Application of the C.45 Algorithm in Measuring the Satisfaction Level of Hotel Visitors

R Purba[1], **F A Sianturi**[2]

[1,2] *Department of Informatics Engineering, STMIK Pelita Nusantara, Medan, Indonesia*

E-mail: rospur97@gmail.com

Abstract. Hotel visitor satisfaction is one of the most important things in assessing the level of service by the hotel to its visitors. The purpose of this study was to determine the satisfaction of hotel visitors based on service criteria (very good, good, good enough, not good), facilities (good, pretty good, not good), quality (high, medium), and classification (first class, business, economy) hotel rooms by applying the C.45 Algorithm method. to visitors who expressed satisfaction or dissatisfaction. In The Crew Hotel, the criteria have not been measured with certainty, so the management is still manually determining visitor satisfaction so it is less effective. Thus, the author tries to measure the four criteria by applying the processing of the questionnaire dataset given to visitors, when staying, either in writing or asking visitors. using the C4.5 algorithm in measuring the level of satisfaction of visitors to The Crew so that a decision tree is formed. to advance the development of The Crew Hotel and improve service at the hotel. After the calculation is done manually, then the proof is done using an application designed by the author according to the existing rules

Keywords— Application, Data Mining, C.45 Algorithm, Hotel, Visitor Satisfaction

1. Introduction

A hotel is a building or company that provides services in the form of rooms, or a place to rest as well as recreation for visitors who come from within the city or outside the city. Every hotel will try to provide the best facilities and services. tricks for every hotel business to strengthen competition in every business. If at any time there are visitors who feel dissatisfied with the services of a hotel, it will cause disappointment and feel less suited to the hotel. One of them, the hotel which is currently under development is The Crew Hotel.

In terms of services and facilities, The Crew still has shortcomings because there are still complaints to the receptionist from visitors about hotel facilities and the quality of service. After all, if visitors are not satisfied with the service they receive, then the visitor will leave the hotel provider and become a visitor at another hotel. which in the end will greatly reduce the reputation of the hotel concerned. Therefore, measuring the level of satisfaction of hotel visitors requires a system that can measure the level of satisfaction of hotel visitors systematically and updated and one of them is using data mining.

Data mining is a process of finding meaningful relationships, patterns, and trends by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques (Santoso, n.d.). One of the algorithms in data mining is the C4.5 algorithm.
According to (Sukma Putri Utari, 2015) the C4.5 algorithm is an algorithm used to form a decision tree. While the decision tree can be interpreted as a way to predict or group very strongly.

2. Research Method

2.1. Data Needs analysis

According to Desyanti (2018), the C4.5 algorithm is one of the algorithms used to classify or group datasets. The basis of the C4.5 algorithm is the formation of a decision tree (Decision Tree). The branches of the decision tree are a classification question while the leaves are the classes or groups. Because the purpose of the C4.5 algorithm is to classify, the result of dataset processing is in the form of grouping data into certain classes.
The stages of design and development include the following:
1. Identification
The identification stage is used to determine the boundaries of the problem, those involved, and the goals to be achieved.

a. Limitations of the problem in this study are: Discussed the level of visitor satisfaction with The Crew Hotel's services by applying the C45 method. The system is designed in the form of a web by utilizing the PHP and MySQL programming languages in database management.
b. As an indicator of assessment to measure customer satisfaction are facilities, service, quality, and classification at The Crew Hotel

2. The criteria
The criteria used in this study are as follows:

Table 1. Criteria

No	Atribut	Variable	Amount Case	Amount Satisfied	Amount Not satisfied
1	Pelayanan	Sangat Baik	69	24	45
		Baik	81	43	38
		Cukup Baik	35	0	35
		Tidak Baik	15	0	15
2	Fasilitas	Bagus	70	35	35
		Lumayan	95	32	63
		Tidak Bagus	35	0	35
3	Kualitas	Tinggi	138	67	71
		Sedang	62	0	62
4	Klasifikasi	First Class	56	27	29
		Bisnis	95	40	55
		Ekonomi	49	0	49

2.2. C45 Algorithm

Based on questionnaire data from The Crew Hotel as shown in the following table :

Table 2. Questionnaire results

Pelayanan	Fasilitas	Kualitas	Klasifikasi
Sangat Baik = 69	Bagus=70	Tinggi=138	First Class=56
Baik = 81	Lumayan =95	Sedang = 62	Bisnis=95

Cukup Baik=35	Tidak Bagus=35		Ekonomi=49
Tidak Baik =15			

1. Process node 1 as root (root)
The root node is obtained by first calculating the Entropy or initializing it as E (all data) to the class composition. With the following formula:

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

Where :
S : Case Collection
A : Features
n : Number of Partitions S
pi : Proportion of Si to S

$$= -\left(\left(\frac{67}{200} x\ \log_2\left(\frac{67}{200}\right)\right) + \left(\frac{133}{200}\ x\ \log_2\left(\frac{133}{200}\right)\right)\right)$$
$$= -\big((0{,}335\ x\ (-1{,}57777)) + (0{,}665\ x\ (-0{,}58857))\big)$$
$$= -\big((-0{,}52855) + (-0{,}3914)\big) = -(-0{,}91995) = (0{,}91995)$$

2. Calculate the entropy and Gain value

Furthermore, it is the same as calculating the previous entropy, except that it is divided according to the criteria so that it is obtained as follows:

A.Pelayanan

$$E\ [Pelayanan - Sangat\ Baik] = -\left(\left(\frac{24}{69} x\ \log_2\left(\frac{24}{69}\right)\right) + \left(\frac{45}{69}\ x\ \log_2\left(\frac{45}{69}\right)\right)\right)$$
$$= -\big((0{,}34783\ x\ (-1{,}52356)) + (0{,}65217\ x\ (-0{,}61667))\big)$$
$$= -\big((-0{,}52993) + (-0{,}40218)\big)$$
$$= -(-0{,}93211) = 0{,}93211$$
$$E[Pelayanan - Baik] = -\left(\left(\frac{43}{81} x\ \log_2\left(\frac{43}{81}\right)\right) + \left(\frac{38}{81}\ x\ log_2\left(\frac{38}{81}\right)\right)\right)$$
$$= -\big((0{,}53086\ x\ (-0{,}91358)) + (0{,}46913\ x\ (-1{,}09192))\big)$$
$$= -\big((-0{,}48499) + (-0{,}51226)\big)$$
$$= -(-0{,}99725)$$
$$= 0{,}99725$$
$$E\ [Pelayanan - Cukup\ Baik] = -\left(\left(\frac{0}{35} x\ \log_2\left(\frac{0}{35}\right)\right) + \left(\frac{35}{35}\ x\ \log_2\left(\frac{35}{35}\right)\right)\right)$$
$$= -\big((0\ x\ (0)) + (1\ x\ (0))\big)$$
$$= -\big((0) + (0)\big)$$
$$= -(0)$$
$$= 0$$
$$E\ [Pelayanan - Tidak\ Baik] = -\left(\left(\frac{0}{15} x\ \log_2\left(\frac{0}{15}\right)\right) + \left(\frac{15}{15}\ x\ \log_2\left(\frac{15}{15}\right)\right)\right)$$
$$= -\big((0\ x\ (0)) + (1\ x\ (0))\big)$$
$$= -\big((0) + (0)\big)$$
$$= -(0)$$
$$= 0$$

Calculate the gain with the following formula:

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

Where :
S : Case Collection
A : Features
n : Number of Partitions S
Si : Number of Cases on the ith certainty
S : Number of Cases in S

$G\ [Total,\ Pelayanan]$ = 0,91995 $- \left((us\ \ 0{,}93211) + \left(\frac{81}{200}\ x\ \ 0{,}99725\right) + \left(\frac{35}{200}\ x\ \ 0\right) + \left(\frac{15}{200}\ x\ 0\right)\right)$
$= 0{,}91995 - \left((0{,}345\ x\ 0{,}93211) + (0{,}405\ x\ 0{,}99725) + (0{,}175\ x\ 0) + (0{,}075\ x\ 0)\right)$
$= 0{,}91995 - \left((0{,}32158) + (0{,}40389) + 0 + 0\right)$
$= 0{,}91995 - (0{,}72546) = 0{,}19449$

B .Fasilitas

$E\ [Fasilitas - Bagus] = -\left(\left(\frac{35}{70} x\ \log_2\left(\frac{35}{70}\right)\right) + \left(\frac{35}{70}\ x\ \log_2\left(\frac{35}{70}\right)\right)\right)$
$= -\left((0{,}5\ x\ (-1)) + (0{,}5\ x\ (-1))\right)$
$= -\left((-0{,}5) + (-0{,}5)\right)$
$= -(-1) = 1$

$E\ [Fasilitas - Lumayan\ Bagus] = -\left(\left(\frac{32}{95} x\ \log_2\left(\frac{32}{95}\right)\right) + \left(\frac{63}{95}\ x\ log_2\left(\frac{63}{95}\right)\right)\right)$
$= -\left((0{,}33684\ x\ (-1{,}56985)) + (0{,}66316\ x\ (-0{,}59257))\right)$
$= -\left((-0{,}52879) + (-0{,}39297)\right)$
$= -(-0{,}92176)$
$= 0{,}92176$

$E\ [Fasilitas - Tidak\ Bagus] = -\left(\left(\frac{0}{35} x\ \log_2\left(\frac{0}{35}\right)\right) + \left(\frac{35}{35}\ x\ \log_2\left(\frac{35}{35}\right)\right)\right)$
$= -\left((0\ x\ (0)) + (1\ x\ (0))\right)$
$= -\left((0) + (0)\right)$
$= -(0)$
$= 0$

$G\ [Total,\ Fasilitas]$ = 0,91995 $- \left(\left(\frac{70}{200} x\ \ 0{,}93211\right) + \left(\frac{95}{200}\ x\ \ 0{,}99725\right) + \left(\frac{35}{200}\ x\ \ 0\right)\right)$
$= 0{,}91995 - \left((0{,}35\ x\ 0{,}93211) + (0{,}475\ x\ 0{,}99725) + (0{,}175\ x\ 0)\right)$
$= 0{,}91995 - \left((0{,}35) + (0{,}43784) + 0\right)$
$= 0{,}91995 - (0{,}78784) = 0{,}13212$

C .Kualitas

$E\ [Kualitas - Tinggi] = -\left(\left(\frac{67}{138} x\ \log_2\left(\frac{67}{138}\right)\right) + \left(\frac{71}{138}\ x\ \log_2\left(\frac{71}{138}\right)\right)\right)$
$= -\left((0{,}48551\ x\ (-1{,}04243)) + (0{,}51449\ x\ (-0{,}95878))\right)$
$= -\left((-0{,}50611) + (-0{,}49238)\right)$
$= -(-0{,}99939)$
$= 0{,}99939$

$$E\ [Kualitas - Sedang] = -\left(\left(\frac{0}{62}x\ \log_2\left(\frac{0}{62}\right)\right) + \left(\frac{62}{62}\ x\ log_2\left(\frac{62}{62}\right)\right)\right)$$
$$= -\big((0x\ (0)) + (1x\ (0))\big)$$
$$= -\big((0) + (0)\big)$$
$$= -(0)$$
$$=\ 0$$
$$G\ [Total,\ Fasilitas] = 0{,}91995 - \left(\left(\frac{138}{200}x\ \ 0{,}99939\right) + \left(\frac{62}{200}\ x\ \ 0\right)\right)$$
$$= 0{,}91995 - \big((0{,}69\ x\ 0{,}93211) + (0{,}31\ x\ 0)\big)$$
$$= 0{,}91995 - \big((0{,}68958) + 0\big)$$
$$= 0{,}91995 - (0{,}68958)$$
$$= 0{,}23037$$

D. Klasifikasi

$$E\ [Klasifikasi - First\ Class] = -\left(\left(\frac{27}{56}x\ \log_2\left(\frac{27}{56}\right)\right) + \left(\frac{29}{56}\ x\ \log_2\left(\frac{29}{56}\right)\right)\right)$$
$$= -\big((0{,}48214\ x\ (-1{,}05215)) + (0{,}57895\ x\ (-0{,}78849))\big)$$
$$= -\big((-0{,}50744) + (-0{,}49164)\big)$$
$$= -(-0{,}99908)$$
$$=\ 0{,}99908$$
$$E\ [Klasifikasi - Bisnis] = -\left(\left(\frac{40}{95}x\ \log_2\left(\frac{40}{95}\right)\right) + \left(\frac{55}{95}\ x\ log_2\left(\frac{55}{95}\right)\right)\right)$$
$$= -\big((0{,}42105\ x\ (-1{,}24792)) + (0{,}57895\ x\ (-0{,}78849))\big)$$
$$= -\big((-0{,}52544) + (-0{,}45649)\big)$$
$$= -(-0{,}98194)$$
$$=\ 0{,}98194$$
$$E\ [Klasifikasi - Ekonomi] = -\left(\left(\frac{0}{49}x\ \log_2\left(\frac{0}{49}\right)\right) + \left(\frac{49}{49}\ x\ \log_2\left(\frac{49}{49}\right)\right)\right)$$
$$= -\big((0\ x\ (0)) + (1\ x\ (0))\big)$$
$$= -\big((0) + (0)\big)$$
$$= -(0) = (0)$$
$$G\ [Total,\ Klasifikasi] = 0{,}91995 - \left(\left(\frac{56}{200}x\ \ 0{,}99908\right) + \left(\frac{95}{200}\ x\ \ 0{,}98194\right) + \left(\frac{49}{200}\ x\ \ 0\right)\right)$$
$$= 0{,}91995 - \big((0{,}280\ x\ 0{,}93211) + (0{,}475\ x\ 0{,}99725) + (0{,}245\ x\ 0)\big)$$
$$= 0{,}91995 - \big((0{,}27974) + (0{,}46642) + 0\big)$$
$$= 0{,}91995 - (0{,}74616)$$
$$= 0{,}17379$$

3. Result and Discussion

The decision process is carried out through the user interface of the system. The system start page can be seen in the explanation below:

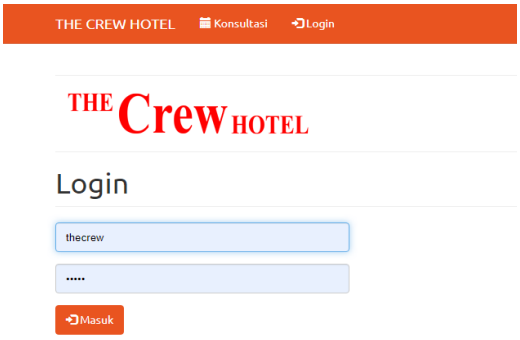

Figure 1. Login

On this login page, the user must enter the specified username and password and enter the system

Figure 2. Main Form

On the main page, the user selects the master and selects the desired form according to what will be done then the user clicks the desired menu to enter the next form

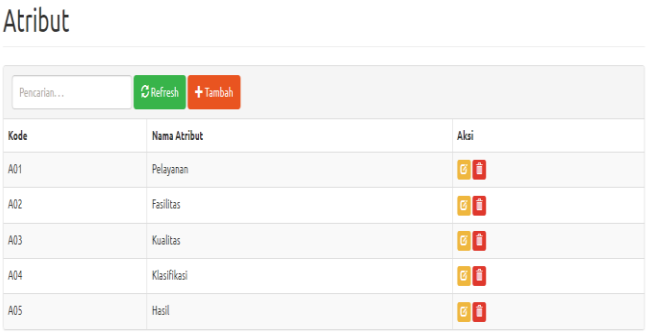

Figure 3. Atribut Form

In the attribute form display, it functions to fill in the attribute name in the provided column.

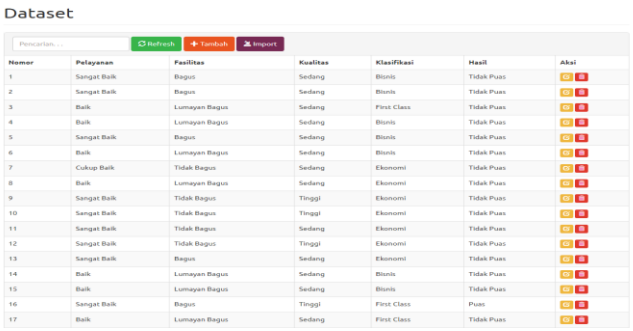

Figure 4. Dataset Form

The dataset form is used to store all the data used in the application, and next is the calculation form

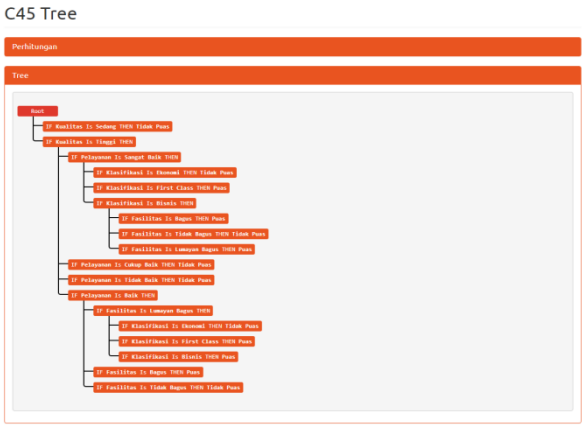

Figure 5. Tree Form

Figure 6. Calculation Form

On the process page Displays the calculation data page based on the dataset that has been created, the final result will determine whether the customer is satisfied or dissatisfied with the services provided by the management of The Crew Hotel.

.

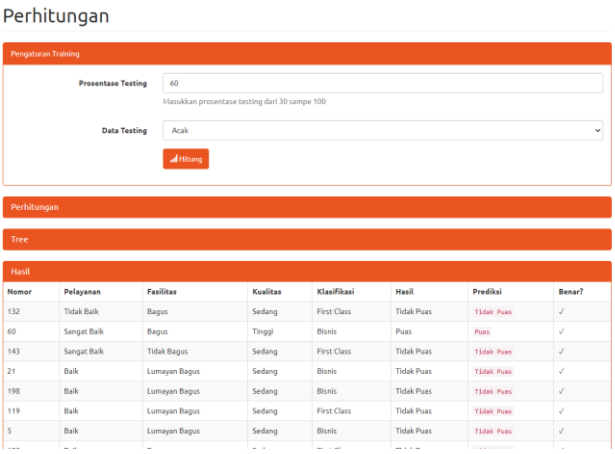

Figure 7. Accurate Form

On the dataset accuracy form page, check the accuracy of the data set that has been inputted.

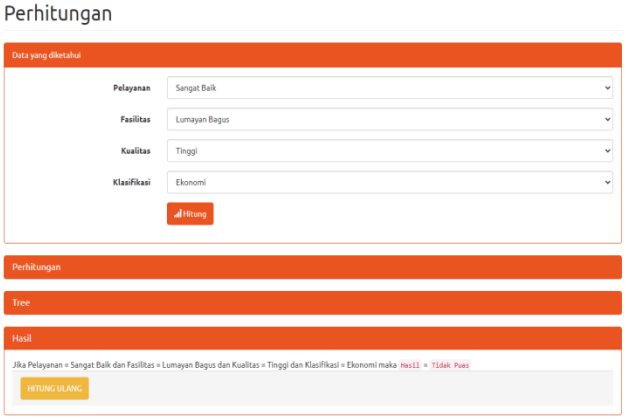

Figure 8. Global Result Form

On the dataset accuracy form page, check the accuracy of the data set that has been inputted.

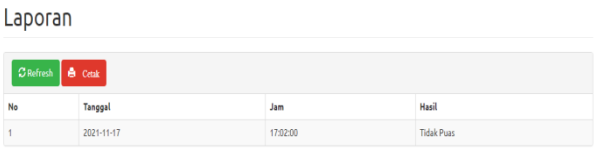

Figure 9. Report Form

4. Conclusion

By the explanation described, to measure the level of visitor satisfaction at The New Hotel using the C.45 method, therefore it can be produced, namely:

1. Applying the C4.5 algorithm in measuring the level of satisfaction of hotel visitors is applied through an assessment with the criteria (service, facilities, quality, classification). With these criteria, a decision tree is obtained for the final result in determining hotel visitor satisfaction.

2. In building data mining applications using the C4.5 algorithm, sublime text applications are used to make programming with HTML, PHP, CSS programming and use the Mysql database as data storage media.

5. References

[1] Destiningrum, M., & Adrian, Q. J. (2017). Sistem Informasi Penjadwalan Dokter Berbassis Web Dengan Menggunakan Framework Codeigniter (Studi Kasus: Rumah Sakit Yukum Medical Centre). Jurnal Teknoinfo, 11(2), 30. https://doi.org/10.33365/jti.v11i2.24

[2] Desyanti, D.-. (2019). Penerapan Data Mining Algoritma C4.5 untuk Mengetahui Tingkat Kepuasan Konsumen di Hotel Grand Zuri Dumai. SATIN - Sains Dan Teknologi Informasi, 4(2), 36. https://doi.org/10.33372/stn.v4i2.403

[3] Dimas Bayu Febriyanto. (2018). Implementasi Algoritma C 4.5 Untuk Klasifikasi Tingkat Kepuasan Pembeli Online Shop. Jurnal Riset Komputer, 5(6), 569–575.

[4] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). Jurnal Media Infotama, 11(2), 130–138.

[5] Heriyanto, Y. (2018). Perancangan Sistem Informasi Rental Mobil Berbasis Web Pada PT.APM Rent Car. Jurnal Intra-Tech, 2(2), 64–77.

[6] Joko Suntoro (2019) Data mining algoritma dan implementasi dengan pemrograman PHP

[7] Jubilee, Enterprise (2018). HTML, PHP & MySql untuk pemula

[8] Nuraini, R. (2015). Desain Algorithma Operasi Perkalian Matriks Menggunakan Metode Flowchart. Jurnal Teknik Komputer Amik Bsi, 1(1), 144–151.

[9] P, G. R., Windarto, A. P., Irawan, E., & Saputra, W. (2020). Penerapan Data Mining Menggunakan Algoritma C 4.5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS. Prosiding Seminar Riset Dan Information Sains (SENARIS), 2, 376–385.

[10] Pahlevi, O., Mulyani, A., & Khoir, M. (2018). Sistem Informasi Inventori Barang Menggunakan Metode Object Oriented Di Pt. Livaza Teknologi Indonesia Jakarta. Jurnal PROSISKO, 5(1). https://livaza.com/.

[11] Rahman, A. A. (2017). Penerapan Pendekatan Realistic Mathematic Education (RME) Pada Materi Statistika Untuk Meningkatkan Pemahaman Konsep Dan Prestasi. Genta Mulia, VIII (2), 1–12.

[12] Riyadli, H., Arliyana, A., & Saputra, F. E. (2020). Rancang Bangun Sistem Informasi Keuangan Berbasis WEB. Jurnal Sains Komputer Dan Teknologi Informasi, 3(1), 98–103. https://doi.org/10.33084/jsakti.v3i1.1770

[13] Standsyah, R. E., & Restu, I. S. (2017). Implementasi Phpmyadmin Pada Rancangan Sistem Pengadministrasian. Jurnal UJMC, Volume 3, Nomor 2, Hal. 38 - 44, 3, 38–44. http://e-jurnal.unisda.ac.id/index.php/ujmc/article/download/467/251