Classification Model Development Student Academic Eligibility In Capstone Project Course Enrollment Use C4.5 Decision Tree Algorithm

Lola Astri Nadita , Muhammad Iqbal

Abstract

Capstone Project course is an important aspect in the implementation of higher education to ensure that students have met the academic requirements and are ready to complete the final project. The process of determining eligibility is generally based on several criteria, such as the number of semester credit units (SKS) taken, Grade Point Average (GPA), and completion of prerequisite courses. This study aims to analyze the application of the *Decision Tree algorithm using the C4.5 method in classifying student eligibility to take the Capstone Project* course . The C4.5 method was chosen because it is able to form a classification model based on relevant attributes and produces decision rules that are easy to understand and interpret. The research stages include collecting student academic data, data preprocessing, attribute selection, forming a decision tree based on the *Gain Ratio value* , the results of the study are expected to produce a classification model that is able to determine student eligibility consistently based on predetermined academic criteria. The model can then be used as a basis for developing a decision support system to assist study programs in evaluating student eligibility more objectively, efficiently, and transparently. Thus, the application of the C4.5 algorithm is expected to support the improvement of the quality of academic governance and provide a basis for more structured decision-making in the implementation of the *Capstone Project course* .

Keywords : C4.5, Classification , *Capstone Project* , System Decision Support

Lola Astri Nadita
Magister Teknologi Informasi, Universitas Pembangunan Pancabudi, Indonesia
e-mail: lolanadita@gmail.com

Muhammad Iqbal
e-mail: muhammadiqbal@dosen.pancabudi.ac.id
3rd International Conference Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Cross-Sector and Cross-Disciplinary Strategies for Achieving the SDGS
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

1.1 Concept and Application of Decision *Tree* Algorithm

The development of information technology in *the big data era* has driven a growing need for data analysis methods capable of producing accurate information and supporting effective decision-making. The use of data in decision-making processes is becoming increasingly important in various fields, such as business, healthcare, education, and information technology. Through appropriate data analysis, organizations can identify patterns, understand relationships between variables, predict the likelihood of an event occurring, and formulate decisions based on empirical evidence. One machine learning method *widely* used to meet these needs is the decision tree *algorithm* .

A decision tree is a machine learning algorithm with a hierarchical structure resembling a tree and is used to solve various data analysis problems. This algorithm works by dividing a set of data into groups based on certain attributes or features deemed relevant to the analysis objective. The division process is carried out in stages until a decision or prediction is obtained at the final node. The structure of a decision tree consists of a root node *representing* the starting point for decision-making, internal nodes *indicating* test attributes, branches *depicting* test results or decision rules, and leaf *nodes* representing classification results or predicted values.

As part of the *supervised learning approach* , decision tree algorithms utilize data that already has labels or target values to build predictive models. Based on the characteristics of the target used, this algorithm can be applied to both classification and regression problems. In classification, the model is used to determine a specific category, while in regression, the model is used to estimate a numerical value. In the formation process, the decision tree selects the best attributes based on certain criteria, such as *Information Gain* , *Gain Ratio* , or *Gini Index* , depending on the algorithm used. The purpose of this attribute selection is to produce an effective data division so that the model can distinguish classes or estimate target values well.

One of the main advantages of decision tree algorithms is their ability to generate decision rules that are relatively easy to understand and interpret. Unlike some machine learning models, which have more complex internal mechanisms, decision trees allow users to trace the decision-making process from the root node to the leaf nodes. This characteristic makes them relevant for a variety of needs, including creditworthiness analysis, early health diagnosis, customer segmentation, student achievement prediction, and decision support system development. For example, in assessing the eligibility of potential bank loan applicants, the model can use attributes such as income, credit history, and number of dependents to classify potential borrowers into specific categories based on patterns learned from historical data.

Despite its advantages in terms of interpretability and ease of implementation, the decision tree algorithm also has several limitations. This model is prone to overfitting, a condition where the tree adapts too closely to the training data, reducing its generalizability to new data. Furthermore, small changes in the data can result in a different tree structure, necessitating careful attention to model stability. To overcome these limitations, several techniques can be applied, such as tree pruning, limiting tree depth, setting a minimum number of samples at each node, and evaluation using test data separate from the training data.

Based on these characteristics, the decision tree algorithm is a potential approach to data analysis and technology-based decision-making. Its ability to identify patterns, form decision rules, and present easily interpretable results makes it useful for both researchers and practitioners. Therefore, understanding the basic concepts, working mechanisms, advantages, limitations, and applications of decision trees is essential in determining the suitability of this algorithm for solving various real-world problems.

Tree decision built using training data , then used For predict new data . Structure tree decision consists of on :

- Knot root node: node beginning place tree decision start analyze data. Node root usually chosen based on the most significant attribute , namely the most influential attribute in divide the data into more groups regular .

- Knot branch (branch node): a functioning node divide data based on attribute certain . Knot branch is part important from the analysis process Because determine how the data from One knot separated into two groups or more based on rule certain .
- Knot leaf node: node final in tree . Node This No own branch Again Because has reach results end in the form of decision or predictions . Every knot leaf usually load output end , such as class labels in classification or mark certain on regression .

This process repeated until No There is Again attributes that can used For share data or until criteria certain fulfilled , for example depth maximum tree (maximum depth) or minimum amount of data in node (minimum sample split).

Literature Review

Previous Research and Relevance of the C4.5 Algorithm in Student Eligibility Classification

The C4.5 decision tree algorithm has been widely applied in various studies to solve classification problems and support data-driven decision-making. In the field of social services, research by Susanto and colleagues demonstrated that the C4.5 algorithm can be used to determine the eligibility of social assistance recipients based on economic characteristics. This approach allows for a more systematic selection process based on predetermined criteria. Similar findings were reported by Mandasari and Hartana through the application of the C4.5 algorithm in the classification of social assistance recipients. The use of this classification model has the potential to increase assessment consistency and reduce subjectivity in the decision-making process.

In education, the C4.5 algorithm also plays a crucial role in supporting the evaluation and academic decision-making process. Research by Fazira and colleagues demonstrates that this algorithm can be used to classify students' eligibility for educational assistance by considering several criteria, including economic conditions and academic achievement. The application demonstrates that the C4.5 algorithm is capable of processing various relevant attributes to generate classifications based on patterns found in the data. Therefore, this method has the potential to assist educational institutions in implementing a more structured and measurable selection process.

In addition to determining eligibility for educational assistance recipients, the C4.5 algorithm has also been utilized to analyze academic achievement and predict student graduation. By processing historical academic data, this algorithm can be used to identify patterns related to study success and identify factors that contribute to student academic achievement. In the context of predicting study duration, the classification model can help universities identify students who are likely to experience delays in graduation and those who have the opportunity to complete their studies on time. This information can serve as a basis for institutions to design mentoring strategies and academic interventions earlier. This demonstrates that the C4.5 algorithm is not only relevant for solving classification problems but can also support predictive analysis in higher education management.

The flexibility of the C4.5 algorithm is also reflected in its application to sectors outside of education, such as creditworthiness analysis and insurance product selection. Research by Yusuf and colleagues shows that the C4.5 algorithm can be used to classify creditworthiness based on a number of relevant attributes. This application demonstrates the algorithm's ability to generate decision rules based on the characteristics of the analyzed data. Its ability to generate tree structures and classification rules that are relatively easy to interpret is a key advantage, especially in problems that require transparency and traceability of the decision-making process.

Based on these various applications, the C4.5 decision tree algorithm has strong potential for use in classification and decision support systems across a wide range of fields. Its advantages in attribute selection based on Gain Ratio, its easy-to-understand decision rule formation, and its ability to handle both numeric and categorical attributes make this algorithm relevant for problems involving diverse assessment criteria. However, the model's effectiveness

still needs to be proven through testing using representative data and appropriate evaluation metrics.

In this study, the C4.5 algorithm is applied to determine student eligibility to take the Capstone Project course based on academic criteria, such as Grade Point Average (GPA), number of semester credit units (SKS) taken, and completion of prerequisite courses. The application of this method is expected to produce a classification model that helps study programs evaluate student eligibility more systematically, consistently, and transparently. In addition to generating eligibility categories, the resulting decision tree also has the potential to identify the academic attributes that most influence decisions. Thus, this research is expected to contribute to the development of data-based academic decision support systems, while strengthening the use of machine learning algorithms in higher education management.

In addition to determining eligibility for educational assistance recipients, the C4.5 algorithm has also been utilized to analyze academic achievement and predict student graduation. By processing historical academic data, this algorithm can be used to identify patterns related to study success and identify factors that contribute to student academic achievement. In the context of predicting study duration, the classification model can help universities identify students who are likely to experience delays in graduation and those who have the opportunity to complete their studies on time. This information can serve as a basis for institutions to design mentoring strategies and academic interventions earlier. This demonstrates that the C4.5 algorithm is not only relevant for solving classification problems but can also support predictive analysis in higher education management.

The flexibility of the C4.5 algorithm is also reflected in its application to sectors outside of education, such as creditworthiness analysis and insurance product selection. Research by Yusuf and colleagues shows that the C4.5 algorithm can be used to classify creditworthiness based on a number of relevant attributes. This application demonstrates the algorithm's ability to generate decision rules based on the characteristics of the analyzed data. Its ability to generate tree structures and classification rules that are relatively easy to interpret is a key advantage, especially in problems that require transparency and traceability of the decision-making process.

Based on these various applications, the C4.5 decision tree algorithm has strong potential for use in classification and decision support systems across a wide range of fields. Its advantages in attribute selection based on Gain Ratio, its easy-to-understand decision rule formation, and its ability to handle both numeric and categorical attributes make this algorithm relevant for problems involving diverse assessment criteria. However, the model's effectiveness still needs to be proven through testing using representative data and appropriate evaluation metrics.

In this study, the C4.5 algorithm is applied to determine student eligibility to take the Capstone Project course based on academic criteria, such as Grade Point Average (GPA), number of semester credit units (SKS) taken, and completion of prerequisite courses. The application of this method is expected to produce a classification model that helps study programs evaluate student eligibility more systematically, consistently, and transparently. In addition to generating eligibility categories, the resulting decision tree also has the potential to identify the academic attributes that most influence decisions. Thus, this research is expected to contribute to the development of data-based academic decision support systems, while strengthening the use of machine learning algorithms in higher education management.

Research Methodology

Study This use approach quantitative with data mining methods , in particular 3797technique classification use algorithm *Decision tree* (C4.5). Main objective study is develop a model that can classify eligibility student For take a Capstone Project based on academic data .

b. Data Sources and Types

Data used in study This is secondary data obtained from system university academics . The data consists of on historical data students . Example attributes used includes :

- GPA (Index Performance Cumulative)
- Amount credit that has been taken
- Currency value studying prerequisite
- Graduation status eye core lectures
- Administrative status
- Class label (Eligible/Uneligible)

2. Stages Study

a. Data Collection

Data collected from academic database or archives institution . Collected data must relevant with determination eligibility take the Capstone Project.

b. Data Preprocessing

Stage This aim increase data quality before analyzed , which includes :

- Data cleaning : removing duplicate data or No complete .
- Data transformation : changing numeric data become category , for example GPA becomes high , medium , and low .
- Data selection : selecting relevant attributes .

c. Data Sharing

Data is shared into two parts :

- Training data : used For building a model, for example 70%.
- Testing data : used For test the model, for example 30%.

3. Implementation C4.5 Algorithm

C4.5 algorithm is used For build tree decision through steps following .

a. Counting Entropy

Entropy used For measure level uncertainty in data.

$$Entropy(S) = -\Sigma_{i=1}^{n} p_i \log_2(p_i)$$

b. Calculating Gain and Gain Ratio

Information Gain is used For determine attribute best :

$$Gain(S, A) = Entropy(S) - \Sigma_i (|S_i| / |S|) \times Entropy(S_i)$$

Gain Ratio is used as improvements :

$$Gain\ Ratio(A) = Gain(S, A) / SplitInfo(A)$$

c. Formation Decision Tree

- Choose attribute with highest Gain Ratio value as roots .
- Share data based on attribute the .
- Repeat the process until all data is classified .

d . Pruning

Pruning done For reduce complexity trees and avoid overfitting.

5. Model Evaluation

resulting model tested using test data with metric evaluation following :

- Accuracy : level accuracy classification .
- Precision : accuracy predictions in class positive .
- Recall: ability find all data is positive .
- Matrix confusion matrix: displays results classification in a way detailed .

6. Implementation System (Optional)

The model created can implemented in form application web- based or system supporters decision .

7. Research Flow

In a way general , flow study can explained as following :

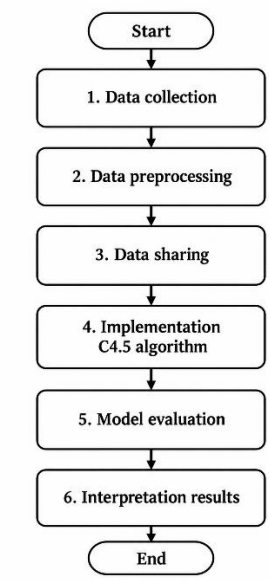


Figure 1. Methodology study

Results

Based on data processing using algorithm Decision Tree C4.5, obtained a classification model that can determine eligibility student take a Capstone Project based on a number of attribute academic , such as GPA, number of credit , and settlement eye studying prerequisites .

1. Formation Results Decision Tree

Based on calculation entropy and Gain Ratio, attributes with mark highest chosen as root tree decision . In research this , the most dominant attribute is :

- Amount credit
- GPA
- Completion status eye studying prerequisite

Tree the resulting decisions form a number of rule decisions , including :

- If the amount credits > 144, GPA > 3.00, and all eye studying prerequisite have passed, then student declared Eligible.
- If the amount credit < 144, then student declared Unfit.
- If GPA < 2.75, then student declared Unfit.
- If there is eye studying prerequisites that have not been passed, then student declared Unfit.

Rules the show that factor academic own significant influence to eligibility student .

2. Model Evaluation Results

resulting model tested using test data with results as following :

- Accuracy : 85%
- Precision : 83%
- Recall: 87%

Evaluation results show that the model has level good accuracy in classify student data .

3. Matrix Confusion

Test results are also displayed in form matrix confusion following :

Condition current	Worthy Prediction	Unworthy Prediction
Actual Worthy	52	8
Actual Not Eligible	7	33

The table show that part big data success classified with Correct .

4. Analysis of Results

Based on results study :

- C4.5 algorithm is capable of produce accurate and easy classification models understood .
- The most influential factor is amount credit and GPA.

- Model can used as system supporters decision For determine eligibility student .

5. Interpretation

Classification results show that students who have fulfil condition academic main tend categorized worthy , whereas students who have not fulfil condition classified No worthy . With Thus , the system This can help party academic take decision in a way more objective and consistent .

Conclusion

Research conclusion This is that results classification show students who have fulfil condition academic main tend categorized worthy For take eye studying *Capstone Project* , whereas students who have not fulfil condition classified No worthy take eye studying *Capstone Project* . With Thus , the system This can help party academic take decision in a way more objective and consistent .

References

[1] Quinlan, J.R. (1993). C4.5: Programs for Machine Learning. San Mateo: Morgan Kaufmann Publishers.

[2] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed .). Elsevier.

[3] Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

[4] Kusrini , & Luthfi , ET (2009). Data Mining Algorithms . Yogyakarta: Andi Publisher .

[5] Larose, D.T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. Wiley Interscience .

[6] Tan, P.N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Pearson Education.

[7] Gorunescu , F. (2011). Data Mining: Concepts, Models and Techniques. Springer.

[8] Suyanto. (2017). Data Mining for Data Classification and Clustering . Informatics .

[9] Prasetyo , E. (2012). Data Mining: Concepts and Applications Using MATLAB. Andi.

[10] Hanifah, S., & Kurniawan, D. (2020). Implementation C4.5 algorithm for Classification Eligibility Recipient Help Social . Journal Technology Information , 14(2), 123–130.

[11] Mandasari , I., & Hartana , A. (2021). Implementation C4.5 algorithm in Determination Recipient Help . Journal Scientific Informatics , 8(1), 45–52.

[12] Susanto, A., Wijaya, R., & Pratama, Y. (2019). Application of Decision Tree C4.5 for Classification of Feasibility Data . Journal System Information , 10(1), 1–8.

[13] Putra, DWT, & Andriani, R. (2019). Unified Modeling Language (UML) in Design System Information . Journal Technology Information and Communication , 7(1), 1–8.

[14] Sari, N., & Fitriani, L. (2022). Data Mining Analysis for Prediction Graduation Student Using C4.5. Journal Informatics , 16(1), 55–63.

[15] Rahman, F., & Hidayat, T. (2020). Implementation Decision Tree Algorithm for Prediction Student Academic Achievement . Journal Computers and Informatics , 12(2), 89–96.

[16] Hartana, & Mandasari, RD (2024). Implementation of the Decision Tree algorithm for the classification of eligibility for the distribution of social assistance funds. Responsive Journal: Science and Informatics Research, 6(2), 240–246. https://doi.org/10.51977/jti.v6i2.1730

[17] Iskandar, I., Hiryanto, L., & Hendryli, J. (2018). Predicting student graduation using the Decision Tree C4.5 algorithm with pruning techniques. Journal of Computer Science and Information Systems, 6(1), 64. https://doi.org/10.24912/jiksi.v6i1.2599

[18] Kamagi, DH, & Hansun, S. (2014). Implementation of data mining with the C4.5 algorithm to predict student graduation rates. Ultimatics: Journal of Informatics Engineering, 6(1), 15–20. https://doi.org/10.31937/ti.v6i1.327

[19] Khasanah, N., Saputri, DUE, Hidayat, T., & Aziz, F. (2025). Predicting student graduation using the C4.5 algorithm with RapidMiner: A case study of XYZ university academic data.