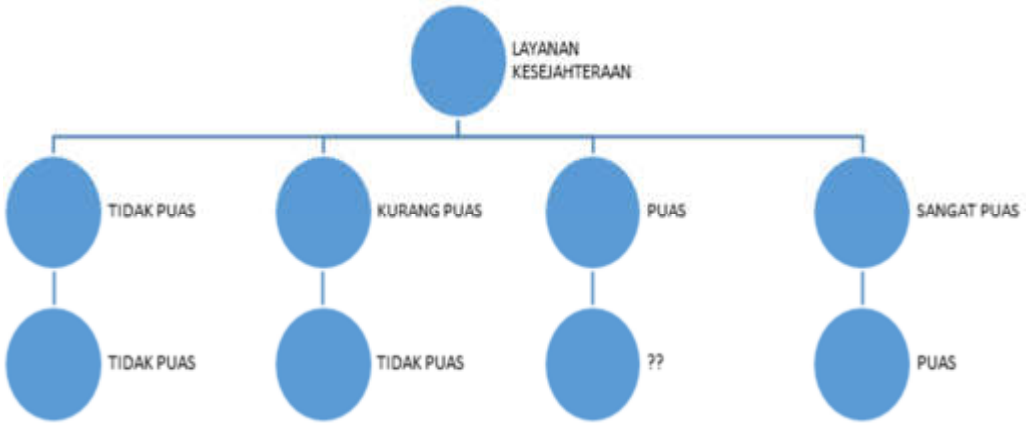

Figure 5 First Iteration Decision Tree

The decision tree above has not been completed because if the Welfare Service is Satisfied, then a decision has not been obtained on what answer, then the calculation is carried out again in iteration 2 specifically for Good Competence.

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

Table 4 Welfare Service Data Satisfied Answers

Welfare Service	Organization development	Interests and Talents	Reasoning and Scientific Services	Development of student creativity programs	Guidance Counseling Services	Satisfaction with student services
3	2	3	1	3	2	1
3	3	2	2	3	3	3
3	2	3	2	3	3	1
3	2	1	3	3	3	3
3	3	3	3	3	3	3
3	3	3	1	3	4	3

Then transform the data to find the entropy and gain values as shown in table 6 before

Table 5. Data transformation to find entropy and gain at the "PUAS" interaction facility

	NUMBER OF CASES	What?	Not satisfied	ENTROPY	GAIN
Welfare Service	6	4	2	0,92	
1	0	0	0	0	
2	3	1	2	0,92	
3	3	3	0	0	
4	0	0	0	0	
Organization development					**0,25**
	0	0	0	0	
1	0	0	0	0	
2	2	2	0	0	
3	4	2	2	1	
4	0	0	0	0	
Interests and Talents					**-0,46**
1	1	0	1	0	
2	3	2	1	0,92	
3	2	2	0	0	
4	0	0	0	0	
Reasoning and Scientific soft skills services					**-0,92**
1	0	0	0	0	
2	0	0	0	0	
3	6	4	2	0,92	
4	0	0	0	0	
Guidance Counseling Services					**-0,53**
TP	0	0	0	0	
KP	1	0	1	0	
P	4	3	1	0,811278124	
SP	1	1	0	0	

Furthermore, in this iteration, the highest gain value is sought to be used as the root or second root and the question that obtains the highest gain is the organizational development aspect with the gain value **0,459148.**

Here is the decision tree from the second iteration

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

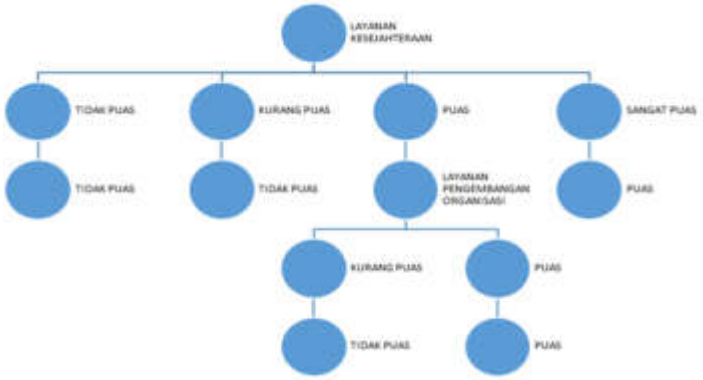

Figure 3 Second Iteration Decision Tree

In the figure above, information or rules can be obtained as follows

1. If Welfare Services Are Not Satisfied Then Student Service Satisfaction Is Not Satisfied
2. If Welfare Services Are Unsatisfied Then Student Service Satisfaction Is Not Satisfied
3. If Welfare Services Are Very Satisfied Then Student Affairs Service Satisfaction Is Satisfied.
4. If Welfare Services Are Satisfied and Organizational Development Services Are Not Satisfied, Student Service Satisfaction Is Not Satisfied
5. If the Welfare Service is Satisfied and the Learning Object Facilities are Satisfied, then the Student Affairs Service Satisfaction is Satisfied

Simulation with Rapid Miner

To Support Manual Calculation Then use*Software Rapid Miner* To find out the results of manual calculations quickly and automatically, and match the results of the manual calculation decision tree using a rapid miner.

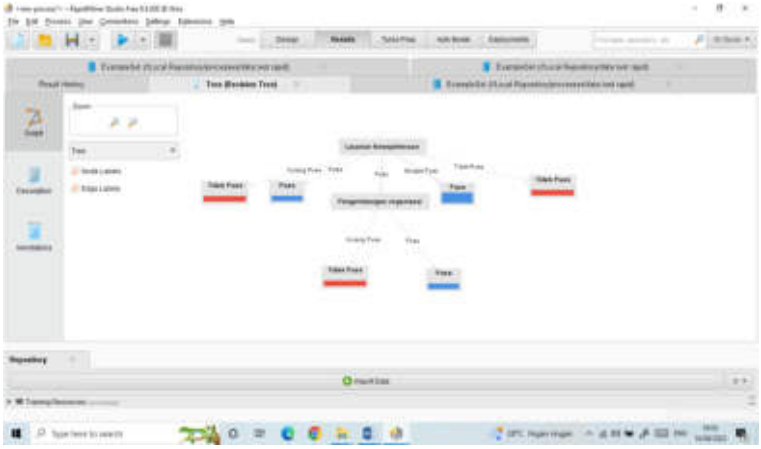

Figure 4. Decision Tree With Rapid Miner

From the decision tree image above using a rapid miner, it can be concluded that the results of manual calculations and those with a rapid miner are the same.

accuracy

The following is the accuracy of predictions of satisfaction with online satisfaction using the Decision Tree algorithm with training data and test data processed by Rapid Miner.

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

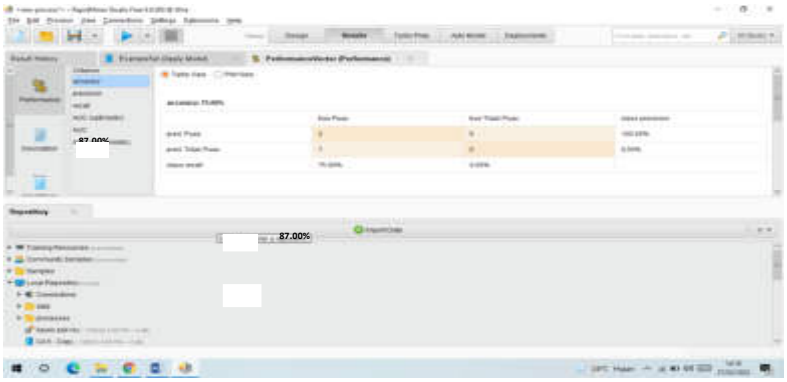

Figure 5. Accuracy of the Decision Tree algorithm with Rapid Miner

From the figure above, it can be seen that by experimenting with test data, a prediction accuracy of 87% was obtained.

4. CONCLUSION

The Decision Tree Algorithm has been successfully applied to analyze the factors of student satisfaction with Student Services. The results obtained from the questionnaire data and the application of the Decision Tree algorithm can obtain information that the most important aspect for obtaining Student Service satisfaction is student welfare services, the next most important aspect is Student Services which regarding organizational development coaching, while the prediction accuracy of satisfaction with student services obtained from test data obtained an accuracy rate of 87%.

REFERENCES

[1] Ali Ma, F., Pratama, A., Sholihin, I., & Rizki Rinaldi, A. (2021). Application of the Prediction Model Using the C4.5 Algorithm for Predicting Graduation of Wahidin Vocational High School Students (Vol. 1, Issue 1).

[2] Ardiansyah Sembiring, M., Fitri Larasati Sibuea, M., Sapta, A., Information Systems Studies, P., & Royal, S. (2018). PERFORMANCE ANALYSIS OF ALGORITHM C4.5 IN PREDICTING LEARNING OUTCOMES. In Journal of Science and Social Research (Vol. 1). http://jurnal.goretanpena.com/index.php/JSSR

[3] Fanny, I. K., Hartama, D., & Perdana, W. A. (2021). C4.5 Classification Analysis of the Factors Causing the Decline of Student Learning Achievement During the Pandemic. JOURNAL OF MEDIA INFORMATION BUDIDARMA, 5(1), 327–331. https://doi.org/10.30865/mib.v5i1.2763

[4] Janu, S., Tyas, S., Febianah, M., Solikhah, F., Kamil, A. L., Arifin, W. A., Marine Affairs, S. I., & Indonesian Education, U. (2021). COMPARISON ANALYSIS OF NAIVE BAYES AND C4.5 ALGORITHMS IN DATA MINING CLASSIFICATION FOR PREDICTING PASSAGE (Vol. 8, Issue 1).

[5] Riandari, F., & Simangunsong, A. (2019). Application of the C4.5 Algorithm to Measure the Level of Student Satisfaction.

[6] Sari, M. (2021). ANALYSIS OF CONSUMER SATISFACTION LEVEL OF NUTRITION HOUSE IN BANJARMASIN AS IMPLEMENTATION OF DATA MINING ALGORITHM C4.5. In Technologia (Vol. 12, Issue 1).

[7] Syahfitri, R. A., Windarto, A. P., & Okprana, H. (2021). Bulletin of Data Science Classification of Prospective New Customers Using C4.5 as the Basis for Providing Insurance Coverage at PT Asuransi Central Asia Pematangsiantar. Bulletin of Data Science (Media Online), 1(1), 40–48. https://ejurnal.seminar-id.com/index.php/bulletinds

[8] Yahya, N., & Jananto, A. (2019). COMPARISON OF THE PERFORMANCE OF C4.5 AND NAIVE BAYES ALGORITHMS FOR PREDICTING NEW STUDENT ADMISSION ACTIVITIES (CASE STUDY: UNIVERSITAS STIKUBANK SEMARANG).

Jurnal Info Sains : Informatikan dan Sains is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)