Application of the Decision Tree Method to Predict Student Achievement Viewed from Final Semester Values

Muhammad Wasil*, Aris Sudianto*, Fathurrahman*
Universitas Hamzanwadi, Lombok, Indonesia

*email: m.wasil@ymail.com

Abstract. Along with the development of the human mindset progress, humans began to develop a system that can help humans in dealing with problems that arise so that they can solve them easily. The decision tree or better known as the decision tree is an implementation of a system that humans develop in finding and making decisions for these problems by taking into account the various factors that are related within the scope of the problem and among about education. Education is a conscious and planned effort to create an atmosphere of learning and learning process so that students can increase their potential. The function of education is very important as one of the factors driving development as human resources to increase the ability of the community to develop science. Thus, the teaching and learning process is one of the determining factors for the success of education in schools/madrasas, the low quality of education is a result of the low quality of the learning process carried out in school

Keyword: decision tree, c4.5 algorithm, learning achievement

1. Introduction
Teaching and learning is one of the determining factors in the success of education in schools/madrasas, the low quality of education is a result of the low quality of the learning process carried out in schools. No exception in Madrasa Aliyah (MA) NW Rempung,[2] one thing that has not been done by many educational institutions or schools/madrasas is to anticipate students who have the potential to experience obstacles or lack of achievement in their learning[3]. This is considered important because the earlier educational institutions or schools/madrasas are aware of the potential of students who are likely to experience[4] obstacles in their learning, then educational institutions or schools/madrasas can take anticipatory steps. [5]Problems that often arise among schools/madrasas (teachers) including those in MA NW Rempung are the lack of anticipation or prevention early on for students who have the potential to experience obstacles in learning or underachievement, expelled from school. These two things are often heard by us and certainly experienced by all schools/madrasas[6]. [7]With these problems, the writer will try to use a method that is with a data mining approach. Meanwhile, [8]data mining is a series of processes to explore the added value of a data set in the form of knowledge that has not been known manually[9]. [10]With a data mining approach, the authors also apply the Decision Tree Algorithm C4.5 method[11] to determine the potential for student achievement[12]. The data used is the data of class X 2015/2016 students and class XI 2016/2017 which will be predicted in class XII. so that it can be a decision support used by the teacher/homeroom teacher at MA NW Rempung. [13]Based on the description above, it is the
underlying reason for the writer to take the title: [14] [15] the application of the decision tree method to predict the achievements of class XII students viewed from the end of semester grades at Madrasah Aliyah (MA) NW Rempung.

2. Methods
The methods used in this study include:

1. Literature Study
   Literature study is carried out by reading and studying books related to this research, as well as books that support the topics that will be discussed in the preparation of this article, as well as visiting (browsing) internet sites related to the researcher's article.

2. Field Study
   • Observation
     This method researchers collected data by conducting observations/observations to Madrasah Aliyah (MA) NW Rempung to meet with the teacher/homeroom teacher regarding the final semester student data, as well as the things needed in conducting this research.
   • Interview
     Interviews or questions and answers were conducted with the homeroom teacher XII Madrasa Aliyah (MA) NW Rempung to find out a picture of students' potentials and what obstacles would hinder the potential of students concerned academically and also to find things that were the obstacles of the teacher.

3. Results and Discussion
   The results obtained from the use of the decision tree method in predicting student achievement are as follows:
   The result of the decision tree

   ![Image: Decision Tree Results and Text View](image.png)

   Figure 1 Decision Tree Results and Text View

   Explanation: The main purpose of analyzing data using the Decision Tree C4.5 Algorithm is to get a rule that will be used for decision making on new data. The rule obtained from the picture above is if $\text{RESULTS} > 0.161$: Increases \{Increases = 21, Declines = 0\} and if $\text{RESULTS} \leq -0.161$: Decreases \{Increases = 0, Declines = 8\}
Accuracy results

Accuracy results tested were 97.22%, following the results of accuracy tests that have been tested.

Table 1. Accuracy results with K-Fold Validation 9

| Predicate Class | Increase | decrease |
|-----------------|----------|----------|
| Observed Class  |          |          |
| Increase        | 21       | 1        |
| decrease        | 0        | 7        |

Explanation: The number of True Positives (TP) is 22 records classified as increasing values and False Negative (FN) only 0 records are classified as increasing values but in reality are values that are decreasing. Next 7 records for True Negative (TN) are classified as declining values, and 1 False Positive (FP) record is classified as increasing but decreasing. Based on table 4.1 shows that the accuracy rate using the Decision Tree Algorithm C4.5 with K-Fold Validation 9 is 97.22% and can be calculated to find the accuracy, sensitivity, specificity, PPV, and NPV values in the equation below.

\[
\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} = \frac{22 + 7}{22 + 7 + 1 + 0} = 97.22\% \\
\text{Sensitivity} = \frac{tp}{tp + fn} = \frac{22}{22 + 0} = 100.00\% \\
\text{Specificity} = \frac{tn}{tn + fp} = \frac{7}{7 + 1} = 87.50\% \\
\text{PPV} = \frac{tp}{tp + fp} = \frac{22}{22 + 1} = 95.45\% \\
\text{NPV} = \frac{tn}{tn + fn}
\]
\[ \frac{7}{7+0} = 100,00\% \]

Table 2. Value of Accuracy, Sensitivity, Specificity, Ppv, Npv

| Value | %     |
|-------|-------|
| Accuracy | 97.22% |
| Sensitivity | 100.00% |
| Specificity | 87.50% |
| Ppv | 95.45% |
| Npv | 100.00% |

From table 2, there is an ROC graph with an AUC (Area Under Curve) value of 0.889% with an accuracy value of good classification.

![Figure 3. AUC values in the ROC Decision Tree Algorithm C4.5 with K-Fold Validation](image)

![Figure 4 Percentage Graph Based on Increasing and Decreasing Classes](image)

Based on Figure 4.5 it is explained that the conclusion of comparison / percentage of class X (ten) 2015/2016 and class XI (eleven) 2016/2017 students can be seen that the number of students is 21 people (72.41%) and the number of students has decreased 8 people (27.58%)
so that it can be used as a reference to predict the achievements of students who are in class XII (twelve).

4. Conclusion
1. Based on data analysis using the Decision Tree C4.5 Algorithm to predict student achievement in class XII (twelve) seen from the end of semester grades in the Madrasa Aliyah (MA) NW Rempung the following results are obtained:
2. By using the Decision Tree madrasas get the desired results, where the results are very different from those used before because the results obtained in more detail and detail are commonly seen from the decision tree obtained after the data is processed with RapidMiner 5
3. The Decision Tree method has a high speed in classifying student achievement predictions. This is evidenced by the travel time for the classification of student achievement quite quickly and accurately. During processing, the Decision Tree method produces a classification of decision trees based on the average value of the end of the semester so that it can find out the percentage/comparison of student achievement increases or decreases.
4. The average value of the end of the semester is a variable that determines the potential of a student to succeed (increase) or not (decrease) student achievement in 2 years. This is evidenced by the variable end of semester results that have been calculated into the Root Node in the Decision Tree that was formed.
5. From the research that has been done with the Decision Tree C4.5 Algorithm, the accuracy results obtained are very good or perfect that is 97.22% while the AUC value obtained from the Decision Tree C4.5 Algorithm is 0.889% with good classification

5. References
[1] G. Ağartioğlu Kundakçi, M. Yılmaz, and M. K. Sözmen, “Determination of the costs of falls in the older people according to the decision tree model,” Arch. Gerontol. Geriatr., vol. 87, no. February 2019, 2020.
[2] L. Sun, B. Zou, S. Fu, J. Chen, and F. Wang, “Speech emotion recognition based on DNN-decision tree SVM model,” Speech Commun., vol. 115, no. August, pp. 29–37, 2019.
[3] A. Trabelsi, Z. Elouedi, and E. Lefevre, “Decision tree classifiers for evidential attribute values and class labels,” Fuzzy Sets Syst., vol. 366, pp. 46–62, 2019.
[4] X. Meng, P. Zhang, Y. Xu, and H. Xie, “Construction of decision tree based on C4.5 algorithm for online voltage stability assessment,” Int. J. Electr. Power Energy Syst., vol. 118, no. July 2019, p. 105793, 2020.
[5] M. A. Adibi, “Single and multiple outputs decision tree classification using bi-level discrete-continues genetic algorithm,” Pattern Recognit. Lett., vol. 128, pp. 190–196, 2019.
[6] R. F. de Mello, C. Manapragada, and A. Bifet, “Measuring the Shattering coefficient of Decision Tree models,” Expert Syst. Appl., vol. 137, pp. 443–452, 2019.
[7] M. Gohari and A. M. Eydi, “Modelling of Shaft Unbalance: Modelling a Multi Discs Rotor
using K-Nearest Neighbor and Decision Tree Algorithms,” *Measurement*, vol. 151, p. 107253, 2019.

[8] M. A. Ganaie, M. Tanveer, and P. N. Suganthan, “Oblique Decision Tree Ensemble via Twin Bounded SVM,” *Expert Syst. Appl.*, vol. 143, p. 113072, 2020.

[9] Y. Ochiai, Y. Masuma, and N. Tomii, “Improvement of timetable robustness by analysis of drivers’ operation based on decision trees,” *J. Rail Transp. Plan. Manag.*, vol. 9, no. March, pp. 57–65, 2019.

[10] J. Obregon, A. Kim, and J. Y. Jung, “RuleCOSI: Combination and simplification of production rules from boosted decision trees for imbalanced classification,” *Expert Syst. Appl.*, vol. 126, pp. 64–82, 2019.

[11] M. Namazkhan, C. Albers, and L. Steg, “A decision tree method for explaining household gas consumption: The role of building characteristics, socio-demographic variables, psychological factors and household behaviour,” *Renew. Sustain. Energy Rev.*, vol. 119, no. November 2019, p. 109542, 2020.

[12] M. Barrios Castellanos, A. L. Serpa, J. L. Biazussi, W. Monte Verde, and N. do Socorro Dias Arrifano Sassim, “Fault identification using a chain of decision trees in an electrical submersible pump operating in a liquid-gas flow,” *J. Pet. Sci. Eng.*, vol. 184, no. February 2019, p. 106490, 2020.

[13] H. A. Cuesta, D. L. Coffman, C. Branas, and H. M. Murphy, “Using decision trees to understand the influence of individual- and neighborhood-level factors on urban diabetes and asthma,” *Heal. Place*, vol. 58, no. November 2018, p. 102119, 2019.

[14] R. Manikandan, R. Patan, A. H. Gandomi, P. Sivanesan, and H. Kalyanaraman, “Hash polynomial two factor decision tree using IoT for smart health care scheduling,” *Expert Syst. Appl.*, vol. 141, p. 112924, 2020.

[15] D. Hao *et al.*, “Application of decision tree in determining the importance of surface electrohysterography signal characteristics for recognizing uterine contractions,” *Biocybern. Biomed. Eng.*, vol. 39, no. 3, pp. 806–813, 2019.