Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2022. Vol. 1. No. 3

Student Satisfaction Level Analysis Of Online Learning During Pandemic Covid 19 Using C5.0 Algorithm

James Hasudungan Sihombing[1]*, Poningsih[2], Irawan[3], Indra Gunawan[4], Zulaini Masruro Nasution[5]

[1,3,4,5]STIKOM Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia
[2]AMIK Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia
**jamessbg20@gmail.com*

Abstract

The Simalungun University Foundation Tourism Vocational School is a private school located in Pematangsiantar. At this time there are problems in the learning process. And currently the data collection used by researchers in obtaining student satisfaction data is by sampling. In this study, data were obtained from giving questionnaires to students of the Simalungun University Foundation Tourism Vocational School which were categorized using five variables, namely teaching methods, learning media, communication, teaching materials/modules, and learning duration. Data Mining using the C5.0 algorithm is proven in analyzing satisfaction in the teaching and learning process. Research result able to find out the results of the analysis of the level of student satisfaction with online learning and find the highest score, can help the Tourism Vocational School of the Simalungun University Pematangsiantar Foundation in optimizing online learning today.

Keywords: *Data Mining, C5.0 Algorithm, Analysis, Online Learning*

1. Introduction

The Simalungun University Foundation Tourism Vocational School is a Private Vocational High School located in North Sumatra Province, Pematangsiantar City with the address West Sisingamangaraja Street, Bah Kapul Village, Siantar Sitalasari District Postal Code 21139 Pematangsiantar. Currently there are 3 majors at the USI Pematangsiantar Foundation Tourism Vocational School, Travel Business, Hospitality Accommodation, Catering. Research related to the analysis of the level of satisfaction has been carried out by [1], [2] namely the Implementation of the C5.0 Algorithm in the Classification of Community Income, the test that has been done, it can be concluded that data mining uses the C5.0 algorithm decision tree method to form a decision tree in classifying people's income using the Weka application and get maximum results[3]. Research related to the analysis of satisfaction levels has been carried out by[4] namely the Analysis of Household Welfare Using the C5.0 Method[5]. From the Wonosobo district SUSENAS data, four classifications were obtained using the C5.0 algorithm. The results obtained are expected to be useful for the government to determine the status of families who deserve assistance.

Research related to the analysis of satisfaction levels has been carried out by namely Predicting the timeliness of student graduation using the C.50 algorithm method[6],[7],[8]. The test results on the test data obtained a recall value of 72.22%, a precision value of 92.85%, an accuracy value of 83.78%, and an error value of 16.21%. Online learning is the right choice during a pandemic like this, but has several obstacles in the online teaching and learning process , namely the limited package or network, the learning time is not right, and the location of their residence is quite far, making students sometimes do not understand the learning provided by the teacher is due to more tasks than learning material.

In this study, the C5.0 Algorithm Method will be used with the application of this method[9],[10], the author can determine how satisfied or dissatisfied students are in carrying out the online teaching and learning process during the Covid 19 pandemic [11]. That way the school will take a policy or action in the offline or online learning process by dividing the time for class entry at the Simalungun University Foundation Tourism Vocational School.

2. Research methodology

The research method provides an overview of the process or procedure used by the author to solve the problem. The method used in this study is data mining using the C5.0 Algorithm Method. This study was conducted to determine the level of student satisfaction with online learning during the pandemic. The data collection used by researchers in obtaining student satisfaction data is by sampling. Sampling is the collection of data that partially or represents all the characteristics of the population. The data obtained by the author in this study is the result of a questionnaire conducted by the author. The results obtained in the research conducted by the author can help the school take policies or actions in the offline or online learning process at the USI Foundation Tourism Vocational School. The following is a Research Work Activity Diagram:

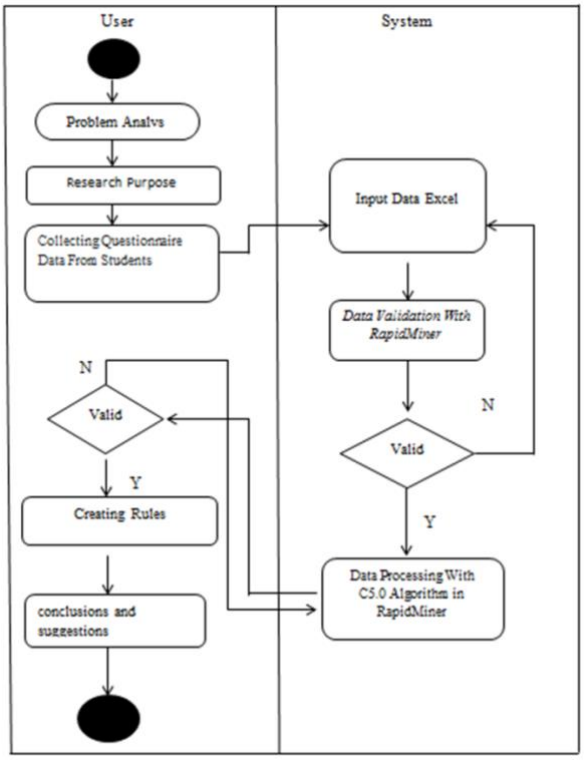

Figure 1. Research Activity Diagram

The method used in this study is one of the classification data mining methods , namely the C5.0 algorithm. In the C5.0 algorithm performed on a sample of data sets that can classify a data effectively. The following are the completion steps taken by the author in determining student satisfaction in online learning e.g: Determine the data to be used in the process of analyzing student satisfaction with online learning. Data obtained from questionnaires given to students, Calculate satisfied and dissatisfied from the data obtained. Furthermore, the number of satisfied and dissatisfied is divided by the number of data, Calculate the number of satisfied and dissatisfied for each criterion based on each data obtained. Next determine the probability of being satisfied and dissatisfied for each criterion used, Calculate one of the variables used in determining student satisfaction to get the value of satisfied and dissatisfied students, In comparing the values of satisfied and dissatisfied by looking at the high accuracy value of the values that have been obtained. Provide conclusions from the data that has been analyzed. In this study the authors use quantitative research methods, the data used will be processed from the results of the questionnaire given to students. In this study, the authors obtained data from the results of the student satisfaction survey with a total of 100 respondents. After the necessary data is obtained then the variables that will be used in the study are determined. The variables used include How to Teach, Learning Media, Communication, Teaching Materials/Modules/Assignments and Study Duration. While the response variable is satisfaction, namely people who are satisfied and dissatisfied with online learning at the Tourism Vocational School of the USI Pematangsiantar Foundation. Each aspect consists of several questions posed to students. The following is data obtained from the Student Satisfaction Survey.

Table 1. Student Questionnaire Results Table

Respondent	C1 (Teaching Method)	C2 (Learning Media)	C3 (Communication)	C4 (Teaching Materials)	C5 (Learning Duration)
1	2	3	2	3	1
2	3	3	3	3	2
3	3	3	3	3	2
4	2	3	2	2	2
5	2	3	2	2	2
6	2	2	2	2	2
7	3	3	3	3	3
8	3	3	3	3	3
9	3	3	3	3	3
10	3	3	3	3	3

…	….	….	….	….	….
100	3	3	3	3	3

To get the results of the research conducted, the following is a description of the manual calculation of satisfied and dissatisfied using the C5.0 algorithm. In this case the data variables that you want to classify are Teaching Methods, Learning Media, Communication, Teaching Materials/Modules/Assignments, Learning Duration. The number of questionnaire data samples that will be used in this study is 100 data. The following is a sample table of case calculations. The calculation of the attribute information gain value from the total of all cases can use the formula:

$$I(S_1, S_2, \dots, S_m) = -\sum_{i=1}^{m} - pi * log_2\,(Pi) \text{ ……..1)}$$

Next is to find the entropy of each attribute by using the equation:

$$(A) = \sum_{j=1}^{y} \frac{S_{1j} + \cdots + S_{mj}}{S} I(S_{1j} \cdots S_{mj}) \text{ ……..2)}$$

The uncertainty or the impurity is represented as the log to base 2 of the probability of a category (pi). The index (i) refers to the number of possible categories. Here, i = 2 as our problem is a binary classification.

3. Results And Discussion

3.1. The Calculation of the C5.0 Algorithm

The calculation of the C5.0 algorithm begins by selecting the root attribute first by looking for the total number of cases, the number of difficult decision cases and the number of easy decision cases. Calculating the entropy of all cases divided by Teaching Method, Learning Media, Facilities, Communication, Teaching Materials/Modules, and Learning Duration. After that, the gain is calculated for each attribute. The calculation results are shown in the following table.

Table 2. Node 1

Node 1		Amount	Satisfied	Not satisfied	Entrophy	Gain
Total		100	72	28	0.855451	
How to Teach						0.441818
	Well	58	56	2	0.216397	
	Enough	21	14	7	0.918296	
	Not enough	21	2	19	0.453716	
Learning Media						0.526433
	Well	53	52	1	0.135036	
	Enough	25	19	6	0.795040	
	Not enough	22	1	21	0.266765	
Communication						0.498787
	Well	55	55	0	0.000000	
	Enough	16	11	5	0.896038	
	Not enough	29	6	23	0.735509	
Teaching Materials/Modules						0.52346
	Well	55	55	0	0.000000	
	Enough	19	13	6	0.899744	
	Not enough	26	4	22	0.619382	
Study Duration						0.713258
	Well	58	58	0	0.000000	
	Enough	11	11	0	0.000000	
	Not enough	31	3	28	0.458686	

The entrophy column in the total row is calculated by equation (1) as follows:

Entrophy (total) = - (72/100)*LOG2(72/100) - (28/100)*LOG2(28/100)
Entrophy (total) = 0.855451

Then on the gain value in the Learning Duration row, equation (2) is calculated as follows:
Gain (total,Learning Duration) = (0.85541) - ((58/100)*0 + (11/100)*0 + (32/150)*0.458686
Gain (total,Learning Duration) = 0.713258

From the results of the calculations in table 1, it is obtained that the attribute that becomes the node (root) is Learning Duration which has the highest gain of 0.713258 which consists of 3 sub-attributes namely Good, Enough and Less. Based on the entrophy values of the three sub-attributes above, the sub-attributes are Good, Enough, which has received a decision, while Less has not yet received a decision. Then it is necessary to do further calculations to determine the next root node. Then perform calculations on the branch nodes of the attributes Learning Duration = Less, Teaching Method = Less, Learning Media = Less and Communication = Less can be shown in table 3 below:

Table 3. Node Calculation

Less Duration, Less Way to Teach	**Amount**	**Satisfied**	**Not satisfied**	**Entrophy**	**Gain**
Less Learning Media,	17	1	16	0.322757	
&Less Communication					0
Well	0	0	0	0.000000	
Enough	0	0	0	0.000000	
Not enough	17	1	16	0.322757	

From the results of the calculations in table 4.5, the attributes that become branch nodes of Less Duration, Less Teaching Method = Less Communication Learning Media = Less where the sub attributes are Good, Enough and Less then all attributes have obtained a decision so that the calculation is complete. Then the decision tree can be described from the table above as follows:

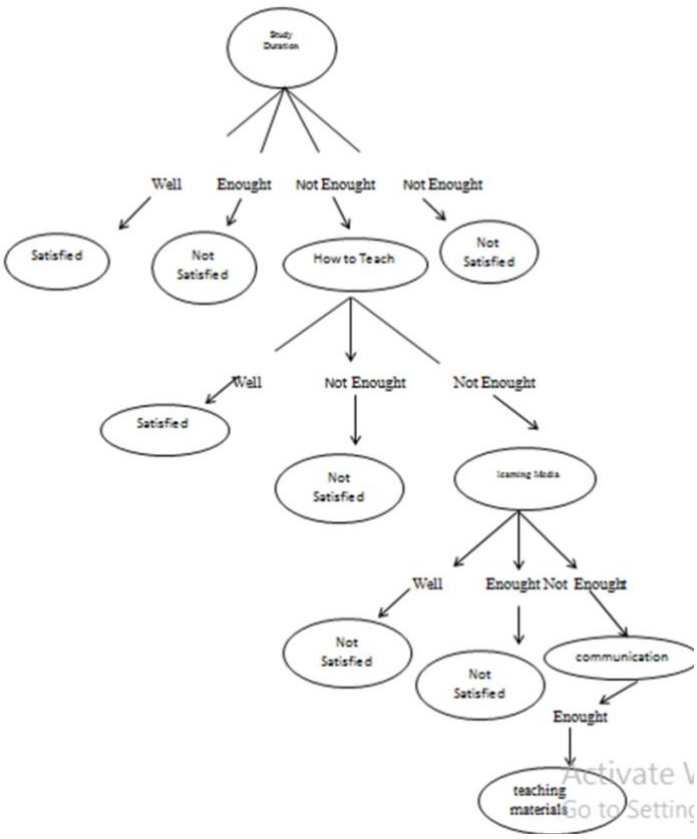

Figure 2. Decision Tree level 5

3.2. Testing Process with Rapidminer

The results of data processing using rapidminer tools can be seen in the following figure 3:

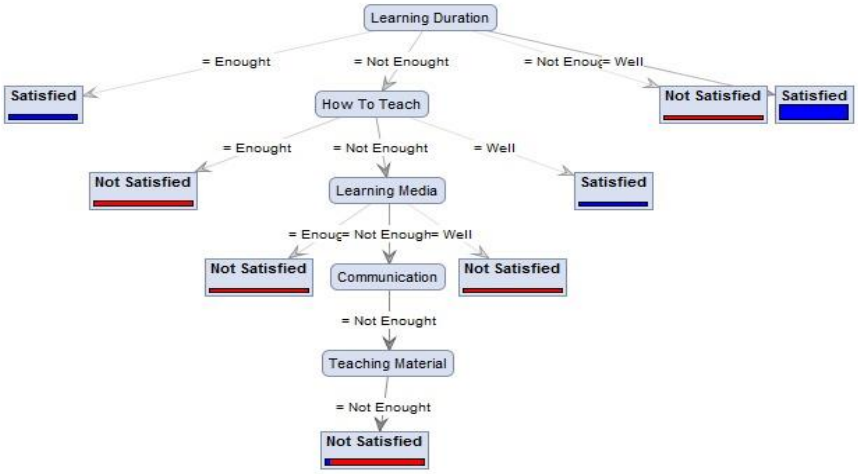

Figure 3. Result of decision tree

After calculating and testing the data on each attribute with the C5.0 algorithm, the final decision tree pattern will be obtained.

Tree

```
Learning Duration = Enought : Satisfied {Satisfied=11, Not Satisfied=0}
Learning Duration = Not Enought
|   How To Teach = Enought : Not Satisfied {Satisfied=0, Not Satisfied=7}
|   How To Teach = Not Enought
|   |   Learning Media = Enought : Not Satisfied {Satisfied=0, Not Satisfied=4}
|   |   Learning Media = Not Enought
|   |   |   Communication = Not Enought
|   |   |   |   Teaching Material = Not Enought: Not Satisfied {Satisfied=1, Not Satisfied=14}
|   |   Learning Media = Well: Not Satisfied {Satisfied=0, Not Satisfied=1}
|   How To Teach = Well: Satisfied {Satisfied=2, Not Satisfied=0}
Learning Duration = Not Enought : Not Satisfied {Satisfied=0, Not Satisfied=2}
Learning Duration = Well: Satisfied {Satisfied=58, Not Satisfied=0}
```

Figure 4. Description of decision tree

The picture above shows the results of a complete description of the decision tree that has been formed using the C5.0 algorithm. From the description also shows that the use of data mining algorithms C5.0 well used in the process of digging the data to draw some conclusions are visualized with trees decision (decision tree) . The following rules are generated from the decision tree:

Table 4. Rules

Rules	Decision
If Learning Duration = Good	Satisfied
If Learning Duration = Enough	Satisfied
Less Duration & Less Teaching Method	Satisfied
Less Duration , Less Teaching Method, Learning Media	Satisfied
Less Duration , Less Teaching Method, Learning Media, Communication = Less	Not satisfied
Less Duration , Less Teaching Method, Less Learning Media, Communication = Less and Teaching Material = Less	Not satisfied

4. Conclusion

Data Mining using the C5.0 Algortima is proven to be able to find out the results of the analysis of the level of student satisfaction with online learning and find the highest score. The results of the trial using the Rapidminer 5.3 software as a tool to prove that the manual data calculation results are the same as the tests carried out using the Rapidminer 5.3 software.

Acknowledgement

Acknowledgments to the supervisors and examiners who are lecturers at AMIK and STIKOM Tunas Bangsa so that this research can be arranged as one of the requirements for completing Bachelor's education (S1) at STIKOM Tunas Bangsa. I hope this research can be a reference for other research related to the methods and algorithms used. I hope for constructive suggestions for the readers for the perfection of this research in the future.

References

[1] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "IMPLEMENTASI ALGORITMA C5 . 0 DALAM KLASIFIKASI PENDAPATAN MASYARAKAT (STUDI KASUS : KELURAHAN MESJID KECAMATAN MEDAN KOTA)," vol. 5, 2018.

[2] R. Pandya and J. Pandya, "C5. 0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning," *Int. J. Comput. Appl.*, vol. 117, pp. 18–21, May 2015, doi: 10.5120/20639-3318.

[3] P. R. Pandurangan, "A Comparative Study of Data Mining Algorithms for Decision Tree Approaches using WEKA Tool," *Adv. Nat. Appl. Sci.*, vol. 11, pp. 230–241, Jul. 2017.

[4] A. Wisri, Y. Susanti, I. Slamet, and K. Kunci, "Analisa Kesejahteraan Rumah Tangga Menggunakan Metode C5 . 0," vol. 3, no. 1, pp. 262–267, 2020.

[5] W. Paper, "Indicators : Experiments Using Predicting Consumption Poverty Using Non-consumption Indicators : Experiments Using Indonesian Data," no. February, 2006.

[6] A. Fitrani, "Prediction of Study Period Students (Bachelor Degree) Muhammadiyah University of Sidoarjo Based on Decision Tree Method using C4.5 Algorithm," *J. Phys. Conf. Ser.*, vol. 1179, p. 12033, Jul. 2019, doi: 10.1088/1742-6596/1179/1/012033.

[7] D. Solomon, "Predicting Performance and Potential Difficulties of University Student using Classification : Survey Paper," vol. 118, no. 18, pp. 2703–2707, 2018.

[8] N. Benediktus and R. S. Oetama, "Algoritma Klasifikasi Decision Tree C5 . 0 untuk Memprediksi Performa Akademik Siswa," vol. XII, no. 1, pp. 14–19, 2020.

[9] D. F. Hardani and Y. P. Purba, "Application of the C5 . 0 Algorithm to Determine the Level of Public Satisfaction with the E-KTP Recording Service at the Bandar Sub-District Office," vol. 1, no. 1, 2021.

[10] M. Fajri, A. Ramadhan, and R. Hidayatullah, "Original article Implementation of boosting on the C5 . 0 algorithm in the health development index data," vol. 10, no. 1, pp. 12–15, 2021, doi: 10.22487/25411969.2021.v10.i1.15466.

[11] A. M. H. Pardede, H. Mawengkang, M. Zarlis, and T. Tulus, "Pemanfaatan Sumber Daya Kesehatan Terbatas Sebagai Pendukung Perawatan Kesehatan Cerdas," *Pros. Semin. Nas. Ris. Inf. Sci.*, 2019, doi: 10.30645/senaris.v1i0.128.