Jurnal Ilmiah Matematika dan Terapan
Volume 21 Number 1 June 2024 (Pages 54 - 64)
doi: https://doi.org/10.22487/2540766X.2024.v21.i1.16958

JIMT
Jurnal Ilmiah Matematika dan Terapan

ISSN : 2540 – 766X

APPLICATION OF THE C4.5 ALGORITHM TO GET CUSTOMER SATISFACTION LEVELS (CASE STUDY : TOKO CRAFT PALU, JL. SETIA BUDI)

Desy Riani Sukma Ningrum[1], Resnawati[2*], Abdul Mahatir Najar[3], Juni Wijayanti Puspita[4]

[1,2,3,4]Study Program of Mathematics, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Tadulako University

[1]desyriani0312@gmail.com, [2]r35n4w4t1@yahoo.com, [3]mahatirnajar@gmail.com, [4]juni.wpuspita@yahoo.com

*(*corresponding author)*

ABSTRACT

Customer satisfaction refers to the response expressed by customers as a result of their evaluation of the perceived difference between their initial expectations before purchase and the performance of the service after purchase. Several specific factors impact the purchasing process and the performance of the product service, such as uncertainty in store operating hours and limited availability of inventory. These related issues have an impact on customer satisfaction, especially at Craft Palu store. The aim of this research is to determine the level of customer satisfaction and accuracy level using the decision tree method, specifically the C4.5 Algorithm. In this study, the measured variables of customer satisfaction at Craft Palu store are Tangibles, Reliability, Responsiveness, Assurance, and Empathy. Based on the results of this research, it is found that Reliability is the most influential variable with an index value is 80,6% of respondents satisfied with the 5th statement, and accuracy test results using the C4.5 Algorithm in python software show an improvement with a decent final accuracy is 90%. Therefore, the C4.5 Algortihm is suitable for measuring customer satisfaction.

Keywords : C.45 Algortihm, Decision Tree, Customer Satisfaction, Python.

I. INTRODUCTION

In its development, the business world requires entrepreneurs to be quick and responsive in making decisions so that the business entity they establish can survive amidst the current situation and conditions, especially after the Covid-19 pandemic that hit. One of the steps that entrepreneurs who have small and medium businesses can take is to pay attention to service quality, especially those related to the level of customer satisfaction, which will influence customer loyalty to the type of service offered by the entrepreneur.

The quality of service provided is the main factor that contributes to customer satisfaction. Companies need to pay attention to important aspects for customers, such as service quality, product quality and price so that customers feel satisfied according to customer expectations. In assessing the level of customer satisfaction, one often compares the added value of the product or service performance received during the purchasing process with other companies. In principle, the level of customer satisfaction involves the difference between the level of perceived significance and the perceived achievement or result. Conversely, a situation of dissatisfaction can arise when the results obtained do not meet customer expectations [2].

One of the small and medium businesses in the city of Palu which is located on Jl. Setia Budi is a Palu Craft shop. The Palu Craft shop is a retail business that provides various materials needed for making dowries and gifts. Apart from providing dowry materials and gifts, the Palu Craft shop provides dowry frame making services. However, there are still several things that influence the level of customer satisfaction, such as uncertain store operating hours and limited stock availability. The related problems will have little effect on customer satisfaction at the Craft Palu shop. Steps that can be taken to improve service quality by finding out and understanding customer needs. Because with feedback from customers, businesses can improve the quality of their services.

To measure customer satisfaction, there are several methods that can be used, one of which is the decision tree. A decision tree is a structure that can be used to divide a large data set into smaller sets of records by applying a series of decision rules [4]. The decision tree process involves transforming data into a tree, changing the tree model into rules and simplifying the rules. There are many ways that can be used to create a decision tree. One of the well-known and effective data mining algorithms is the C4.5 Algorithm. This algorithm is used to create decision trees that can be used for classification and prediction. In the context of customer satisfaction, a decision tree built using the C4.5 Algorithm can help identify factors that have a significant influence on the level of customer satisfaction. Therefore, the purpose of the research is to apply the C4.5 Algorithm to determine the level of customer satisfaction and gain a deeper understanding of the elements that influence the level of customer satisfaction.[7]

II. METHODS

2.1. Population

Population is the entire object of research, population involves not only individual humans but also other natural entities and elements. Population is also more than just the number of entities being investigated, but includes all the attributes or characteristics possessed by the subject or object, and the sample represents a portion of that population.[5]

This research was carried out by collecting data from respondents who provided responses. The data collected comes from a sample that reflects the entire population, therefore, the sample selected must fully reflect the population. The population in this research are customers who buy at the Toko Craft Palu.

2.2. Sampling technique

The sample selection in this research was carried out by accidental sampling, which according to Sugiyono (2013), is a sample selection method based on chance, where consumers are coincidentally at the location of the incident. The accident sampling technique was used because Palu Craft Shop customers were very difficult to identify one by one and required longer research time. Therefore, samples were taken using the formula according to Wibisono in Akdon and Ridwan (2013)[1] as follows.

$$n = \left[\frac{Z_{\frac{a}{2}}\sigma}{e}\right]^2$$

Where

n = number of sample or minimum sample size

$Z_{\frac{a}{2}}$ = Z table value (value obtained from the normal table for the level of confidence, where the confidence level is 95%)

σ = Population standard deviation (0.25 = already stipulated)

e = Sampling error rate (in this study taken 5%)

III. RESULTS AND DISSCUSSION

3.1. Data Collection

The data used is the result of a questionnaire distributed to 96 respondents. Some of the attributes used are as follows.

a. Tangibles

b. Reliability

c. Responsiveness

d. Assurance

e. Empathy

Questionnaire data obtained from respondents in the form of questions about customer satisfaction at the Palu Craft Shop. For each attribute, a value is given to determine whether the respondent is satisfied or dissatisfied, calculated based on the resulting value, as follows.[6]

Very Satisfied = 5
Satisfied = 4
Quite Satisfied = 3
Dissatisfied = 2
Very Dissatisfied = 1

3.2. Data Transformation

In the process of collecting data from distributing questionnaires, the data obtained is data in numerical form so data transformation needs to be carried out to obtain categorical data.

There are no definite rules regarding how many categories need to be created and the score limits that must be given to each category. So, in this study the researchers divided it into 2 categories, namely High and Low.[3]

Then the category interval is determined using the following equation.

$$R = NT - NR \tag{1}$$

Where
R = Range
NT = Highest Value
NR = Lowest Value

The formula for finding the length of the class/interval is:

$$i = \frac{R}{K} \tag{2}$$

Where
$i =$ Length of Class/Interval
$R =$ Range
$K =$ Number of Class

3.3. C.45 Algorithm Training

To determine the attribute that will be the root of the decision tree with the largest gain ratio, the first step is to count the number of cases. Next, continue calculating the entropy value, gain information, split information and gain ratio for each attribute. The following are the results of manual calculations from the C4.5 Algorithm. for thisThe calculation is denoted as follows.

$$\begin{aligned}
Total\ Entropy &= \textstyle\sum_{i=1} -p_i \times \log_2 p_i \\
&= \left(-\frac{x}{z}\right) \times \log_2\left(\frac{x}{2}\right) + \left(-\frac{y}{z}\right) \times \log_2\left(\frac{y}{z}\right) \\
&= \left(\left(\frac{-50}{77}\right) \times \log_2\left(\frac{50}{77}\right)\right) + \left(\left(\frac{-27}{77}\right) \times \log_2\left(\frac{27}{77}\right)\right) \\
&= 0{,}934646644.
\end{aligned}$$

$$
\begin{aligned}
Entropy\ Tangibles_{tinggi} &= \textstyle\sum_{i=1}^{n} -p_i \times \log_2 p_i \\
&= \left(-\frac{x}{z(Tt)}\right) \times \log_2\left(\frac{x}{z(Tt)}\right) + \left(-\frac{y}{z(Tt)}\right) \times \log_2\left(\frac{y}{z(Tt)}\right) \\
&= \left(\left(\frac{48}{57}\right) \times \log_2\left(\frac{48}{57}\right)\right) + \left(\left(-\frac{9}{57}\right) \times \log_2\left(\frac{9}{57}\right)\right) \\
&= 0{,}629249224.
\end{aligned}
$$

$$
\begin{aligned}
Entropy\ Tangibles_{rendah} &= \textstyle\sum_{i=1}^{n} -p_i \times \log_2 p_i \\
&= \left(-\frac{2}{20}\right) \times \log_2\left(\frac{2}{20}\right) + \left(-\frac{18}{20}\right) \times \log_2\left(\frac{18}{20}\right) \\
&= 0{,}468995594.
\end{aligned}
$$

$$
\begin{aligned}
Gain\ (S,A) &= Entropy(S) - \textstyle\sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i) \\
&= Total\ Entropy - \left(\frac{z(Tt)}{z} \times Entropy(Tt)\right) - \left(\frac{z(Tt)}{z} \times Entropy(Tr)\right) \\
&= (0{,}934646644) - \left(\frac{57}{77} \times 0{,}629249224\right) - \left(\frac{20}{77} \times 0{,}468995594\right) \\
&= 0{,}34702174.
\end{aligned}
$$

$$
\begin{aligned}
SplitInfo(S,A) &= -\textstyle\sum_{j=1}^{k} \frac{S_j}{S} \times \log_2\left(\frac{S_j}{S}\right) \\
&= \left(-\left(\left(\frac{z(Tt)}{z}\right) \times \log_2\left(\frac{z(Tt)}{z}\right)\right)\right) + -\left(\left(\frac{z(Tr)}{z}\right) \times \log_2\left(\frac{z(Tr)}{z}\right)\right) \\
&= \left(-\left(\frac{57}{77}\right) \times \log_2\left(\frac{57}{77}\right)\right) + \left(-\left(\frac{20}{77}\right) \times \log_2\left(\frac{20}{77}\right)\right) \\
&= 0{,}826354168.
\end{aligned}
$$

$$
\begin{aligned}
GainRatio(S,A) &= \frac{Gain\ (S,A)}{SplitInfo(S,A)} \\
&= \frac{0{,}34702174}{0{,}826354168} \\
&= 0.419943111,
\end{aligned}
$$

Where

x : Number of satisfied case

y : Number of not satisfied case

z : Total case

r : Low

t : High

T : Tangibles

Re : Realibility

Rs : Responsiveness

As : Assurance

Em : Empathy

The calculation of each node in these paper given as follow.

Table 1 : Calculation for Node 1

Node	**Faktor**	**Jumlah Kasus**	**Tidak Puas**	**Puas**	***Entrophy***	***Gain***	***Split Info***	***Gain Ratio***
1	**Total**	**77**	**50**	**27**	**0,934646644**			

	Tangibles					0,34702174	0,826354168	0,419943111
	Tinggi	57	48	9	0,629249224			
	Rendah	20	2	18	0,468995594			
	Reliability					0,428135666	0,761587787	0,562161938
	Tinggi	60	50	10	0,650022422			
	Rendah	17	0	17	0			
	Responsiveness					0,215659306	0,655023991	0,32923879
	Tinggi	64	49	15	0,785560292			
	Rendah	13	1	12	0,391243564			
	Assurance					0,304721688	0,655023991	0,465206912
	Tinggi	64	50	14	0,757878463			
	Rendah	13	0	13	0			
	Empathy					0,276753924	0,624274101	0,443321169
	Tinggi	65	50	15	0,779349837			
	Rendah	12	0	12	0			

From the results in Table 1 it can be seen that the total number of known cases is 77, the number of dissatisfied responses is 27 and the number of satisfied responses is 50. So that the calculation results shown in Table 3.1 are obtained, it can be seen that the attribute with the highest gain ratio is Reliability, namely 0.562161938. This Reliability becomes the root node. There are two attribute values for Reliability, namely low and high. In the low Reliability instance, 1 case has been classified as dissatisfied. Meanwhile, high reliability still requires further calculations. The results of the decision tree formed from the calculation of node 1 are depicted in Figure 1.

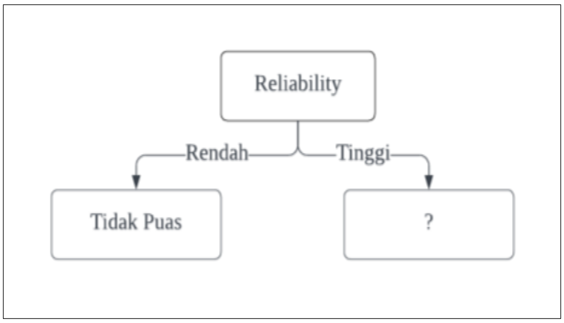


Figure 1 : Decision Tree for Node 1

The calculation continues by looking for branch nodes with high attribute values by looking for attribute values other than the root node (Reliability). Next, calculate the number of cases for satisfied responses, the number of cases for dissatisfied responses, entropy, information gain, split information, and gain ratio for each attribute. Do the calculations until all cases have includes in the class. The final branch can be seen in the following figure.

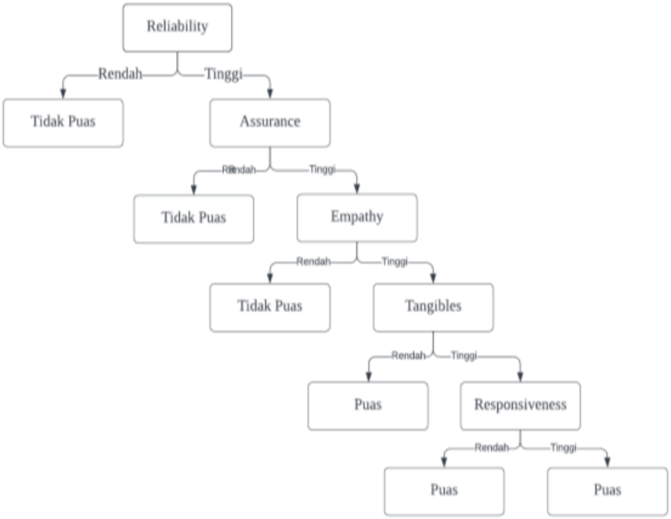


Figure 2 : Decision Tree for Node 1.2.2.2.2

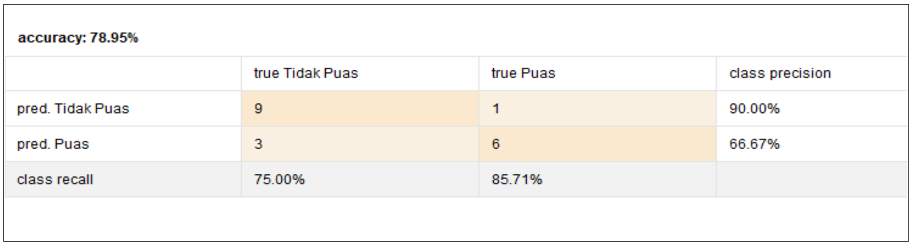
The decision tree that is formed up to the final stage is shown in Figure 2. By looking at the decision tree in Figure 2, it is known that all cases are included in the class. Thus, the decision tree in Figure 2 is the final decision tree formed.

From Figure 2, a rule is obtained for customer satisfaction using the C4.5 Algorithm as follows.

1. If Reliability is Low Then You Are Not Satisfied.
2. If Reliability is High and Assurance is Low Then You Are Not Satisfied.
3. If Reliability is High and Assurance is High and Empathy is Low Then You Are Not Satisfied.

4. If Reliability is High and Assurance is High and Empathy is High and Tangibles are Low Then You are Satisfied.
5. If Reliability is High and Assurance is High and Empathy is High and Tangibles are High and Low Responsiveness So Satisfied.
6. If Reliability is High and Assurance is High and Empathy is High and Tangibles are High and High Responsiveness So Satisfied.

3.4. Manual Testing Prediction Results



accuracy: 78.95%

	true Tidak Puas	true Puas	class precision
pred. Tidak Puas	9	1	90.00%
pred. Puas	3	6	66.67%
class recall	75.00%	85.71%	

Figure 3 : C4.5 Algorithm Accuracy Calculation Results

In this step, the 6 rules that have been obtained from the mining process will then be used as a reference in predicting the target class (Satisfied and Dissatisfied) on the test data. 19 data were used. Based on the results of precision, recall, and testing accuracy produces an accuracy of 78.95%.

Results based on Figure 3 show that from a total of 19 data, there were 15 data that were predicted correctly and 4 data with incorrect predictions. Based on the accuracy formula, we obtain:

$$level\ of\ accuracy = \frac{The\ number\ of\ the\ correct\ predictions}{The\ total\ number\ of\ prediction} \times 100\%$$

$$= \frac{15}{19} \times 100\%$$

$$= 78{,}95\%.$$

3.5. Consumer Response Index Analysis

The analysis was carried out to obtain an overview of the services at the Palu Craft Shop. In this research, questionnaire data in the form of qualitative data was converted into quantitative data. by providing scoring to the respondent's questionnaire. Therefore, the calculation of the respondent response index is expressed using the following formula.

$$Index\ value = \frac{(F_1 \times 1) + (F_2 \times 1) + (F_3 \times 1) + (F_4 \times 1) + (F_5 \times 1)}{5}$$

Where

F_1 = Number of responden choosing option 1

F_2 = Number of responden choosing option 2

$F_3 =$ Number of responden choosing option 3

$F_4 =$ Number of responden choosing option 4

$F_5 =$ Number of responden choosing option 5

The result for one variable (tangibles) can be seen in the following Table 2.

Table 2 : Consumer Index value for Tangibles Variable

No	Variable Tangibles	Respond of Responden					Quantity	Index	criteria
	Questions	**SP**	**P**	**CP**	**TP**	**STP**			
1.	Anda merasa puas dengan kemudahan dalam menemukan Lokasi Toko Craft Palu	29	44	20	2	1	386	77,2	*Tinggi*
2.	Anda merasa puas dengan kenyamanan yang ditawarkan oleh Toko Craft Palu	*30*	*41*	*21*	*4*	*0*	*385*	*77*	*Tinggi*
3.	Anda merasa puas dengan Fasilitas yang disediakan Toko Craft Palu	*29*	*29*	*33*	*4*	*1*	*369*	*73,8*	*Tinggi*
Jumlah								**1140**	**228**
Rata-rata								**380**	**76**

3.6. Discussion

The prediction process using the C4.5 Algorithm was conducted using 96 data points in this research. The research utilizes five attributes: Tangibles, Reliability, Responsiveness, Assurance, and Empathy. The data was divided into two datasets: the training dataset, which contained 77 data points, and the test dataset, which contained 19 data points.

During the data training phase using the C4.5 Algorithm, computations are performed to identify the attribute with the greatest gain ratio, which will serve as the root of the decision tree. Initially, the computation involves determining the overall count of cases, the entropy of each characteristic, the information gain, the split information, and the gain ratio. Upon completion of all calculations, it is determined that the Reliability attribute exhibits the highest gain ratio, therefore making it the selected root node for the decision tree. The computation persists until all decisions are categorized without any remaining unclassified branch nodes.

Next, testing is conducted utilizing the C4.5 Algorithm in the Python programmer. The purpose of this test is to assess the program's efficiency in executing the detection process and the accuracy level of the decision tree that was previously developed. Out of the 19 data points included in the testing process, 15 had accurate forecasts and 4 had inaccurate predictions.

The manual testing yielded an accuracy rate of 78.95%. Nevertheless, when conducting tests with python, the resulting accuracy results were 90%. The discrepancy arises from the inadequate allocation of training data in manual testing. Specifically, out of the 100-questionnaire data utilized, there were 50 positive responses and 27 negative responses. Consequently, the data utilized is biassed, leading to an accuracy level of only 78.95%. In order to improve precision, the training data was redistributed in a more equitable manner, resulting in an accuracy rate of 90%. By improving the distribution of training data, the test results in Python demonstrate notable enhancements in performance.

Next, we will examine the analysis of client happiness, which has been generated through manual computations and aided by software. The data indicates that 80.6% of respondents expressed satisfaction with the Reliability dimension, specifically regarding the dexterity of Palu Craft Shop staff in handling customer requests.

These results demonstrate that the service quality at Toko Craft Palu is excellent, positively impacting consumer happiness and providing tangible evidence of the store's ability to deliver promised services.

IV. CONCLUSION

Based on the research conducted, it can be concluded that the results of calculating the level of satisfaction using the C4.5 method showed that the Reliability variable was high with an index value of 80.6% of respondents who were satisfied with the statement of the dexterity of Palu Craft Shop employees in handling customer orders.

The manual accuracy level using Rapid Miner software is 78.95% and the accuracy using Python software is 90%, because data distribution is very important in determining the accuracy of the C4.5 Algorithm model.

REFERENCES

[1]. Akdon dan Riduwan. (2013). Rumus dan Data Dalam Analisis dan Statistika. Bandung: Alfabeta

[2]. Ariyani, E. R., dan Nurcahyo, B. (n.d.). (2009). *Service Quality Effect Analysis of Customer Sstisfaction in Restaurant*. http://www.gunadarma.ac.id.

[3]. Azwar, S. (2012). Penyusunan Skala Psikologi edisi 2. Yogyakarta: Pustaka Pelajar.

[4]. Berry, Michael J.A. dan Gordon S. Linoff. (2004). *Data Mining Technique for Marketing, Sales and Customer Relationship Management, Second Edition, Wiley Publishing, Inc.*

[5]. Sugiyono. (2013). Metode Penelitian Kuantitatif Kualitatif dan R&D. (Bandung: Penerbit Alfabeta, Hal. 85).

[6]. Telaumbanua, D. dan Kurniawati, I. (2022). Penerapan Algoritma C4.5 Untuk Klasifikasi Kepuasan Pelanggan Pada Jasa Layanan Pengiriman INFORMASI ARTIKEL ABSTRACT. *Jl. Kramat Raya*, *06*(01). https://doi.org/10.46961/jommit.v6i1.

[7]. Witten, I.H dan Frank, E.I. (2005). Data Mining Practical Machine Learning Tools and Techniques, second edition. San Fransisco : Morgan KJayanthi, P., Salvagopal, P., & Sundaram, S. S. 2012. "Some $C3-$ Supermagic Graphs." Until. Math. 89, 357-366.