

TESTING C4.5 ALGORITHM USING RAPID MINER APPLICATIONS IN DETERMINING CUSTOMER SATISFACTION LEVELS

Fithrie Soufitri^{1*}, Ellanda Purwawijaya², Eka Hayana Hasibuan³, Roy Nuary Singarimbun⁴

Universitas Battuta Medan, Sumatera Utara

fithriesoufitri@gmail.com¹, ellanda.purwa.wijaya@gmail.com², ekahayana@man1medan.sch.id³,
roy90singarimbun@gmail.com⁴

Article Info

Received, 01 Mei 2021

Revised, 20 Mei 2021

Accepted, 28 June 2021

Data mining is a series of processes to extract added value in the form of information that has not been known manually from a database. The resulting information is obtained by extracting and recognizing important or interesting patterns from the data contained in the database. One part of the service of an agency is the service to customers, customers or those related to certain services. The quality of service is assessed by what has been done and how to treat from those who serve, some provisions used to ensure service optimization are Confiden, Integrity, Pride and Passion with the main purpose or output is Customer Satisfaction. The process of Forming a Pattern of Satisfaction Level by Utilizing the C4.5 Algorithm Penelitian process is carried out by data collection, testing with applications, exposure of pattern results or Knowledge. Pola yang terbentuk after the extraction is 1.Integrity = low: quite satisfied {very satisfied=0, quite satisfied=3}Integrity = tall, Passion = low: quite satisfied {very satisfied =0, quite satisfied =2}, Passion = tall: very satisfied {very satisfied =11, quite satisfied =0}.

Keywords: Satisfaction Level, Data Mining, C4.5 algorithm, Pattern

1. INTRODUCTION

The quality of customer service must get special attention so that the company can continue to maintain an image in the eyes of customers. By prioritizing good service, customers will make it easier for the company to achieve the goal of obtaining maximum profit through increasing the number of customers who stay continuously. Customer satisfaction according to Tjiptono(2008) is the extent to which the presumption of product or service performance meets buyer expectations. If the product performance is lower than the customer's expectations, then the customer is not satisfied, if the performance matches or exceeds the customer's expectations, then the customer is satisfied. Satisfaction can be interpreted as a response to the perceived evaluation between previous expectations and accepted reality. Of course, if the expectations are higher than the reality received then the customer feels dissatisfied and complains. Customer satisfaction or dissatisfaction is a function of perceived nonconformity due to the difference between expectations and actual performance.[1]–[5].

In this research, activities are carried out to extract data in determining customer satisfaction with a particular service by mandating data mining techniques. Data mining is a series of processes to extract added value in the form of information that has not been known manually from a database. The resulting information is obtained by extracting and recognizing important or interesting patterns from the data contained in the database. Data mining is mainly used to find knowledge contained in large databases so often called Knowledge Discovery Database (KDD). [3],[6] Data mining is the mining or discovery of new information by looking for specific patterns or rules of a very large

 INFOKUM is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

amount of data. Data mining is an interactive and interactive process of finding new patterns or models that can be generalized for the future, useful and understandable in a massive database. Data mining on the other hand is an activity that includes the collection, use of historical data to find regularities, patterns or relationships in large data sets.[7], [8].

The process used to explore the potential of a set of data by following steps and rules 1) Data selection is a set of operational data needs to be done before the stage of extracting information in knowledge discovery in database (KDD) begins. The selection data to be used for the data mining process, stored in a file, separate from the operational database followed by 2) The preprocessing process includes, among others, discarding data duplication, checking inconsistent data, and correcting errors in the data, such as printing (typography). Enrichment process is also carried out, which is the process of "enriching" existing data with data or other information that is relevant and necessary for KDD, such as data or external information. 3) The transformation phase of data forms that do not yet have a clear entity into a valid data form or are ready for the Data Mining process. 4) Data Mining is this phase that is done is applying algorithms or methods of knowledge search. 5) Evaluation In this last phase that is done is an easy to understand output formation process that is sourced from the data mining process of information patterns. [9]–[11]

C4.5 is a widely known algorithm used for the classification of data that has numerical and categorical attributes. The result of the classification process in the form of rules can be used to predict the value of discrete attributes of the new record. Algorithm C4.5 itself is a development of the ID3 algorithm, where development is done in terms of, can overcome missing data, can overcome continuous data and pruning. In general the C4.5 algorithm for building decision trees with the main working step 1) Select attributes as roots. 2) Create a branch for each value 4). Divide cases in branch 4) Repeat the process for each branch until all cases on the branch have the same class. The C4.5 algorithm has the main advantage of being able to produce a tree model or rule that is easily interpreted, has an acceptable level of accuracy, can handle discrete and numerical type attributes. C4.5 algorithm is an algorithm used to form decision trees based on decision forming criteria.[1]–[3]

2. METHOD

To help in the preparation of this research, a framework is prepared for a clear framework (Frame Work) at its stages. This framework is the steps that will be taken in solving the problem that will be discussed. Here are the steps taken in accordance with the flow of the author's research design as follows:

1. Identifying problems
This stage is the stage of identifying or introducing the problem to be discussed. In this section it is determined what is the problem that is the determination of the right position or position for employees and workers in supporting the success of a job.
2. Data Collection, This stage is the stage of collecting data needed to help solve problems in this study, using secondary data, namely from previous research that has produced a whitish level of Customer satisfaction.
3. Data Analysis This stage is the stage of converting data to a form that is easier to implement to a form that is easier to implement with the algorithm to be used.
4. Data Transformation, at this stage, processes the data generated in accordance with the needs of rapid miner applications.
5. Testing using the Rapid Miner App
6. Present the results found and give conclusions against.

3. RESULTS AND DISCUSSIONS

3.1 Analysis of the Use of Algorithm C4.5

1. Data Needs

Data used as part of the extraction in the determination of the most appropriate position of position to workers in a particular agency with the following test data:

Table 1. Test Data with C4.5 Algorithm

Objek	Confiden	Integrity	Pride	Passion	Customer Satisfaction
A	good	Tall	Tall	Tall	Very satisfied
B	good	Tall	Tall	Tall	Quite satisfied
C	Good	Tall	Tall	Tall	Very satisfied
D	Good	Tall	Tall	Tall	Very satisfied
E	enough	Low	Tall	Low	Quite satisfied
F	enough	Tall	Tall	Low	Quite satisfied
G	Good	Tall	Tall	Tall	Very satisfied
H	enough	Tall	Tall	Tall	Very satisfied
I	Good	Tall	Tall	Tall	Very satisfied
J	Good	Tall	Tall	Tall	Very satisfied

2. Entropy and Gain Value Formation

The first step of the decision formation is to determine the values of Entropy and Gain with the formula and direction that have been outlined in algorithm C4.5. The description of the results of the calculation of Entropy Value and gain value is calculated in the following table:

Table 2. Entropy and Gian Value

Total		Number of case	very satisfied	quite satisfied	Entropy	Gain
		12	4	8	0,918295834	
Confiden						0,5559195
	good	8	7	1	0,543564443	
	enough	4	0	4	0	
Integrity						0,3166891
	tall	10	8	2	0,721928095	
	low	2	0	2	0	
Pride						0,3166891
	tall	10	8	2	0,721928095	
	low	2	0	2	0	
Passion						0,5408521
	tall	9	8	1	0,503258335	
	low	3	0	3	0	

From the calculation of the table above, the process of comparison of the terTall value of each attribute and heard as the root in the formation of the Decision tree, with the Confiden attribute with a value of 0.5559195 so that the graph is formed with a description if the condifen value is enough then based on the calculation user satisfaction is quite satisfied and If the confiden is good then tested on other attributes.

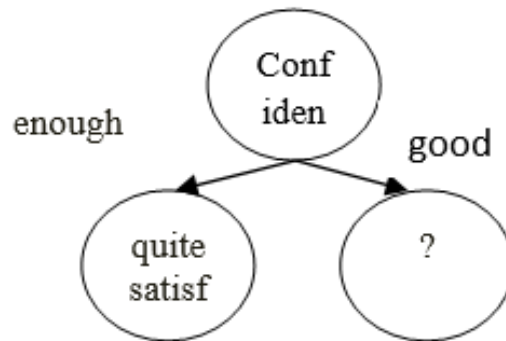


Figure 1. First Node Of Vaginal Discharge

Based on the first node, the search is carried out by helping new tables, by eliminating the root node and will be done more in-depth, the description of the new data is explained below:

Table 3. New Data Formation for 2nd Node Testing

Total	Jumlah Kasus	very satisfied	quite satisfied	Entropy	Gain
Integrity					0,3166891
tall	10	8	2	0,721928095	
low	2	0	2	0	
Pride					0,3166891
tall	10	8	2	0,721928095	
low	2	0	2	0	
Passion					0,5408521
tall	9	8	1	0,503258335	
low	3	0	3	0	

From the calculation of the table above, the process of comparison of the terTall value of each attribute and expressed as the 2nd node in the formation of the Decision tree, with the passion attribute with a value of 0.5408521 so that the graph is formed with a description if the value of confiden enough then based on the calculation user satisfaction is quite satisfied and If confiden Good then tested the passion value whether tall or low, for the picture with the following graph:

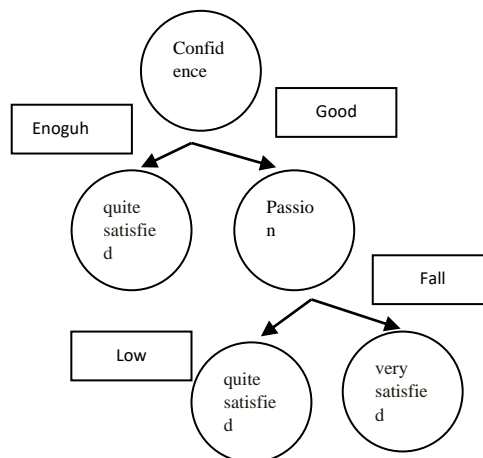


Figure 2. 2nd Node Decision Tree

From figure 2 will be formed information that can be used as knowledge is if confidence enough then the level of satisfaction is low, if good assessment will be tested on passion and has provided new information if passion falls then tall satisfaction level and if low passion it will lead to insatiation so that testing for each node activity will produce some of the following patterns:

1. If confidence enough then low customer satisfaction level
2. If confidence is good and Passion Low then the level of customer satisfaction is Low
3. If confidence enough and Passion Fall Then High Customer Satisfaction Level
4. If confidence enough and Passion Fall and Integrity fall Then Customer Satisfaction Level Tall
5. If confidence enough and Passion Fall and Integrity low then customer satisfaction levels are low.

3.2 Testing With Rapid Miner Application

The test data that has been provided is prepared as an input from the needs of C4.5 Algorithm by utilizing Microsoft Excel and stored in a certain format and carried out load against the rapid miner application, the display of the results of the data load in the following image:

Table 4. Data Load View

Row No.	Confiden	Integrity	Pride	Passion	Customer S...
1	good	tall	tall	tall	very satisfied
2	enough	tall	tall	tall	quite satisfied
3	good	tall	tall	tall	very satisfied
4	good	tall	tall	tall	very satisfied
5	enough	low	tall	low	quite satisfied
6	enough	tall	tall	low	quite satisfied
7	good	tall	tall	tall	very satisfied
8	enough	tall	tall	tall	very satisfied
9	good	tall	tall	tall	very satisfied
10	good	tall	tall	tall	very satisfied

Figure 1. Load Data View

Based on the data in figure 1 with statistical information, namely attributes as input as much as 4 and 1 attribute as external or decision results needed with the overall data type is polynominal with a missing value of 0. Description of these statistics in the following image:

Name	Type	Missing	Statistics	Filter (5 / 5 attributes):
Confiden	Polynominal	0	Least enough (4) Most good (6)	Search for Attributes
Integrity	Polynominal	0	Least low (1) Most tall (9)	
Pride	Polynominal	0	Least tall (10) Most tall (10)	
Passion	Polynominal	0	Least low (2) Most tall (8)	
Customer Satisfaction	Polynominal	0	Least quite satisfied (3) Most very satisfied (7)	

Figure 3. Statistical Value Against Satisfaction Level Attributes

1. Testing Architecture

For the C4.5 algorithm testing architecture with the preparation of retrivel determination, which is the process of determining the data used, then formed a set role with the selection of external results

INFOKUM is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

from the test. The next design is to form a tree with the selection of Decision Tree operators. The description of the test architecture is described in the following image:

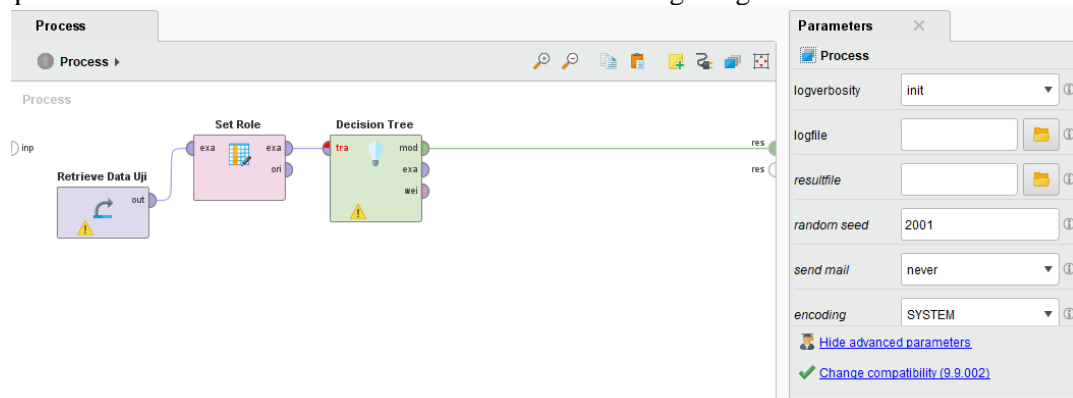


Figure 4. Algorithm Testing Architecture C4.5

2. Pembentukan Pohon Keputusan

The decision tree as knowledge generated using Algoritam C4.5 is described in the following image:

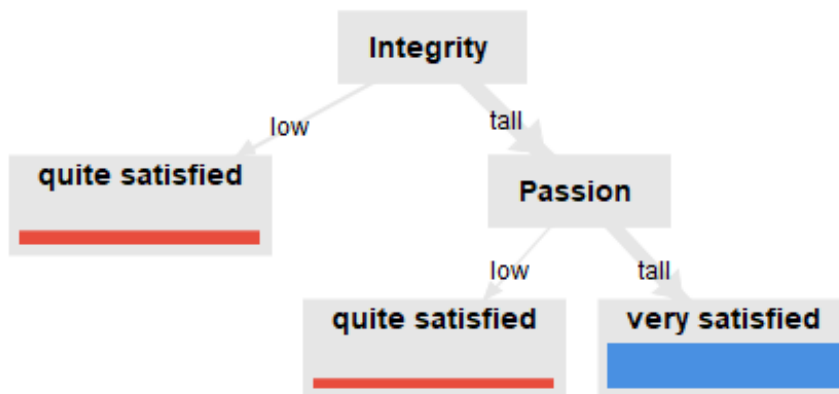


Figure 5. Decision Tree

Based on figure 4 found results and knowledge with some of the following features and patterns:

1. Integrity = low: quite satisfied {very satisfied=0, quite satisfied=3} Integrity = tall
2. Passion = low: quite satisfied {very satisfied=0, quite satisfied=2}
3. Passion = tall: very satisfied {very satisfied=11, quite satisfied=0}

3. Customer Satisfaction Pattern Graph

The spread of the pattern of formation of customer satisfaction levels is formed based on the following 5 figures:

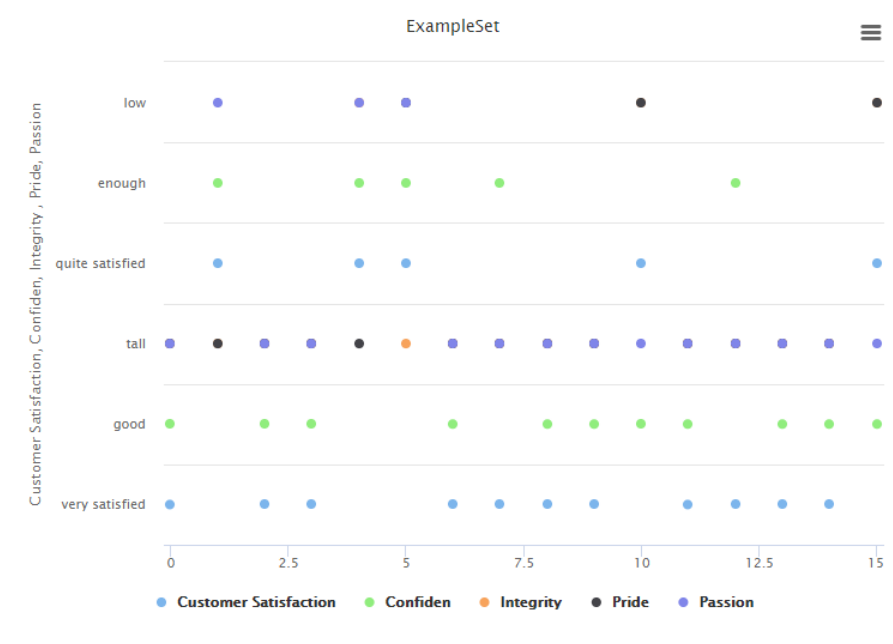


Figure 5. Customer Satisfaction Level Deployment Graph.

4. CONCLUSION

Algoritma C4.5 is able to provide information about user satisfaction patterns by proving using the Rapid Miner application, a pattern formed as information that can be used as part of knowledge is if confidence enough then the level of satisfaction is low, if good judgment will be tested against passion and has provided new information if passion falls then the level of satisfaction is high and if passion is low it will lead to disgy.

Reference

- [1] E. S. Palupi and S. M. Pahlevi, "Klasifikasi Opportunity Menggunakan Algoritma C4.5, C4.5 Dan Naïve Bayes Berbasis Particle Swarm Optimization," *Inti Nusa Mandiri*, vol. 14, no. 2, 2020.
- [2] S. Haryati, A. Sudarsono, and E. (2015) Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, vol. 11, no. 2, 2015.
- [3] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, 2015.
- [4] Supardi, "Peran Kepemimpinan Dan Keterlibatan Group Decission Making Dalam Perubahan Organisasi," *Tarbawi*, vol. 1, no. 1, 2015.
- [5] I. P. Sari and R. Harman, "Decission Tree Technique Dalam Menentukan Penjurusan Siswa Menengah Kejuruan," *J. Inf. Syst. Res.*, vol. 1, no. 4, 2020.
- [6] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [7] M. J. Hamid Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, 2018, doi: 10.14569/IJACSA.2018.090630.
- [8] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *Int. J. Inf. Technol.*, vol. 12, no. 4, 2020, doi: 10.1007/s41870-020-00427-7.
- [9] R. H. Pambudi, B. D. Setiawan, and Indriati, "Penerapan Algoritma C4.5 Dalam Program

- Untuk Memprediksi Kinerja Siswa Sekolah Menengah,” *Look. Forward, Look. Back Draw. Past To Shape Futur. Mark.*, vol. 2, no. 7, 2018.
- [10] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, “Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, 2018.
- [11] R. Novita, “Teknik Data Mining : Algoritma C4.5,” *Ilmu Komputer.com*, 2016.