International Journal of Information System & Technology
Akreditasi No. 36/E/KPT/2019 | Vol. 3, No. 2, (2020), pp:323-331

The Application of C4.5 Algorithm to Classify the User Satisfaction of Online Learning System

Didi Supriyadi[1], S. Thya Safitri*[2]

[1,2] *S1 Information System, Institut Teknologi Telkom Purwokerto*

e-mail: [1]didisupriyadi@ittelkom-pwt.ac.id, [2]sisil@ittelkom-pwt.ac.id

Abstract

Telkom Institute of Technology Purwokerto (ITTP) is one of the universities that always strives to improve the quality of services to stakeholders. One of the services provided is an online learning system known as e-learning. Currently, ITTP does not have an obvious measurement to determine the level of user satisfaction of the e-learning system. System user satisfaction is determined by the quality of the system. Therefore, the purpose of this research is to determine the user satisfaction of the ITTP e-learning system by referring to the attributes of the success information system model developed by DeLone & McLean. Attributes used include the Ease of use, Response time, Reliability, Flexibility, and Security. Respondent data related to the assessment of the e-learning system were further classified by applying the C4.5 Decision Tree algorithm. Based on the results of calculations and tests using Rapid Miner software, it is known that the classification of users who have the SATISFIED classification is 46 respondents and 27 respondents are DISSATISFIED. Also, it is known that the flexibility of the ITTP e-learning system is the main determinant indicator of user satisfaction followed by response time.

Keywords: *C4.5 algorithm, user satisfaction, e-learning*

1. Introduction

During the Industrial Revolution 4.0 era, students were learners who were familiar with digital media. The educational paradigm is changing rapidly due to the adoption of technology among a new generation of students [1]. Even in recent years, investment in online learning systems has occurred significantly. This investment should meet the needs of users of many online learning systems, or it will be poorly adapted, discontinued, abandoned, or even rejected by its main users, which is students [2]. The individual's level of technology acceptance has been widely researched in various literature that shows the intention to use technology is measured from user perceptions of technology attributes such as usefulness and ease of use [3].

An online learning system is often known as online learning, or electronic learning (e-learning) is a digital-based learning technology device such as desktop computers, laptops, tablets, and smartphones[4]. Telkom Institute of Technology Purwokerto, one of the higher education institutions in Indonesia focused on technology has also developed an e-learning system developed by the Information Systems Unit to facilitate technology-based learning. The investment provided is expected to meet the needs and expectations of each stakeholder. Various advantages are provided by online-learning such as ease and flexibility [5].

Telkom Institute of Technology Purwokerto is committed to continuously improving the quality of services in general to every stakeholder, including a commitment to improving the quality of online learning systems. However, recently ITTP has not been able to measure how far the satisfaction level of the stakeholder towards the quality of the e-learning system services has been given. System quality is an intensity that users can feel, that the system is easy to operate, connect, and learn, and also enjoy when used [6].

ISSN : 2580-7250

According to[7], service quality can be explained by some attributes such as Ease of Use (EU), Response Time (RT), Reliability (R), Flexibility (F), and Security(S).

The user's response to the high quality of the e-learning system in large amounts of data can be extracted and classified into useful information and knowledge. This knowledge is commonly known as Data Mining. Data mining is the most popular knowledge acquisition technique[8]. To be able to map user satisfaction of the ITTP e-learning system and to measure the user satisfaction, it is necessary to research the analysis of user satisfaction online learning system using the data mining approach with Decision Tree C 4.5 Algorithm. The Decision Tree algorithm is generally selected because the decision tree is the best algorithm for data classification purposes with a minimum error rate [9].

2. Research Methodology

2.1. Data Mining

Data mining is the process of performing automatic extraction and generating predictive information from large databases [10]. Tennis's approach to data mining consists of several things, such as machine learning, statistics, and database systems. The purpose of data mining is to dig into and know the knowledge of large databases, which is converted into a format that can be understood and provide information to humans. The discovery of knowledge and data mining by an organization is an important component for the organization in its strategic decision-making[8]. Some of the utilization of information and knowledge from data mining results is currently widely implemented in several sectors ranging from trade, economy, social, politics, and tourism. Examples include market analysis, fraud detection, and control of the production amount of a product and also other science explorations [11]. One of the previous studies related to data mining for Education is the application of the Decision Tree C4.5 Algorithm for Evaluation of Educational Scholarships [12]. In previous research, the Decision Tree C4.5 Algorithm was rated highly effective and accurate.

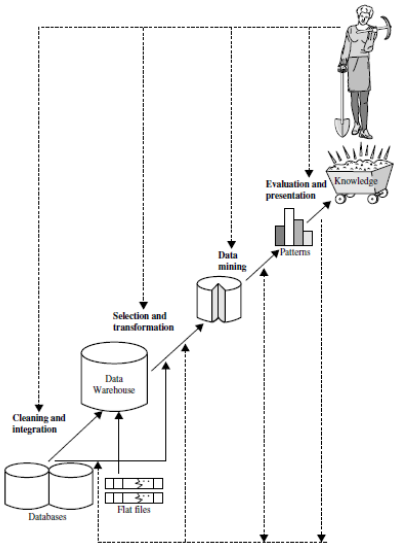


Figure 1. Data Mining as a Stage in the Knowledge Discovery Process[11]

Most people think of data mining as another more popular equation than Knowledge Discovery from Data or commonly abbreviated as KDD, and some others view data mining as a knowledge discovery process. Figure 1 illustrates the stages of the iterative knowledge discovery process as follows [11]:

a) Data cleaning, step to remove the noise and inconsistent data.
b) Data integration, step to combine the data from multiple data source.
c) Data selection, step to retrieve the relevant data from the database to be analysed.

d) Data transformation, step to transform and consolidate the data into the appropriate form of the data mining.
e) Data mining, step to extract the data pattern using the intelligent methods
f) Pattern Evaluation, the step to identify the most attractive pattern and represent the knowledge based on the interest measurement.
g) Knowledge presentation, step to present the knowledge by visualizing and representing the knowledge.

2.2. Decision Trees

The decision tree is a simple but powerful classification technique for analyzing multiple variables. The decision tree provides a unique ability to supplement, complement, and substitute a form of traditional statistical analysis (such as multiple linear regressions), various data mining techniques such as neural networks, and the most common is the development of a form of multidimensional analysis reporting on business intelligence[13]. The decision tree technique is generated from an algorithm that identifies how to split data into segments of a single branch as shown in Figure 2.

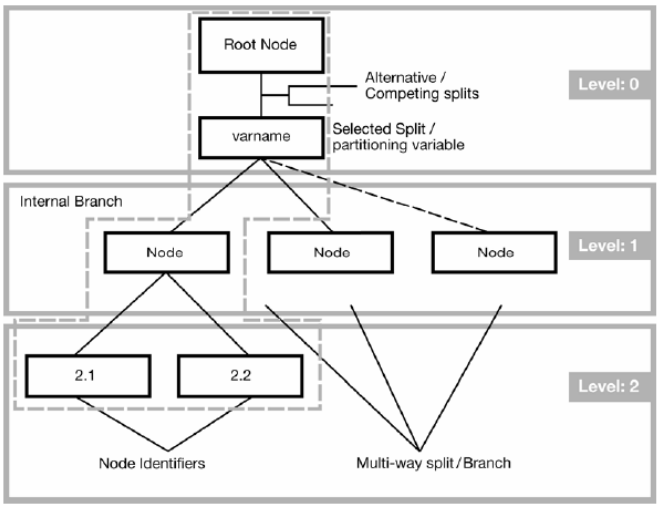


Figure 2. Decision Tree Illustration[13]

2.3. Algorithm C4.5

Decision Tree C4.5 algorithm is an algorithm that has been widely used in machine learning. This algorithm is a development of the ID3 algorithm introduced by J. Ross Quinlan in 1993[12]. In making decision trees with the C4.5 algorithm there are several stages, namely[14]:

a) Set up a training data set. The data can be data from responses from a system that has been grouped into specific classes. The data is the attributes and labels of event history.
b) Specifies the node of the decision tree. To determine which node to select, it must previously be calculated the entropy value and gain of each attribute, the attribute that has the highest gain value, will be the first node.

Entropy values are calculated by using the following formula:

$$Entropy\ (S) = \sum_{i=1}^{n} -Pi * Log2\ Pi \quad (1)$$

In which:
S = case set
n = number of S partition
Pi = Proportion S1 to S

Once each entropy value of each attribute is known, further is to calculate the gain

value using the following formula:

$$Gain(S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|si|}{|S|} * Entropy(Si) \quad (2)$$

Formula (2) is a formula commonly used in gain calculations after performing entropy calculations. Here's the description:

S = case set
n = number of A partition
|Si| = number of case on i-partition
|S| = number of case in s

By knowing the above formulas, the data that has been obtained can be entered and processed with the C4.5 algorithm for the decision tree creation process

The research stages performed are represented in Figure 3. Research stages are carried out from problem analysis, data collection, data testing by applying the C4.5 algorithm to the classification of user satisfaction of the ITTP e-learning system, and the last stage is to analyze and conclude the results of the study.

2.4. Problem Analysis

This stage is an early stage in research. This stage aims to identify research problems in research objects, literature review, and formulate problem-solving solutions by determining one of the methods of problem-solving.

2.5. Data Collection

After identifying problems and designing solutions from this research, the next stage is data collection. This stage begins by compiling a user satisfaction questionnaire for the ITTP e-learning system. The questionnaire refers to the attributes of the success of the information system developed by Delone & McLean. The attributes in question are Ease of Use, Response time, Flexibility, Security, and Reliability with the conclusion label is "SATISFIED" and "DISSATISFIED".

2.6. Testing the Data Using C4.5 Algorithm

Calculating the respondent's satisfaction assessment data on the ITTP e-learning system by implementing formulas 1 and 2 of the Decision Tree C4.5 algorithm by looking for entropy values and gain from each attribute. After manual calculation, the respondent's data was also tested using Rapid Miner software.

2.7. Result Analysis and Conclusion

The next stage is analyzing the user satisfaction data of the ITTP e-learning system. In addition to analyzing user satisfaction, the analysis is also carried out by referring to the calculation results between the manual and the results of the simulation using Rapid Miner software. After analyzing the data, the next step is concluding the results of the study.

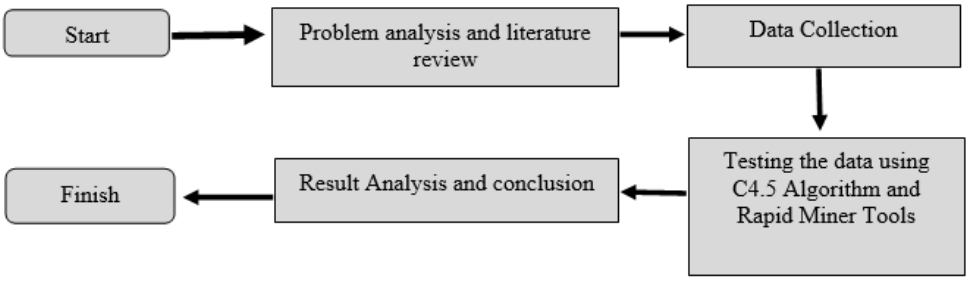


Figure 3. Research Stages

4. Result and Discussion

The sample data used are 73 respondents who are students from one of the major at Telkom Institute of Technology Purwokerto. The data was taken from 2019 to January 2020. The variables used contain Ease of Use (EU), Response Time (RT), Reliability (R), Flexibility (F), Security(S), and Satisfied (St) and Dissatisfied (DSt). As for the data obtained as shown in Table 1. Furthermore, the sample data was used to design the decision tree model.

Table 1. The user satisfaction data of the ITTP e-learning system

NO	EU	RT	R	F	S	RESULT
1	Low	Sufficient	Sufficient	Low	low	DSt
2	Sufficient	Sufficient	Sufficient	Sufficient	Sufficient	St
3	Low	enough	Sufficient	Low	Low	DSt
4	Low	Low	Low	Sufficient	Low	DSt
5	Sufficient	Sufficient	Sufficient	Sufficient	Low	St
6	Sufficient	Sufficient	high	Sufficient	Low	St
7	Sufficient	Sufficient	Low	Low	Low	DSt
8	High	Sufficient	Sufficient	Sufficient	High	St
9	Low	Sufficient	Sufficient	Low	Low	DSt
10	High	Sufficient	High	Sufficient	High	St
11	Low	Sufficient	Sufficient	Sufficient	Low	St
12	High	Sufficient	Sufficient	Sufficient	High	St
...	...	...	...	...	...	...
...	...	...	...	...	...	...
72	Sufficient	Sufficient	Sufficient	Sufficient	Sufficient	St
73	High	Low	Sufficient	Sufficient	Low	St

Table 2. The Calculation of entropy and gain on user satisfaction of the ITTP E-Learning system

NODE	SCALE	NUMBER OF CASES	DST	ST	ENTROPY	GAIN
TOTAL		73	27	46	0.95056685	
EU						0.280040784
	Low	19	16	3	0.62924922	
	Sufficient	38	10	28	0.83147439	
	High	16	1	15	0.33729007	
RT						0.165301878
	Low	14	11	3	0.74959526	
	Sufficient	53	16	37	0.88358509	
	High	6	0	6	0	
R						0.411505787
	Low	15	15	0	0	
	Sufficient	49	12	37	0.80309098	
	High	9	0	9	0	
F						0.633591851
	Low	22	22	0	0	
	Sufficient	48	5	43	0.48206615	
	High	3	0	3	0	
S						0.414292878
	Low	31	24	7	0.77062907	
	Sufficient	39	3	36	0.39124356	
	High	3	0	3	0	

From Table 2 above, it is known that the attribute with the highest gain is Flexibility (F) with a gain value of 0.633591851; the flexibility node is then used as the root node. Tree decisions resulting from the calculation of entropy values and gain on node 1 can be described in Figure 4.

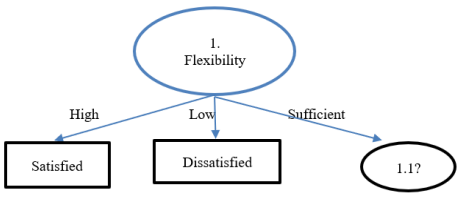


Figure 4. Decision Tree on Node 1

Flexibility attributes that are of sufficient value have not been able to indicate the value to be a decision, so it is necessary to calculate entropy and further gain on node 1.1.

Table 3. The Calculation on Node 1.1 based on Flexibility attribute

NODE	SCALE	NUMBER OF CASES	DST	ST	ENTROPY	GAIN
TOTAL		48	5	43	0.48206615	
EU						0.207647232
	Low	7	4	3	0.98522814	
	Sufficient	29	1	28	0.21639693	
	High	12	0	12	0	
RT						0.322993814
	Low	8	5	3	0.954434	
	Sufficient	36	0	36	0	
	High	4	0	4	0	
R						0.244963815
	Low	3	3	0	0	
	Sufficient	39	2	37	0.29181826	
	High	6	0	6	0	
S						0.129753773
	Low	11	4	7	0.9456603	
	Sufficient	34	1	33	0.19143325	
	High	3	0	3	0	

From Table 3 above, it is known that the attribute with the highest gain is Response Time (RT) with a gain value of 0.322993814; next node Response time is used as the root node. Tree decisions resulting from the calculation of entropy values and gain on node 1.1 can be described in Figure 5.

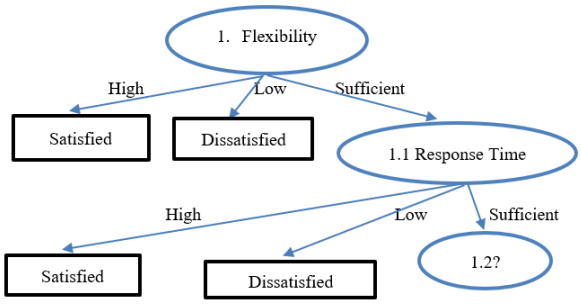


Figure 5. Decision Tree on Node 1.1

Response Time attribute that is of sufficient value has not been able to indicate the value into a decision, so it is necessary to calculate entropy and further gain on node 1.2.

Table 4. The Calculation on Node 1.2 based on Response Time attribute

NODE	SCALE	NUMBER OF CASES	DST	ST	ENTROPY	GAIN
TOTAL		8	5	3	0.954434	
EU						0.704434003
	Low	4	4	0	0	
	Sufficient	2	1	1	1	
	High	2	0	2	0	
R						0.347589881
	Low	3	3	0	0	
	Sufficient	5	2	3	0.97095059	
	High	0	0	0	0	
S						0.158868006
	Low	5	4	1	0.72192809	
	Sufficient	3	1	2	0.91829583	
	High	0	0	0	0	

Table 5. The Calculation on Node 1.3 based on Ease of Use Attribute

NODE	SCALE	NUMBER OF CASES	DST	ST	ENTROPY	GAIN
TOTAL		2	1	1	1	
R						1
	Low	1	1	0	0	
	Sufficient	1	0	1	0	
	High	0	0	0	0	
S						1
	Low	1	1	0	0	
	Sufficient	1	0	1	0	
	High	0	0	0	0	

The next step of the author uses Rapid Miner to form a Decision Tree. Based on the results of the data shown in Table 2, it is known that node F which is the Flexibility indicator of the ITTP e-learning system has the highest gain of 0.633591851. Node F is used as the root node of the decision tree.

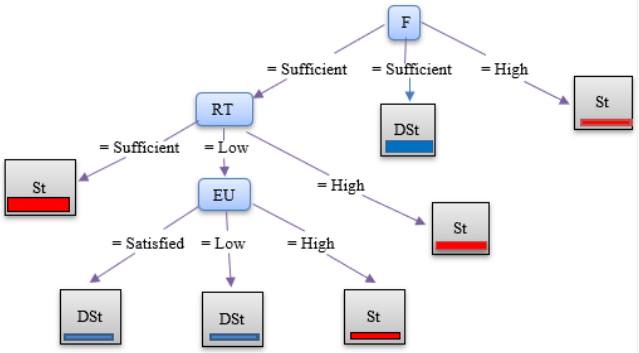


Figure 6. Decision Tree on User Satisfaction of the ITTP E-learning system

4. Conclusion

Based on the results of mathematical analysis and testing using Rapid Miner software can be concluded as follows:

a) The C4.5 Algorithm Decision Tree can be implemented to classify the user satisfaction of the ITTP e-learning system.
b) From 73 respondents of the ITTP e-learning system users, 46 or about 63% of respondents expressed "SATISFACTION" and 27, or about 37% of respondents stated, "DISSATISFIED". This result indicates that the system is still acceptable to the user.
c) There are 5 gain attributes that become the benchmark of the ITTP e-learning system users including Ease of use, Response time, Reliability, Flexibility, and Security attributes.
d) Besides, it is known that the flexibility of the ITTP e-learning system becomes the main indicator of the determinant of user satisfaction followed by response time.

Further research that can be done is to test the application of other methods in data mining such as ID3, Fuzzy, or others to classify user satisfaction with the ITTP e-learning system and other systems. This further research is intended to be able to identify the level of accuracy of the classification methods available in Data Mining.

Acknowledgment

The author thanked the Head of IT Support and Head of LPPM Section of the Telkom Institute of Technology Purwokerto, who has given permission and support to this research.

References

[1] L. Y. K. Wang, S. L. Lew, S. H. Lau, and M. C. Leow, "Usability factors predicting continuance of intention to use cloud e-learning application," *Heliyon*, vol. 5, no. 6, p. e01788, 2019, doi: 10.1016/j.heliyon.2019.e01788.

[2] A. Mehta, N. P. Morris, B. Swinnerton, and M. Homer, "The Influence of Values on E-learning Adoption," *Comput. Educ.*, vol. 141, no. July, 2019, doi: 10.1016/j.compedu.2019.103617.

[3] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Manage. Sci.*, vol. 35, no. 8, pp. 982–1003, 1989, doi: 10.1287/mnsc.35.8.982.

[4] R. C. Clark and R. E. Mayer, *E-Learning and the Science of Instruction Proven Guidelines for Consumers and Designers of Multimedia Learning*, 3rd ed. San Francisco: Jhon Wiley & Sons, 2008.

[5] O. Isaac, A. Aldholay, Z. Abdullah, and T. Ramayah, "Online learning usage within Yemeni higher education: The role of compatibility and task-technology fit as mediating variables in the IS success model," *Comput. Educ.*, vol. 136, no. November 2017, pp. 113–129, 2019, doi: 10.1016/j.compedu.2019.02.012.

[6] S. Petter and E. R. McLean, "A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level," *Inf. Manag.*, vol. 46, no. 3, pp. 159–166, 2009, doi: 10.1016/j.im.2008.12.006.

[7] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, 2003, doi: 10.1080/07421222.2003.11045748.

[8] S. Sharma, J. Agrawal, and S. Sharma, "Classification Through Machine Learning Technique: C4. 5 Algorithm based on Various Entropies," *Int. J. Comput. Appl.*, vol. 82, no. 16, pp. 28–32, 2013, doi: 10.5120/14249-2444.

[9] A. Viloria *et al.*, "Determinating student interactions in a virtual learning environment using data mining," *Procedia Comput. Sci.*, vol. 155, no. 2018, pp. 587–592, 2019, doi: 10.1016/j.procs.2019.08.082.

[10] J. Havrda and F. Charvat, "Quantification Method of Classification Processes. Concept of Structural a Entropy," *Kybernetika*, pp. 1–4, 1967.

[11] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. USA: Elsevier Inc., 2012.

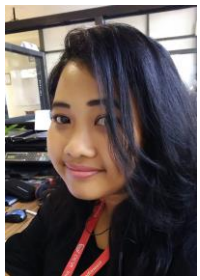
[12] X. Wang, C. Zhou, and X. Xu, "Application of C4.5 decision tree for scholarship evaluations," *Procedia Comput. Sci.*, vol. 00, no. 2018, 2019, doi: 10.1016/j.procs.2019.04.027.

[13] B. DeVille, *Decision Trees for Business Intelligence and Data Mining Using SAS Enterprise Miner*. USA: SAS Institute Inc., 2006.

[14] D. Ardiansyah and W. Walim, "Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner," *Infokar*, vol. 1, no. 2, pp. 5–12, 2018.

Authors

1st Author
Didi Supriyadi
I'm a lecturer of Information System Study Program at Institut Teknologi Telkom Purwokerto.



2nd Author
S. Thya Safitri
I'm a lecturer of Information System Study Program at Institut Teknologi Telkom Purwokerto.