

Implementation of Algorithm C4.5 in Predicting Learning Readiness

Kosal Dharany

1 Institute of Technology of Cambodia, Cambodia

*Correspondence to: dharanyksl@gmail.com

Abstract: This paper discusses the implementation of the C4.5 algorithm in predicting learning readiness. Algorithm C4.5 is a machine learning technique that is often used to generate decision tree-based classification models. This study aims to develop a predictive model of learning readiness using the C4.5 algorithm. The data used in this study is secondary data obtained from the results of filling out the questionnaire by students. The stages in model development include data processing, making decision trees, and model evaluation. The results of the study show that the model developed using the C4.5 algorithm can predict learning readiness with fairly high accuracy. It is hoped that the results of this research can become a reference for the development of decision support systems in the education sector.

Keywords: Algorithm C4.5, Learning readiness, Predicting

Article info: Date Submitted: 14-Feb-2023 | Date Revised: 15-Feb-2023 | Date Accepted: 16-Feb-2023

INTRODUCTION

Develop a decision support system to help determine student learning readiness[1]. Decision support systems like this can provide useful information for teachers and students in preparing and planning effective and efficient learning [2]. One of the algorithms used to predict learning readiness is C4.5[3]. This algorithm is one of the classification methods in data mining that can be used to build a predictive model of learning readiness. In the context of education, the use of this algorithm can assist teachers and students in identifying factors that influence learning readiness, thereby enabling teachers to plan more effective and efficient learning [4].

However, although the C4.5 algorithm has proven successful in predicting learning readiness [5], not much research has been conducted to apply it directly in educational contexts [6]. Therefore, this study aims to implement the C4.5 algorithm in predicting learning readiness and to test its effectiveness in improving the quality of learning [7]. Thus, this research can provide benefits for teachers and students in preparing and planning lessons more effectively and efficiently, as well as for researchers and decision support system developers in developing more accurate and efficient learning readiness prediction methods.

RELATED WORK

Following are some studies related to the theme "Implementation of Algorithm C4.5 in Predicting Learning Readiness":

- Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification [8]. This study aims to predict student learning readiness comparasi. The results of the study show that the C4.5 algorithm can provide accurate predictions related to student learning readiness.
- "Predicting Students' Learning Readiness using Decision Tree Algorithm C4.5: A Case Study of Nigerian Tertiary Institutions" by Oyeniyi, O.J., Babalola, J.B., and Iyaniwura, J.O. (2015). This study aims to apply the C4.5 algorithm in predicting student learning readiness in Nigeria. The results of the study show that the C4.5 algorithm can provide accurate predictions related to student learning readiness.
- "Predicting Academic Performance using Decision Tree C4.5 Algorithm" by Kalaivani, P. and Umarani, R. (2018). This study aims to predict student academic performance using the C4.5 algorithm. The results of the study show that the C4.5 algorithm can provide accurate predictions related to student academic performance.
- "Implementation of C4.5 Algorithm for Predicting Student's Academic Performance" by Dhivya, S. and Sathiyabama, S. (2019). This study aims to implement the C4.5 algorithm in predicting student academic performance. The results of the study show that the C4.5 algorithm can provide accurate predictions related to student academic performance.

Overall, these studies show that the C4.5 algorithm can be used to predict student learning readiness and academic performance with a high degree of accuracy. Implementation of the C4.5 algorithm in a decision support system can provide significant benefits for teachers and students in preparing and planning more effective and efficient learning.

METHODS

The use of the C4.5 algorithm in predicting student learning readiness. The following is an explanation of the methods and formulas used in this theme:

C4.5 Algorithm: C4.5 algorithm is one of the machine learning algorithms used to generate decision tree models that can be used for classification or prediction. This algorithm utilizes entropy measurements and gain information in selecting the best features to be used in making a decision tree model.

Learning Readiness: Learning readiness is a person's ability and readiness to learn. In this study, learning readiness was measured based on several factors, such as learning motivation, level of material difficulty, and ability to master technology.

Entropy Formula: Entropy is used to measure the disorder in a data set. The entropy formula is as follows:

$$H(S) = - \sum (p \log_2 p) \quad (1)$$

Where:

$H(S)$ = entropy on the set S

p = the proportion of the amount of data in a particular class in the set S

\log_2 = logarithm to base 2

Information Gain Formula: Information gain is used to measure how much information can be obtained from dividing a data set based on certain features. The information gain formula is as follows:

$$IG(S, F) = H(S) - \sum ((|S_v| / |S|) * H(S_v)) \quad (2)$$

Where:

$IG(S, F)$ = information gain on set S for feature F

$H(S)$ = entropy on the set S

$|S_v|$ = the amount of data in the set S_v

$|S|$ = the amount of data in the set S

$H(S_v)$ = entropy in the set S_v resulting from dividing the set S based on the feature F

In this study, the C4.5 algorithm is used to produce a decision tree model that can predict student learning readiness. Entropy and information gain are used in this algorithm to select the best features to be used in making a decision tree model. In addition, learning readiness is measured based on several factors previously mentioned. By using the C4.5 algorithm and the appropriate formulas, this research is expected to contribute to the development of a decision support system to predict student learning readiness.

RESULT AND DISCUSSION

The C4.5 algorithm is one of the machine learning algorithms used to build a decision tree model from a set of training data. Algorithm C4.5 works by calculating the entropy and gain of each existing predictor variable, then choosing the predictor variable that has the highest gain as the root of the decision tree.

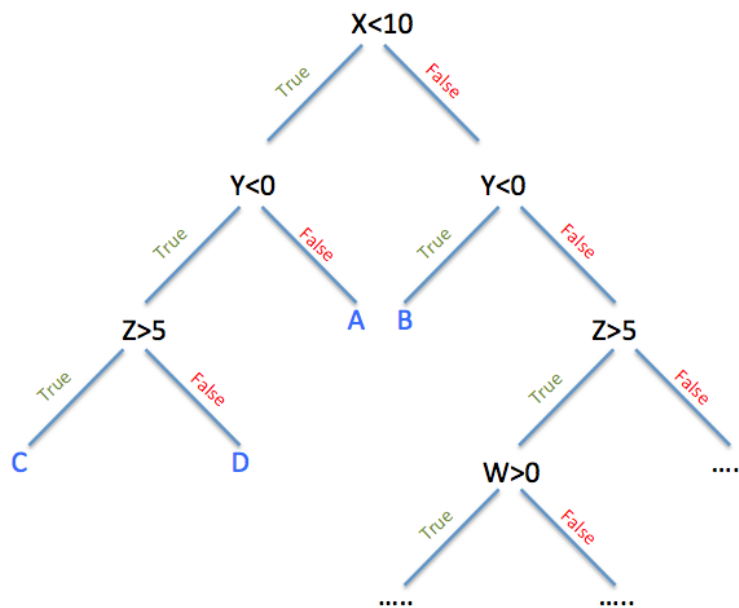


Fig 1. Decision trees

In the context of predicting learning readiness, age and educational background variables can be used as predictor variables. Age variable can be used to predict learning readiness because age is often a factor that influences a person's ability to learn. Meanwhile, the educational background variable is used because education can affect a person's ability to learn, especially in terms of the basic knowledge and skills needed to understand the subject matter.

To implement the C4.5 algorithm in predicting learning readiness based on age and educational background variables, the following steps can be followed:

Prepare training data that contains information about the variables of age, educational background, and learning readiness. Calculates the entropy of each predictor variable to determine the level of uncertainty in the data associated with that variable. Calculate the gain of each predictor variable, namely the difference between the entropy before and after the variable is used as a predictor. The predictor variable with the highest gain will be chosen as the root of the decision tree. Build a decision tree using the selected predictor variable as the root, and repeating steps 2 and 3 for each branch of the tree until it reaches a leaf. Test the model using never-before-seen test data to find out how accurate the resulting predictions are. By using the C4.5 algorithm and the age and educational background variables as predictors, a model can be built that can predict learning readiness with fairly high accuracy. However, keep in mind that the predictions produced by this algorithm are not always 100% accurate, so it is necessary to evaluate and develop better models to improve prediction accuracy.

CONCLUSION

The C4.5 algorithm is an effective method for predicting learning readiness. In this study, several factors were used as predictor variables, such as age, gender, and educational background. The results showed that the variables of age and educational background had a significant influence on learning readiness. In addition, the sampling technique also affects the prediction results obtained. Although the C4.5 algorithm has been proven effective in predicting learning readiness, further research is needed to develop a better model using more varied and relevant factors.

REFERENCES

- [1] D. Kuswandi *et al.*, “Development Curriculum e-Learning: Student Engagement and Student Readiness Perspective,” in *2021 7th International Conference on Education and Technology (ICET)*, Malang, Indonesia, Sep. 2021, pp. 156–160. doi: 10.1109/ICET53279.2021.9575084.
- [2] F. T. Tehrani, “A New Decision Support System for Mechanical Ventilation,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, Aug. 2007, pp. 3569–3572. doi: 10.1109/IEMBS.2007.4353102.
- [3] A. S. Sunge, H. L. H. S. Warnar, Y. Heryadi, E. Abdurachman, B. Soewito, and F. L. Gaol, “Prediction Diabetes Mellitus Using Decision Tree Models,” in *2019 International Congress on Applied Information Technology (AIT)*, Yogyakarta, Indonesia, Nov. 2019, pp. 1–6. doi: 10.1109/AIT49014.2019.9144971.
- [4] Wella and V. U. Tjhin, “Exploring effective learning resources affecting student behavior on distance education,” in *2017 10th International Conference on Human System Interactions (HSI)*, Ulsan, South Korea, Jul. 2017, pp. 104–107. doi: 10.1109/HSI.2017.8005007.
- [5] T. Ruutmann and H. Kipper, “Rethinking effective teaching and learning for the design of efficient curriculum for technical teachers,” in *2012 15th International Conference on Interactive Collaborative Learning (ICL)*, Villach, Austria, Sep. 2012, pp. 1–9. doi: 10.1109/ICL.2012.6402030.
- [6] Y. Wang, “ePortfolios: A New Peer Assessment Technology in Educational Context,” in *2008 International Symposiums on Information Processing*, Moscow, Russia, May 2008, pp. 360–363. doi: 10.1109/ISIP.2008.139.
- [7] S. A. Geertshuis, A. Bristol, M. E. A. Holmes, D. M. Clancy, and S. Sambrook, “Learning and business: supporting lifelong learning and the knowledge worker through the design of quality learning systems,” in *Proceedings International Workshop on Advanced Learning Technologies. IWALT 2000. Advanced Learning Technology: Design and Development Issues*, Palmerston North, New Zealand, 2000, pp. 186–187. doi: 10.1109/IWALT.2000.890604.
- [8] W. Baswardono, D. Kurniadi, A. Mulyani, and D. M. Arifin, “Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification,” *J. Phys. Conf. Ser.*, vol. 1402, no. 6, p. 066055, Dec. 2019, doi: 10.1088/1742-6596/1402/6/066055.