C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject

Wiwiek Katrina¹, Habibah Jayanti Damanik¹, Friskila Parhusip¹, Dedy Hartama², Agus Perdana Windarto², Anjar Wanto²

¹Students of STIKOM Tunas Bangsa Pematangsiantar, North Sumatera, Indonesia
²STIKOM Tunas Bangsa Pematangsiantar, North Sumatera, Indonesia

*agus.perdana@amiktunasbangsa.ac.id

Abstract. This study aims to classify the level of understanding of students at STIKOM Tunas Bangsa. Data obtained from the results of even semester semester student questionnaires, with data samples of 165 students. The attributes used are 5, namely communication (C1), learning atmosphere (C2), learning media (C3), appearance (C4) and teaching methods (C5). The method used in this research is C4.5 Algorithm and processed using RapidMiner software to make decision trees. From the results of the study obtained eight rules to determine the level of student understanding, 4 rules understand and 4 rules do not understand. The C4.5 algorithm is accurately applied to determine the level of understanding of students with an accuracy rate of 87.50%. With this analysis, it is expected that it can help institutions in improving the quality of education through the performance of lecturers in the teaching and learning process. So as to produce STIKOM Tunas Bangsa graduates who are rich in knowledge and useful for the community.

1. Introduction

Understanding is very important in learning something because it relates to one's intelligence or level of knowledge. A person's ability is determined by the level of intelligence he has. The interesting or not a lecture material is also very much influenced by the learning method used by the lecturers. The use of media in learning by lecturers is also not optimal. This is thought to affect student learning outcomes that are not optimal. Learning media is a tool that serves to convey learning messages. Communication will not run without the help of message or media delivery suggestions. In the process of teaching and learning media presence has an important meaning. Because in the activity the material ambiguity delivered by the lecturer can be helped by presenting the media as an intermediary. The complexity of the material that will be delivered to students can be simplified with the help of media. The media can represent what the lecturer is unable to say through certain words or sentences. The difficulty of students understanding certain concepts and principles can be overcome with the help of learning media.

The teaching and learning process is inseparable from the role of lecturers and good student understanding. In the teaching and learning process, it is necessary to have a two-way relationship between students and teaching staff. That is intended to have good cooperation during the teaching and learning process. Analysis carried out by the campus towards the teaching and learning process is very necessary to be done at the end of the semester which aims as an assessment of students and lecturers. For students, this assessment aims to assess the level of understanding and absorption of the subjects taught and for this assessment lecturer aims to assess the extent to which the lecturer can channel his knowledge to students related to the courses delivered for 1 (one) semester. So that the campus can provide objective decisions in improving the quality of education, especially at STIKOM Tunas
Bangsa. Comprehension is the ability to understand or Content comprehend anything once they are known and remembered. Students are supposed to get a lot of opportunities to use their logical reasoning abilities, practice, formulate concepts, and participate in solving complex problems that require their great effort. Then, the students are encouraged to reflect their thoughts in drawing an accurate conclusion[1].

Based on the descriptions above, it can see that understanding is a form of a statement of learning outcomes. Intelligence is a level higher than knowledge or memory. Therefore, to improve understanding requires a good and right learning process. That does evidence by several studies in the field of data mining [2]–[7], field of artificial neural networks [8]–[12], in the field of decision support systems [13]–[17]. Data Mining is one of the fastest growing fields because of the vast demands for large-scale value-added databases that are in line with the growth of information technology. Data mining can extract large data sets into new knowledge [18]. The use of the C4.5 algorithm is expected to help institutions in improving the quality of education through the performance of lecturers in the teaching and learning process. To produce STIKOM Tunas Bangsa graduates who are rich in knowledge and useful for the community.

2. Methodology
2.1. Data Mining
Data mining is the process of discovering interesting knowledge, such as associations, patterns, changes, significant structures and anomalies, from large amounts of data stored in databases or data warehouses or other information repositories[2][19][20].

2.2. C4.5 Algorithm
C4.5 is a algorithm used to generate a decision trees. It is an extension of the ID3 algorithm used to overcome its disadvantages. The decision trees generated by the C4.5 algorithm can be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier. The C4.5 algorithm made a number of changes to improve ID3 algorithm[20][21].

2.3. Data Used
This research conducted at STIKOM Tunas Bangsa Pematang Siantar by using the C4.5 algorithm and processed using RapidMiner software to classify students to determine the level of students' understanding of the subject. The form of a flowchart can clearly describe the process of stages and steps in the classification using the C4.5 algorithm. Can be seen in Figure 1 in the form of a flowchart as follows:

Figure 1. Process Flowchart in C4.5 Algorithm
The steps to classify the Decision Tree C4.5 Algorithm are:

a. Prepare data
b. Select attributes as root
c. Calculate the value of gain and entropy
   \[ \text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i) \] 
   \[ \text{Entropy}(A) = \sum_{i=1}^{n} -p_i \cdot \log_2 p_i \] 
d. Share cases in branches
e. Repeat the process for each branch until all cases in the branch have the same class and or there are no attributes in the tuples that are partitioned again and/or there are no tuples in the empty branch.

This research method is carried out systematically to get a good workflow and is translated into several steps, namely:

1. Literature study is conducted to help researchers trace the developing theories about data mining and C4.5 Algorithm to obtain significant methods with existing problems.
2. The selection of research objects is done to facilitate researchers in classifying the level of understanding of students at STIKOM Tunas Bangsa, given the increasing number of students each year so that the data becomes useful.
3. In this study, 5 attributes were used in classifying students' level of understanding. The attributes used are as follows:
   a. Communication (C1), which is considered as the main factor between students and lecturers so that it influences students to understand courses.
   b. Learning atmosphere (C2), lecturers who can build a classroom atmosphere will make it easier for students to understand each material given.
   c. Learning media (C3), In the learning process the learning media plays an important role. Aside from being a tool for delivering material, learning media can also influence the attention of students. Lecturers must be able to provide the best teaching media so that the focus of students remains on the subject matter.
   d. Appearance (C4), lecturer readiness in teaching can be seen from the performance given. Tidiness symbolizes a strong spirit to give each lecture material. With a good appearance, it can attract students to focus more on studying courses.
   e. How to teach (C5), In delivering the teaching material, the lecturer must have a separate method so that the teaching material gets the attention of the students. Not monotonous and always creative is what expected of every student in the lecture.
4. Determine the class value of each attribute. For class values in attribute (C) can be seen in the table below:

| Criteria       | Field Name         | Data Class Type | Data Class Used          |
|----------------|--------------------|-----------------|--------------------------|
| C1 Communication | Nominal            | Open, Sometimes, Closed |
| C2 Learning atmosphere | Nominal            | Support, Not Support |
| C3 Instructional Media | Nominal           | Print, Non-Print, InFocus / Projector |
| C4 Appearance | Nominal            | Interesting, ordinary |
| C5 How to teach | Nominal            | Serious and relaxed, serious, relaxed, boring |

Analyze the data to classify the level of understanding of students in STIKOM Tunas Bangsa according to entropy based on each attribute using the C.45 method.

3. Result and Discussion

The researcher obtained the questionnaire data for the STIKOM Tunas Bangsa even semester 2017-2018 academic year, which was recorded in the excel format to be processed to obtain information. The table of student questionnaire data is as follows:

| Respondents | Communication | Learning atmosphere | Instructional Media | Appearance | How to teach |
|-------------|---------------|---------------------|---------------------|------------|--------------|
| A1          | Overt         | Support             | InFocus / Projector | Interesting| Serious and relaxed |
| A2          | Overt         | Support             | Print               | Interesting| Serious and relaxed |
The International Conference on Computer Science and Applied Mathematics
IOP Publishing
IOP Conf. Series: Journal of Physics: Conf. Series 1255 (2019) 012005 doi:10.1088/1742-6596/1255/1/012005

Preprocessing results obtained are then processed into the Decision tree using Rapid Miner software to determine the level of understanding of students who have completed the study, as for the decision tree Figure as follows:

| Communication | Learning atmosphere | Instructional Media | Appearance | How to teach | Results |
|---------------|---------------------|---------------------|------------|-------------|---------|
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Print               | Interesting | Serious and relaxed | Understand |
| Sometimes     | Support             | Non Print           | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Print               | Interesting | Serious and relaxed | Understand |
| Sometimes     | Support             | Print               | Ordinary   | Serious and relaxed | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious and relaxed | Understand |
| Overt         | Support             | Non Print           | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Print               | Interesting | Serious and relaxed | Understand |
| Sometimes     | Support             | Non Print           | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Print               | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Not Support          | Non Print           | Ordinary   | Serious      | Don’t Understand |
| Support       | Support             | Print               | Ordinary   | Serious      | Don’t Understand |
| Support       | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Support       | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Support       | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Support       | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Support       | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |

The preprocessing results of student data are as follows:

| Communication | Learning atmosphere | Instructional Media | Appearance | How to teach | Results |
|---------------|---------------------|---------------------|------------|-------------|---------|
| Sometimes     | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Print               | Interesting | Serious and relaxed | Understand |
| Sometimes     | Support             | Print               | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Print               | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Support             | Infocus / Projector | Ordinary   | Serious      | Don’t Understand |
| Overt         | Support             | Infocus / Projector | Interesting | Serious and relaxed | Understand |
| Overt         | Not Support          | Non Print           | Ordinary   | Serious      | Don’t Understand |

Table 3. Data preprocessing results
When viewed based on the results of the student understanding level classification decision tree contained in Figure 2, that the attribute that has the main influence to get the cause of the level of understanding is the variable B1 (Communication) which occupies as the root node.

It can be seen from the results of the decision tree and the text-shaped rule model that:

a) If Communication = Overt and the atmosphere of learning = Support and how to teach = relaxed and appearance of Interesting then Results = Understand.

b) If Communication = Overt and atmosphere of learning = Support, and way of teaching = Serious and relaxed and appearance = Interesting, then Results = Understand

c) If Communication = Overt and atmosphere of learning = Support, and way of teaching = Serious and relaxed and appearance = Ordinary media learning = in focus / Projector then Result = Understand

d) If Communication = Overt and learning atmosphere = Support and teaching methods = Serious and relaxed and appearance = Ordinary and learning media = Print, Understand Results (tryzha anggi) and Don't Understand (Nurliana)

4. Conclusion
The results of the study concluded that there was a classification of students' level of understanding at STIKOM Tunas Bangsa. Variables that have the highest priority to the level of understanding of students are lecturers who have Communication Overt, able to create an atmosphere of learning that supports, how to teach that Serious and relaxed, appearance of Interesting, in focus learning media / Projector then Understand students on the delivery of material provided by the lecturer compared to Communication sometimes (Overt or closed), learning atmosphere is not Support, Ordinary appearance. This influence can be seen from how a lecturer can build a classroom atmosphere so that students can understand each teaching material

References
[1] M. F. Goes, M. A. Sinhereiti, S. Consani, and M. A. Silva, “Morphological effect of the type, concentration and etching time of acid solutions on enamel and dentin surfaces,” Brazilian dental journal, vol. 9, no. 1, pp. 3–10, 1998.
[2] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” International Journal of artificial intelligence research, vol. 1, no. 2, pp. 26–33, 2017.
[3] U. R. Raval and C. Jani, “Implementing and Improvisation of K-means Clustering,” International Journal of Computer Science and Mobile Computing, vol. 5, no. 5, pp. 72–76, 2016.
[4] M. K. Arzoo, A. Prof, and K. Rathod, “K-Means algorithm with different distance metrics in spatial data mining with uses of NetBeans IDE 8.2,” International Research Journal of Engineering and Technology(IRJET), vol. 4, no. 4, pp. 2363–2368, 2017.
[5] S. Kumar and S. K. Rathi, “Performance Evaluation of K-Means Algorithm and Enhanced Mid-point based K-Means Algorithm on Mining Frequent Patterns,” International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 10, pp. 545–548, 2014.
[6] S. Sudirman, A. P. Windarto, and A. Wanto, “Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia,” IOP Conference Series: Materials Science and Engineering, vol. 420, no. 12089, pp. 1–8, 2018.

[7] A. Yadav and S. Dhingra, “An Enhanced K-Means Clustering Algorithm to Remove Empty Clusters,” IJEDR, vol. 4, no. 4, pp. 901–907, 2016.

[8] A. P. Windarto, L. S. Dewi, and D. Hartama, “Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm,” International Journal of Recent Trends in Engineering & Research (IJRTER), vol. 3, no. 10, pp. 1–12, 2017.

[9] A. Wanto, M. Zarlis, Sawaluddin, and D. Hartama, “Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process,” Journal of Physics: Conference Series, vol. 930, no. 1, pp. 1–7, 2017.

[10] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, “Implementation of Neural Networks in Predicting the Understanding Level of Students Subject,” International Journal of Software Engineering and Its Applications, vol. 10, no. 10, pp. 189–204, 2016.

[11] B. Febriadi, Z. Zamzami, Y. Yunefri, and A. Wanto, “Bipolar function in backpropagation algorithm in predicting Indonesia’s coal exports by major destination countries,” IOP Conference Series: Materials Science and Engineering, vol. 420, no. 12089, pp. 1–9, 2018.

[12] N. Nasution, A. Zamsuri, L. Lisnawita, and A. Wanto, “Polak-Ribiere updates analysis with binary and linear function in determining coffee exports in Indonesia,” IOP Conference Series: Materials Science and Engineering, vol. 420, no. 12089, pp. 1–9, 2018.

[13] A. N. D. J. D. Fadhilah, “Perancangan Aplikasi Pakar Penyakit Kulit Pada Anak Dengan Metode Expert System Development Life Cycle,” Jurnal Algoritma Sekolah Tinggi Teknologi Garut, vol. 9, no. 13, pp. 1–7, 2012.

[14] S. Fekri-Ershad, H. Tajalizadeh, and S. Jafari, “Design and Development of an Expert System to Help Head of University Departments,” International Journal of Science and Modern Engineering, vol. 1, no. 2, pp. 45–48, 2013.

[15] M. Min, “A rule based expert system for analysis of mobile sales data on fashion market,” 2013 International Conference on Information Science and Applications, ICISA 2013, 2013.

[16] M. Mohammadi and S. Jafari, “An expert system for recommending suitable ornamental fish addition to an aquarium based on aquarium condition,” arXiv preprint arXiv:1405.1524, vol. 3, no. 2, pp. 1–7, 2014.

[17] I. Chen and B. L. Poole, “Performance Evaluation of Rule Grouping on a Real-Time Expert System Architecture,” vol. 6, no. 6, pp. 883–891, 2014.

[18] C. Series, “Implementation of Data Mining to Analyze Drug Cases Using C4 . 5 Decision Tree Implementation of Data Mining to Analyze Drug Cases Using C4 . 5 Decision Tree,” pp. 0–6, 2018.

[19] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, “Classification of natural disaster prone areas in Indonesia using K-means,” International Journal of Grid and Distributed Computing, vol. 11, no. 8, pp. 87–98, 2018.

[20] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao, “Predicting Students’ Performance Using Id3 and C4.5 Classification Algorithms,” International Journal of Data Mining and Knowledge Management Process, vol. 3, no. 5, pp. 39–52, 2013.

[21] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. P. Windarto, “Application of Classification Method C4 . 5 on Selection of Exemplary Teachers,” in IOP Conference Series, 2018, pp. 1–6.