The IJICS (International Journal of Informatics and Computer Science)
Vol 5 No 2, July 2021, Page 118-122
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v5i2.3129

Analysis of Online Learning Understanding Patterns at Budi Darma University Using the C5.0 Algorithm

Firman Telaumbanua, Johanes Mario Purba, Dito Putro Utomo*

Fakultas Ilmu Komputer dan Teknologi Informasi, Prodi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: [1]firmantelaumbanua96@gmail.com, [2]johanes@gmail.com, [3,*]ditoputro12@gmail.com
Coressponding author: ditoputro12@gmail.com
Submitted: **06/07/2021**; Accepted: **15/07/2021**; Published: **31/07/2021**

Abstract−Online learning is a learning process that is carried out without face to face, but at a distance that utilizes communication and information technology that utilizes the internet. The emergence of the online learning process provides considerable benefits, both for the manager of the information system and for students and students. Currently, the development of online learning is becoming popular among schools and universities. This is caused by the corona virus or what is often called Covid-19. Covid-19 is one of the viruses that first emerged and developed in China, especially in the city of Wuhan. Initially this virus only infects a few people, but in fact this virus can move places or can be transmitted through direct contact with the sufferer. This virus has infected the whole world so this is an international problem that must be addressed. Various efforts have been made, one of which is in the field of education. The closure of schools and colleges is an effort made to break the chain of development of COVID-19, so that the teaching and learning process is carried out online. In this study, an analysis of student satisfaction assessment will be carried out on the online learning process using the existing methods in data mining. Data mining has several algorithms that are used to solve these problems, one of which is the C5.0 algorithm. The result of this study is to find out the results of the satisfaction assessment of online learning during the Covid-19 pandemic.

Keywords: Data Mining; Corona Virus; Online Learning; Algorithm C5.0

1. INTRODUCTION

The emergence of the corona virus or Covid-19 has a negative impact on people's lives, especially in Indonesia. The corona virus has caused enormous losses in various fields, one of which is education. Covid-19 is a virus that is transmitted from animals to humans[1]. However, this is still under investigation regarding Covid-19. The learning process is usually carried out face-to-face in the classroom, but due to Covid-19 it is carried out online. Changes in the learning system carried out online or more precisely online learning at colleges and schools are temporarily a good solution to stop the spread of the corona virus. Online is a learning process that is carried out without face to face between students and lecturers, but it is done online using several media platforms. One of the universities that conduct online learning process is Budi Darma University.

Budi Darma University is a higher education institution located in the city of Medan. The learning system at Budi Darma University was initially carried out directly or face to face in the classroom. However, in the last 4 months, the learning system has changed. Changes in the learning system are carried out based on regulations issued by the national education minister. The application of the online learning system is carried out until the situation is supportive and Covid-19 is successfully handled. However, the problem that must be known is whether online learning is carried out efficiently in the teaching and learning process. In this study, it will be discussed and analyzed about the application of online learning to students. Whether or not online learning is carried out can be seen from the responses given by students, where students understand or not the lecture material delivered can be seen from several variables such as the presence or absence of a material module provided by the lecturer, the explanation in the form of a video made by the lecturer. either directly or through other media, there are questions and answers that are carried out during online lectures and assignments or exercises given by the supporting lecturer.

Based on previous research conducted by Eko Kuntarto and his friends about online lectures or more precisely E-learning, said that e-learning is very effective in certain learning processes such as Indonesian language courses[2]. Then in the research conducted by Kuwat Sentiyanto on laboratory practicum based on E-learning, concluded that the use of E-learning is effective so that students can easily do the exercises given and the assessment is faster in the learning process for certain needs. Subsequent research also concluded the same thing by Warsun Najib, that E-learning is more practical and efficient in subjects that require online learning, one of which is Computer Networking courses[3]. The problem is how to use online services that have been carried out for the past 4 months in all courses as an effort to break the chain of Covid-19 developments at Budi Darma University. Is it effective and efficient and can students understand the material delivered online.

These problems can be solved by using data mining. Data mining is a process that uses techniques such as statistical, mathematical, artificial intelligence, and machine learning techniques to identify those that can be useful. Data mining has several algorithms that can be used to solve problems related to data[4][5]. one of the algorithms that can be used to solve the above problem is the C5.0 algorithm. The C5.0 algorithm is an improvement algorithm from the C4.5 algorithm where the process is almost the same, it's just that the C5.0 algorithm has advantages over the previous algorithm. With this algorithm, the assessment of the application of the online learning system is analyzed[6][7].

2. RESEARCH METHODOLOGY

2.1 Research Methodology

In this study, researchers took several approaches to data collection carried out at the Budi Darma University environment in understanding the online learning process, in a study data is needed as an object to be developed for the system to be built.

a. Identification of Problems
   At this stage identify problems that may be faced in this research, the problem faced is whether online learning or the system in learning has been effective so that the material presented can be understood by students.
b. Literature Review
   In conducting this research, the researcher cannot do research without a literature review of what is concerned in the research, the authors collect data by looking at research sites or trusted journals (Internet), books.
c. Data Collection
   Collecting data by distributing online questionnaires to Budi Darma University students.
d. Analysis and Discussion
   At this stage an analysis of the pattern of understanding online learning at Budi Darma University uses the C5.0 Algorithm
e. Conclusion and Results
   After passing through the stages of analysis and discussion, the next step is drawing conclusions from the results obtained in the study.

2.2 Data Mining

Data mining is an attempt to dig up valuable and useful information on a very large database[8]. Data Mining is also known as Knowledge Discovery in Database (KDD) which can be interpreted as the extraction of unknown implicit potential information from a set of data. The KDD process involves the results of the data mining process or the process of extracting the tendency of a data pattern, then the results are converted appropriately into information something that is easier to understand[9][10][11].

2.3 Online Learning

Online learning is a learning process that is carried out without face-to-face as well as in a classroom, but is carried out online through the help of certain services[2]. The benefits of online learning are that it becomes a means of learning that is carried out online, facilitates interaction between lecturers and students and becomes a research tool.

2.4 Algorithm C5.0

The C5.0 algorithm is one of the algorithms which is a refinement of the C4.5 algorithm which uses a tree-shaped representation where each node represents an attribute and then a branch represents the value of the attribute and has a leaf where the function is a class[8][12]. Decision making is based on the largest Gain value from the calculation of all attributes. Here are the steps for using the C5.0 algorithm[12][13]:

1. Make a decision system that includes condition attributes and decision attributes. Then describe a decision system consisting of only n objects
2. Counting the number of column data, where the amount of data must be based on certain attribute members whose results are based on certain conditions.
3. Choose the attribute that is used as the Node.
4. Create a branch for each member of the Node.
5. Checks whether the entropy value of each Node member has a value of zero. If the value is 0, then determine the leaf that has been formed. If the entropy value of each Node member is entirely zero, then the process stops.

Node members that have a value greater than zero, then the process is repeated from the beginning with the condition that all members of the Node are zero. Node is an attribute that has the highest gain value of the existing attributes. The process of calculating the gain value of an attribute must use a formula. The following is the formula for calculating the gain value used in the C4.5 algorithm:

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

The explanation of the above formula is:
S = Case Set
A = Attribute
n = Number of attribute partitions A
|Si| = Proportion of Si to S
|S| = Number of cases in S
In addition, the formula for calculating the entropy value is as follows:

$$\text{Entropy(S)} = \sum_{i=1}^{n} - pi + log_2 \text{ pi} \quad (2)$$

The explanation of the formula above is:

S= Case Set
n = Number of partitions S
Pi = Proportion of Si to S

3. RESULT AND DISCUSSION

The sample that will be used to assess how much student satisfaction is in the online learning process is the daily data of students during the online learning process. Assessment of student satisfaction is grouped into several criteria. The following are the criteria that will be used to analyze the satisfaction of Budi Darma University students with online learning services during the Covid-19 pandemic:

Table 1. Student Satisfaction Data

No	Respondents	V1	V2	V3	V4	Results
1	R1	Yes	Yes	Yes	Often	Understand
2	R2	Yes	No	Yes	Rare	Not Understand
3	R3	No	No	No	No	Not Understand
4	R4	Yes	Yes	No	Rare	Understand
5	R5	No	No	Yes	No	Not Understand
6	R6	No	No	No	No	Not Understand
7	R7	No	No	No	Often	Not Understand
8	R8	Yes	Yes	Yes	Rare	Understand
9	R9	No	No	Yes	Rare	Not Understand
10	R10	Yes	Yes	Yes	No	Understand
11	R11	Yes	No	No	Rare	Understand
12	R12	No	Yes	Yes	No	Not Understand
13	R13	Yes	Yes	No	Often	Not Understand
14	R14	Yes	No	No	No	Understand
15	R15	No	Yes	Yes	Rare	Not Understand

Then calculate the number of cases, the number of cases for the Satisfied result, the number of cases for the Dissatisfied result, and the entropy of all cases and the cases are divided based on the attributes of Discipline Variables, Teaching and Learning Quality Variables, Attendance Variables, and Behavioral Variables. The next step, yes, the gain is calculated for each attribute.

1. Entropy Value
   a. Total Entropy = Entropy (S) = $\sum - n\ i{=}l$ pi*$\log_2$ pi
      Total Entropy = ((-6/15 * log 2 (6/15) + (-9/15 * log 2 (9/15)) = 0.9709506
   b. Entropy V1
      Attribute Value "Yes" = ((-6/8*log 2 (4/8)) + (-2/8 * log 2 (2/8) = 0.8112781
      Attribute Value "Nothing" = ((-7/7*log 2 (7/7)) + (-0/7 * log 2 (0/7) = 0
   c. Entropy V2
      Attribute Value "Yes" = ((-4/7*log 2 (4/7)) + (-3/7 * log 2 (3/7) = 0.9852281
      Attribute Value "Nothing" = ((-6/8*log 2 (6/8)) + (-2/8 * log 2 (2/8) = 0.8112781
   d. Entropy V3
      Attribute Value "Yes" = ((-3/8*log 2 (3/8)) + (-5/8 * log 2 (5/8) = 0.954434
      Attribute Value "Nothing" = ((-3/7*log 2 (3/7)) + (-4/7 * log 2 (4/7) = 0.9852281
   e. Entropy V3
      Attribute Value "Often" = ((-1/3*log 2 (1/3)) + (-2/3 * log 2 (2/3) = 0.9182958
      Attribute Value "Rare" = ((-3/6*log 2 (3/6)) + (-3/6 * log 2 (3/6) = 1
      Attribute Value "Nothing" = ((-2/6*log 2 (2/6)) + (-4/6 * log 2 (4/6) = 0.9182958
2. Then look for the value of Gain
   a. Gain (Total V1) = 0.9709506- ((8/15 * 0.8112781) + (7/15 * 0)) = 0.5382689
   b. Gain (Total V2) = 0.9709506- ((7/15 * 0.9852281) + (8/15 * 0.8112781)) = 0.0784958
   c. Gain (Total V3) = 0.9709506- ((8/15 *0.954434) + (7/15 * 0.9852281)) = 0.014746
   d. Gain (Total V4) = 0.9709506- ((3/15 * 0.9182958) + (6/15 * 1) + (6/15 * 0.9182958)) = 0.0199731

Based on the results of the above calculations, it can be seen that V1 (material module) has a higher gain than the other gains, namely 0.9709506. Therefore, V1 is used as the root of the decision tree.

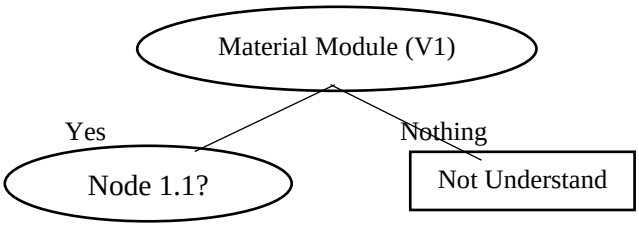

Figure 1. Decision Tree

Based on the decision tree above, it is known that the Nothing partition produces the value of Do not Understand while the Yes partition has two values, namely between understanding and not understanding. Therefore, the process of searching for the next node is carried out using the reduced data Yes filter Module Material = Yes. Here's the calculation Yeslah:

Table 2. Attributes of Material Module

No	Responden	V1	V2	V3	V4	Results
1	R1	Yes	Yes	Yes	Often	Understand
2	R2	Yes	Nothing	Yes	Rare	Not Understand
3	R4	Yes	Yes	Nothing	Rare	Understand
4	R8	Yes	Yes	Yes	Rare	Understand
5	R10	Yes	Yes	Yes	Nothing	Understand
6	R11	Yes	Nothing	Nothing	Rare	Understand
7	R13	Yes	Yes	Nothing	Often	Not Understand
8	R14	Yes	Nothing	Nothing	Nothing	Understand

1. Total Entropy Value
   a. Material Module Total Entropy Value Modul = ((-6/8*log 2 (6/8)) + (-2/8 * log 2 (2/8) = 0.8112781
   b. Entropy V2
      Attribute Value "Yes" = ((-4/5*log 2 (4/5)) + (-1/5 * log 2 (1/5) = 0.7219281
      Attribute Value "Nothing" = ((-2/3*log 2 (2/3)) + (-1/3 * log 2 (2/3) = 0.9182958
   c. Entropy V3
      Attribute Value "Yes" = ((-3/4*log 2 (3/4)) + (-1/4 * log 2 (1/4) = 0.8112781
      Attribute Value "Nothing" = ((-3/4*log 2 (3/4)) + (-1/4 * log 2 (1/4) = 0.8112781
   d. Entropy V3
      Attribute Value "Often" = ((-1/2*log 2 (1/2)) + (-1/2 * log 2 (1/2) = 1
      Attribute Value "Rare" = ((-3/4*log 2 (3/4)) + (-1/4 * log 2 (1/4) = 0.8112781
      Attribute Value "Nothing" = ((-2/2*log 2 (2/2)) + (-0/2 * log 2 (0/2) = 0
2. Calculating the Attribute Gain Value
   a. Gain (Total V2) = 0.8112781- ((5/8 * 0.7219281) + (3/8 * 0.9182958)) = 0.0157121125
   b. Gain (Total V3) = 0.8112781- ((4/8 *0.8112781) + (4/8 * 0.8112781)) = 0
   c. Gain (Total V4) = 0.8112781- ((2/8 * 1) + (4/8 * 0.8112781) + (4/8 * 0)) = 0.15563905

The results of the above calculations can be seen that V4 (Task/Practice) has a higher gain than the other gains, namely 0.15563905. Therefore, V4 is used as the Yes branch of V1:

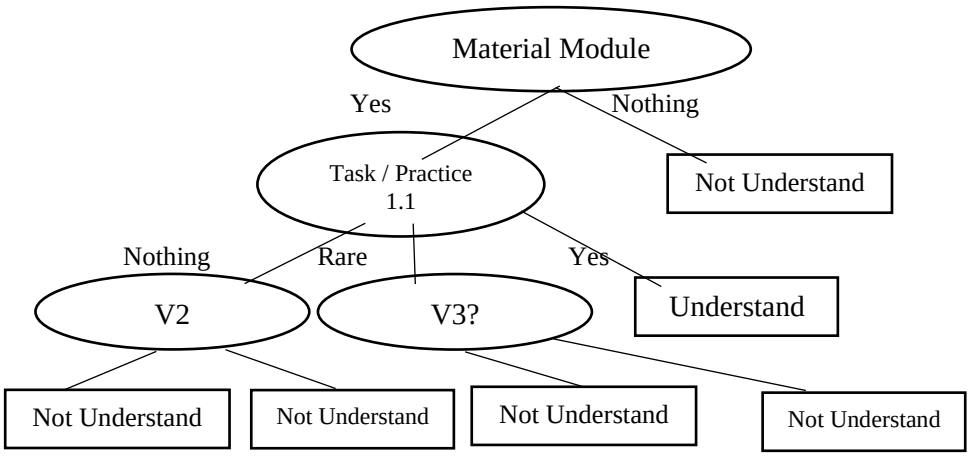

Figure 3. Decision Tree

The decision tree above shows that the Nothing partition produces an Understand value while the other partitions are unknown. For that do the search for the next node:

1. Total Entropy Value
   a. Total Entropy Value of Tasks/Practice = ((-1/2*log 2 (1/2)) + (-1/2 * log 2 (1/2) = 1
   b. Entropy V2
      Attribute Value "Yes" = ((-2/2*log 2 (2/2)) + (-0/2 * log 2 (0/2) = 0
   c. Entropy V3
      Attribute Value "Yes" = ((-1/1*log 2 (1/1)) + (-0/1 * log 2 (0/1) = 0.
      Attribute Value "Nothing" = ((-0/1*log 2 (0/1)) + (-1/0 * log 2 (1/0) = 0
2. Attribute Gain Value
   a. Gain (Total V2) = 1- ((2//2 * 0) + 0) = 1
   b. Gain (Total V3) = 1- ((1/ 2*0) + (1/2 * 0)) = 1

Based on the results of the analysis above, it was concluded that learning that utilizes online is not very inefficient, it can be seen from the results that have been analyzed that students do not understand the material delivered online.

4. CONCLUSION

Based on the discussion above, it can be concluded that the C5.0 algorithm is very suitable to be used to analyze the understanding of online learning at Budi Darma University during the Covid-19 pandemic. This algorithm provides results that can be used as consideration to note that the learning process by using online services is very inefficient to use in the teaching and learning process. Because from the results of the analysis above, it can be concluded that the online learning system is poorly understood by Budi Darma University students.

REFERENCES

[1] Y. Yuliana, "Corona virus diseases (Covid-19): Sebuah tinjauan literatur," *Wellness Heal. Mag.*, vol. 2, no. 1, pp. 187–192, 2020, doi: 10.30604/well.95212020.
[2] E. Kuntarto, "Keefektifan Model Pembelajaran Daring dalam Perkuliahan Bahasa Indonesia di Perguruan Tinggi," *J. Indones. Lang. Educ. Lit.*, vol. 3, no. 1, pp. 53–65, 2017, [Online]. Available: https://www.syekhnurjati.ac.id/jurnal/index.php/jeill/article/view/1820.
[3] M. Sulistiono, "Implementasi Hybrid Learning Menggunakan Aplikasi Edmodo Pada Matakuliah Metode Penelitian Kualitatif," *Elem. J. Ilm. Pendidik. Dasar Islam*, vol. 1, no. 1, p. 57, 2019, doi: 10.33474/elementeris.v1i1.2794.
[4] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: Deepublish, 2020.
[5] E. Prasetyo, *Data Mining, Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi, 2012.
[6] D. P. Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Media Inform. Budidarma*, vol. 4, no. 2, pp. 437–444, 2020.
[7] D. P. Utomo, P. Sirait, and R. Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5. 0," *Media Inform. Budidarma*, vol. 4, no. 4, pp. 994–1006, 2020.
[8] F. Hadi, "Penerapan Data Mining Dalam Menganalisa Pemberian Pinjamana Dengan Menggunakan Metode Algoritma C5 . 0 (Studi Kasus : Koperasi Jasa Keuangan Syariah Kelurahan Lambung Bukik)," *J. KomTekInfo*, vol. 4, no. 2, pp. 214–223, 2017.
[9] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, p. 1, 2018, doi: 10.23917/khif.v4i1.5975.
[10] D. Nofriansyah, *Konsep Data Mining vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish, 2015.
[11] D. Nofriansyah and G. W. Nurcahyo, *Algoritma Data Mining Dan Pengujiannya*. Yogyakarta: Deepublish, 2017.
[12] D. Dalbergio, M. N. Hayati, and Y. N. Nasution, "Klasifikasi Lama Studi Mahasiswa Menggunakan Metode C5.0 pada Studi Kasus Data Kelulusan Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Tahun 2017," *Pros. Semin. Nas. Mat. Stat. dan Apl. 2019*, vol. 1, no. 1, pp. 36–42, 2019.
[13] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "Implementasi Algoritma C5 . 0 Dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus : Kelurahan Mesjid Kecamatan Medan Kota)," *Inf. dan Teknol. Ilm.*, vol. 13, pp. 192–198, 2018.