The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00036(2024) https://doi.org/10.29103/icomden.v2.765
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Application of Decision Tree C4.5 to determine post-Madrasah Aliyah students' college majors (case study: Madrasah Aliyah Misbahul Ulum)

Tasya Mulyani[✉1] **Bustami**[2] **Sujacka Retno**[3]
[1]Informatics Engineering, Universitas Malikussaleh, Aceh, Lhokseumawe, 24353, Indonesia, tasya.200170003@mhs.unimal.ac.id
[2]Informatics Engineering, Universitas Malikussaleh, Aceh, Lhokseumawe, 24353, Indonesia, bustami@unimal.ac.id
[3]Informatics Engineering, Universitas Malikussaleh, Aceh, Lhokseumawe, 24353, Indonesia, sujacka@unimal.ac.id
[✉]Corresponding Author: **tasya.200170003@mhs.unimal.ac.id | Phone: +6282274470617**

Abstract
This research aims (1) To apply Decision Tree C4.5 in determining post-Madrasah Aliyah students' college majors (2) To find out the results of applying the C4.5 Decision Tree Method in determining post-Madrasah Aliyah students' college majors (3) To design a system decision support to simplify and speed up the process of analyzing student data to determine post-Madrasah Aliyah students' college majors. This system was developed using the PHP programming language by applying the C4.5 Decision Tree algorithm method. Decision Tree is a machine learning algorithm that uses a set of rules to make decisions with a tree-like structure with the concept of presenting the algorithm with conditional statements, which includes branches to represent decision-making steps that can lead to profitable results by determining the value of entropy, split info and he gain ratio of this research produces objective results. Involving students from class 12 odd semesters, this research resulted in the identification of engineering majors for each student.
Keywords: C4.5 algorithm, decision tree, php, system

Introduction
Background
The development of computerized systems cannot be separated from technological developments which affect almost all aspects of human life. In the development of computerized systems, the use of information technology in the field of knowledge is very necessary. One of the advantages of a computerized system is that it can produce all kinds of information in a relevant, timely and accurate manner. Education now requires applications that can process data quickly, to make it easier for students to get information and help and make it easier for teachers to process student data, so that teachers know and can guide students' desired college decisions at school.

Therefore, a decision support system is needed to determine the best choice from all criteria. It takes quite a long time to evaluate a student's college decision, because the evaluation process is still done manually, so this makes the process results slow, therefore a computerized decision support system is needed to help the decision making process be faster and more accurate.

Decision trees are machine learning algorithms that use a set of rules to make decisions with a tree-like structure that models possible outcomes, resource costs, utilities and possible consequences or risks. The concept is to present an algorithm with conditional statements, which include branches to represent decision-making steps that can lead to a favorable outcome. Where each branch represents the results for an attribute, while the path from the leaves to the root represents the rules for classification.

Therefore, the Decision Tree Method is needed to identify students' college plans. The application of the Decision Tree Method is also a form of technology development in education. The Decision tree method usually imitates human thinking abilities when making decisions, so it is easy to understand.

With the problems above, the author felt interested in conducting research using the decision tree method.

Formulation of the problem
Based on the background that has been described above, the problems that the author formulated are as follows:
(1) How to apply the Decision Tree C4.5 Method in determining the study majors of post-Islamic Senior High School students?
(2) What are the results of applying the Decision Tree C4.5 Method in determining the study majors of post-Islamic Senior High School students?

Materials & Methods

Place and Time of Research

The location of this research was conducted at the Misbahul Ulum Private Islamic High School located at Jalan Tgk.Chiek Dipaloh, Muara Satu District, Lhokseumawe City, with the research period starting from January 2024.

Research Steps

The research steps taken are:

(1) Data Collection

At this stage, the data collection process used for research purposes will be explained, such as what data is used in the research, and how the data is used.

(2) Data Processing Method

Data processing uses various basic statistical methods such as finding Mean, Mode, Median and Percentage, Prediction results using various software such as Excel and Google Spreadsheet.

(3) System Design

System design uses Context Diagrams, DFD and ERD. Where this design will help the application development process that uses a programming language.

(4) System Implementation

System implementation is the process of developing an application using a programming language. In this case, a programming language such as PHP will be used.

(5) System Testing

System testing is carried out by carrying out the stages of testing and debugging the program to ensure that it can run properly according to the previously made design.

Research methods

This research method uses a research flow that starts from data collection and ends with conclusions.

(1) Data Collection

Data collection is important to do so that information related to the research object is obtained to support the research process.

(2) Determining Attribute Values

In the sample data, the selected nodes are first determined, namely by calculating the information gain value of each attribute to determine the selected node, use the largest information gain value. Then the entropy and gain values of each attribute will be calculated.

(3) Graph View

Graph view shows the results of the classification with branches that can produce conclusions.

(4) Conclusion

The rules obtained from the results of the decision tree and the description of the tree above to analyze career recommendation factors based on attributes.

System Scheme

The following is a schematic illustration of the system for implementing the decision tree c4.5 to determine the study majors for post-Islamic high school students.

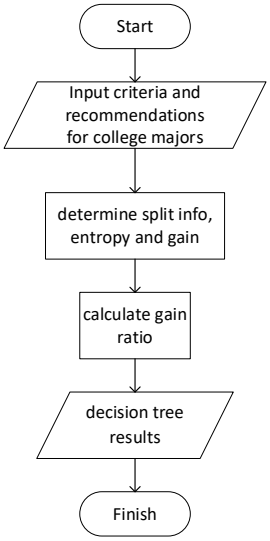

Figure 1. System Scheme

Description:
(1) The first step is to input the criteria values and recommendations for college majors
(2) Next, the application will determine SplitInfo, Entropy and Gain
(3) Then, the application will calculate the gain ratio
(4) When finished, the application will display the results of the Decision Tree value.

Results and Discussion

Design Analysis

In this study, the criteria for determining the study program of post-Islamic Senior High School students at Misbahul Ulum Islamic Senior High School were identified through the Decision Tree method, considering 5 subjects, namely Mathematics, English, Biology, Physics, and Chemistry. The students who were the subjects of the study came from class 12 semester 1 (odd) with a total of 91 students.

Stages of the Decision Tree Method C4.5

For the subject scores of grade 12 students in semester 1 (odd), the dataset used is as follows:

Table 1. Student grades dataset

Student Name	Math	English	Biology	Physics	Chemistry	Department of Engineering
Agustina	96	96	96	96	96	Chemical Engineering
Alfia Rahmi	96	96	96	85	85	Informatics Engineering
Aulia Farissa	96	96	87	85	82	Informatics Engineering
Ayu Nabila	94	83	89	95	82	Architectural Engineering
Cut Fara in	97	97	97	97	97	Industrial Engineering
Cut Rahil	96	86	97	97	97	Chemical Engineering
Cut Rahimah	97	97	86	97	97	Industrial Engineering
Dinda Septia Annisa	97	97	97	97	97	Industrial Engineering
Farhatul Fadillah	96	90	87	96	86	Civil Engineering
Misratul Husna	97	97	97	95	95	Chemical Engineering
Muamar	96	96	85	96	85	Civil Engineering
Muhammad Alwi Zahrial	95	84	84	95	83	Architectural Engineering
Muhammad Ausid Addari	97	87	87	97	97	Mechanical Engineering
Muhammad Husnul Qamar	93	93	82	95	82	Architectural Engineering
Muhammad Nazar	96	96	96	96	96	Industrial Engineering
Muhammad Rafi Akbar	97	86	86	87	97	Materials Engineering
Muhammad Ramadhan	93	93	88	83	82	Informatics Engineering
Muhammad Rusdi	95	84	82	95	82	Civil Engineering
Muhammad Saprizal	96	96	86	87	96	Materials Engineering
Anonymous	96	96	86	87	96	Materials Engineering
Anonymous	93	93	88	82	81	Informatics Engineering
Novi	97	97	97	96	96	Materials Engineering
Nur Fira	93	93	88	82	81	Informatics Engineering
Nurbaiti	93	93	82	93	81	Architectural Engineering

Here are the steps for calculating the C4.5 Algorithm:
Attributes: Subject Values for Mathematics, English, Biology, Physics and Chemistry.
Class Label: College Major (Engineering: Civil, Industrial, Chemical, Informatics, Architecture, Mechanical and Materials).

Calculating the initial entropy:

Before dividing the dataset, we calculate the initial entropy (the entropy of the entire dataset). There are 7 different majors, so we calculate the probability distribution of each major.

Major distribution:

- Civil Engineering: 3 students
- Industrial Engineering: 4 students
- Chemical Engineering: 3 students
- Information Engineering: 5 students
- Mechanical Engineering: 2 students
- Architectural Engineering: 4 students

• Materials Engineering: 3 students

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi$$

$$Entropy(S) = -(\frac{3}{24}\log_2\frac{3}{24} + \frac{4}{24}\log_2\frac{4}{24} + \frac{3}{24}\log_2\frac{3}{24} + \frac{5}{24}\log_2\frac{5}{24} + \frac{2}{24}\log_2\frac{2}{24} + \frac{4}{24}\log_2\frac{4}{24} + \frac{3}{24}\log_2\frac{3}{24})$$
$$= -(\frac{3}{24}(-3) + \frac{4}{24}(-2{,}585) + \frac{3}{24}(-3) + \frac{5}{24}(-2{,}263) + \frac{2}{24}(-3{,}585) + \frac{4}{24}(-2{,}585) + \frac{3}{24}(-3))$$
$$= 2{,}758$$

Mathematics subject:

Calculate information gain for Math attribute. We divide the scores into two categories

High scores (≥ 90)

Low scores (< 90)

Calculating Entropy of Each Branch.

(1) High branch entropy (≥ 90)

There are 24 students with high Math scores.

$$E(high) = -(\frac{3}{24}\log_2\frac{3}{24} + \frac{4}{24}\log_2\frac{4}{24} + \frac{3}{24}\log_2\frac{3}{24} + \frac{4}{24}\log_2\frac{4}{24} + \frac{2}{24}\log_2\frac{2}{24} + \frac{4}{24}\log_2\frac{4}{24} + \frac{3}{24}\log_2\frac{3}{24})$$
$$= -(\frac{3}{24}(-3) + \frac{4}{24}(-2{,}585) + \frac{3}{24}(-3) + \frac{4}{24}(-2{,}585) + \frac{2}{24}(-3{,}585) + \frac{4}{24}(-2{,}585) + \frac{3}{24}(-3))$$
$$= 2{,}758$$

(2) Low branch entropy (< 90)

There are 0 students with low Mathematics scores, so E (low) = 0

After calculating the entropy, then calculate the Mathematics attribute gain.

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

$$Gain\ (S, Mathematics) = Entropy(S) - (\frac{24}{24} * E(high) + \frac{0}{24} * E(low))$$
$$= 2{,}758 - (\frac{24}{24} * 2{,}748 + \frac{0}{24} * 0)$$
$$= 2{,}758 - (2{,}758+0)$$
$$= 0$$

After the gain is obtained, the next step is to calculate the split info and gain ratio.

$$SplitInfo(S, A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} * \log_2 \frac{|Si|}{|S|}$$

$$SplitInfo(Mathematics) = -(\frac{24}{24}\log_2(\frac{24}{24}) + \frac{0}{24}\log_2(\frac{0}{24}))$$
$$= -(\frac{24}{24}(0) + \frac{0}{24}(0))$$
$$= 0$$

And the Gain Ratio is as follows.

$$Gain\ Ratio = \frac{Gain\ (A)}{SplitInfo(S, A)} = \frac{0}{0} = 0$$

English subject:

Calculate information gain for English attribute. We divide the scores into two categories

High scores (≥ 90)

Low scores (< 90)

Calculating Entropy of Each Branch.

(1) High branch entropy (≥ 90)

There are 18 students with high English scores.

$$E(high) = -(\frac{2}{18}\log_2\frac{2}{18} + \frac{4}{18}\log_2\frac{4}{18} + \frac{2}{18}\log_2\frac{2}{18} + \frac{5}{18}\log_2\frac{5}{18} + \frac{0}{18}\log_2\frac{0}{18} + \frac{2}{18}\log_2\frac{2}{18} + \frac{3}{18}\log_2\frac{3}{18})$$
$$= -(\frac{2}{18}(-3{,}170) + \frac{4}{18}(-2{,}170) + \frac{2}{18}(-3{,}170) + \frac{5}{18}(-1{,}848) + \frac{0}{18}(0) + \frac{2}{18}(-3{,}170) + \frac{3}{18}(-2{,}585))$$
$$= 2{,}483$$

(2) Low branch entropy (< 90):

There are 6 students with low English scores.

$$E(low) = -(\frac{1}{6}\log_2\frac{1}{6} + \frac{0}{6}\log_2\frac{0}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{0}{6}\log_2\frac{0}{6} + \frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6} + \frac{0}{6}\log_2\frac{0}{6})$$
$$= -(\frac{1}{6}(-2{,}621) + \frac{0}{6}(0) + \frac{1}{6}(-2{,}621) + \frac{0}{6}(0) + \frac{2}{6}(-1{,}621) + \frac{2}{6}(-1{,}621) + \frac{0}{6}(0))$$
$$= 1{,}954$$

After calculating the entropy, then calculate the English attribute gain.

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

Gain (S, English) = *Entropy(S)* – ($\frac{18}{24}$ * E(high) + $\frac{6}{24}$ * E(low))
= 2,758 – ($\frac{18}{24}$ * 2,483 + $\frac{6}{24}$ * 1,954)
= 2,758 – (1,862+0,4885)
= 0,4075

After the gain is obtained, the next step is to calculate the split info and gain ratio.

$$SplitInfo(S,A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} * \log_2 \frac{|Si|}{|S|}$$

SplitInfo(English) = -($\frac{18}{24}\log_2(\frac{18}{24}) + \frac{6}{24}\log_2(\frac{6}{24})$)
= -($\frac{18}{24}$ (-0,415) + $\frac{6}{24}$ (-2))
= 0,811

And the Gain Ratio is as follows.

$$Gain\ Ratio = \frac{Gain\ (A)}{SplitInfo(S,A)} = \frac{0{,}4075}{0{,}811} = 0{,}502$$

Biology subject:

Calculate information gain for Biology attribute. We divide the scores into two categories

High scores (≥ 90)

Low scores (< 90)

Calculating Entropy of Each Branch.

(1) High branch entropy (≥ 90)

There are 8 students with high Biology scores.

E (high) = -($\frac{0}{8}\log_2\frac{0}{8} + \frac{3}{8}\log_2\frac{3}{8} + \frac{3}{8}\log_2\frac{3}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{0}{8}\log_2\frac{0}{8} + \frac{0}{8}\log_2\frac{0}{8} + \frac{1}{8}\log_2\frac{1}{8}$)
= -($\frac{0}{8}$ (0) + $\frac{3}{8}$ (-1,415) + $\frac{3}{8}$ (-1,415) + $\frac{1}{8}$ (-3) + $\frac{0}{8}$ (0) + $\frac{0}{8}$ (0) + $\frac{1}{8}(-3)$)
= 1,81125

(2) Low branch entropy (< 90)

There are 16 students with low Biology scores.

E (low) = -($\frac{3}{16}\log_2\frac{3}{16} + \frac{1}{16}\log_2\frac{1}{16} + \frac{0}{16}\log_2\frac{0}{16} + \frac{4}{16}\log_2\frac{4}{16} + \frac{2}{16}\log_2\frac{2}{16} + \frac{4}{16}\log_2\frac{4}{16} + \frac{2}{16}\log_2\frac{2}{16}$)
= -($\frac{3}{16}$ (-2,415) + $\frac{1}{16}$ (-4) + $\frac{0}{16}$ (0) + $\frac{4}{16}$ (-2) + $\frac{2}{16}$ (-3)+ $\frac{4}{16}$ (-2) + $\frac{2}{16}(-3)$)
= 2,45

After calculating the entropy, then calculate the Biology attribute gain.

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

Gain (S, Biology) = *Entropy(S)* – ($\frac{8}{24}$ * E(high) + $\frac{16}{24}$ * E(low))
= 2,758 – ($\frac{8}{24}$ * 1,81125 + $\frac{16}{24}$ * 2,45)
= 2,758 – (0,60375+1,6333)
= 0,52095

After the gain is obtained, the next step is to calculate the split info and gain ratio.

$$SplitInfo(S,A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} * \log_2 \frac{|Si|}{|S|}$$

SplitInfo(Biology) = -($\frac{8}{24}\log_2(\frac{8}{24}) + \frac{16}{24}\log_2(\frac{16}{24})$)
= -($\frac{8}{24}$ (-1,585) + $\frac{16}{24}$ (-0,585))
= 0,9183

And the Gain Ratio is as follows.

$$Gain\ Ratio = \frac{Gain\ (A)}{SplitInfo(S,A)} = \frac{0{,}52095}{0{,}9183} = 0{,}567$$

Physics Subject:

Calculate information gain for Physics attribute. We divide the scores into two categories

High scores (≥ 90)

Low scores (< 90)

Calculating Entropy of Each Branch.

(1) High branch entropy (≥ 90)

There are 17 students with high Physics scores.

E (high) = -($\frac{3}{17}\log_2\frac{3}{17} + \frac{4}{17}\log_2\frac{4}{17} + \frac{3}{17}\log_2\frac{3}{17} + \frac{0}{17}\log_2\frac{0}{17} + \frac{2}{17}\log_2\frac{2}{17} + \frac{4}{17}\log_2\frac{4}{17} + \frac{1}{17}\log_2\frac{1}{17}$)

$= -(\frac{3}{17}\ (-2{,}502) + \frac{4}{17}(-2{,}0875) + \frac{3}{17}(-2{,}502) + \frac{0}{17}\ (0) + \frac{2}{17}(-3{,}087) + \frac{4}{17}(-2{,}0875) + \frac{1}{17}(-4{,}0875))$
$= 2{,}46834$

(2) Low branch entropy (< 90)

There are 7 students with low Physics scores.

E (low) $= -(\frac{0}{7}\log_2\frac{0}{7} + \frac{0}{7}\log_2\frac{0}{7} + \frac{0}{7}\log_2\frac{0}{7} + \frac{5}{7}\log_2\frac{5}{7} + \frac{0}{7}\log_2\frac{0}{7} + \frac{0}{7}\ \log 2\ \frac{0}{7} + \frac{2}{7}\log_2\frac{2}{7})$
$= -(\frac{0}{7}\ (0) + \frac{0}{7}\ (0) + \frac{0}{7}\ (0) + \frac{5}{7}\ (-0{,}4855) + \frac{0}{7}\ (0) + \frac{0}{7}\ (0) + \frac{2}{7}(-1{,}8074))$
$= 0{,}86247$

After calculating the entropy, then calculate the Physics attribute gain.

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

Gain (S, Physics) = *Entropy(S)* – ($\frac{17}{24}$ * E(high) + $\frac{7}{24}$ * E(low))
$= 2{,}758 - (\frac{17}{24} * 2{,}46834 + \frac{7}{24} * 0{,}86247)$
$= 2{,}758 - (1{,}74757 + 0{,}25139)$
$= 0{,}75904$

After the gain is obtained, the next step is to calculate the split info and gain ratio.

$$SplitInfo(S, A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} * \log_2 \frac{|Si|}{|S|}$$

SplitInfo(Physics) = $-(\frac{17}{24}\log_2(\frac{17}{24}) + \frac{7}{24}\log_2(\frac{7}{24}))$
$= -(\frac{17}{24}(-0{,}498) + \frac{7}{24}(-1{,}778))$
$= 0{,}871$

And the Gain Ratio is as follows.

$$Gain\ Ratio = \frac{Gain\ (A)}{SplitInfo(S, A)} = \frac{0{,}75904}{0{,}871} = 0{,}871$$

Chemistry Subject:

Calculate information gain for Chemistry attribute. We divide the scores into two categories

High scores (≥ 90)

Low scores (< 90)

Calculating Entropy of Each Branch.

(1) High branch entropy (≥ 90)

There are 12 students with high Chemistry scores.

E (high) $= -(\frac{0}{12}\log_2\frac{0}{12} + \frac{4}{12}\log_2\frac{4}{12} + \frac{3}{12}\log_2\frac{3}{12} + \frac{0}{12}\log_2\frac{0}{12}$
$+ \frac{2}{12}\log_2\frac{2}{12} + \frac{0}{12}\log_2\frac{0}{12} + \frac{3}{12}\log_2\frac{3}{12})$
$= -(\frac{0}{12}\ (0) + \frac{4}{12}(-1{,}585) + \frac{3}{12}(-2) + \frac{0}{12}\ (0)$
$+ \frac{2}{12}(-2{,}585) + \frac{0}{12}(0) + \frac{3}{12}(-2))$
$= 1{,}95916$

(2) Low branch entropy (< 90)

There are 12 students with low English scores.

E (low) $= -(\frac{3}{12}\log_2\frac{3}{12} + \frac{0}{12}\log_2\frac{0}{12} + \frac{0}{12}\log_2\frac{0}{12} + \frac{5}{12}\log_2\frac{5}{12}$
$+ \frac{0}{12}\log_2\frac{0}{12} + \frac{4}{12}\log_2\frac{4}{12} + \frac{0}{12}\log_2\frac{0}{12})$
$= -(\frac{3}{12}\ (-2) + \frac{0}{12}(0) + \frac{0}{12}(0) + \frac{5}{12}\ (-1{,}2631) + \frac{0}{12}(0)$
$+ \frac{4}{12}(-1{,}585) + \frac{0}{12}(0))$
$= 1{,}5555$

After calculating the entropy, then calculate the Chemistry attribute gain:

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

Gain (S, Chemistry) = *Entropy(S)* – ($\frac{12}{24}$ * E(tinggi) + $\frac{12}{24}$ * E(rendah))
$= 2{,}758 - (\frac{12}{24} * 2{,}483 + \frac{12}{24} * 1{,}954)$
$= 2{,}758 - (1{,}95916 + 1{,}5555)$
$= -0{,}75666$

After the gain is obtained, the next step is to calculate the split info and gain ratio.

$$SplitInfo(S, A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} * \log_2 \frac{|Si|}{|S|}$$

SplitInfo(Chemistry) = -($\frac{12}{24}\log_2(\frac{12}{24}) + \frac{12}{24}\log_2(\frac{12}{24})$)

= -($\frac{12}{24}$ (-1) + $\frac{12}{24}$ (-1)) = 1

And the Gain Ratio is as follows.

$$Gain\ Ratio = \frac{Gain\ (A)}{SplitInfo(S,A)} = \frac{-0{,}75666}{1} = -0{,}756$$

The following table shows the calculation results for all attributes:

Tabel 2. calculation results

Atribut	*Split Point*	*Gain*	*Entropy*	*Split Info*	*Gain Ratio*
Mathematics		0		0	0
	≥ 90		2,758		
	< 90		0		
English		0,4075		0,811	0,502
	≥ 90		2,483		
	< 90		1,954		
Biology		0,52095		0,9183	0,567
	≥ 90		1,81125		
	< 90		2,45		
Physics		0,75904		0,871	0,871
	≥ 90		2,46834		
	< 90		0,86247		
Chemistry		-0,75666		1	-0,756
	≥ 90		1,95916		
	< 90		1,5555		

From the table above, it can be seen that the highest gain ratio value that will be used as the first root is the Physics subject with a value of 0.871.

Conclusions

Based on the results of the research and discussion that have been discussed in the previous chapter, the conclusions that can be drawn from the decision tree application system to determine the study majors of post-Islamic Senior High School students Misbahul Ulum are as follows:

(1) This research has succeeded in designing a system to classify study majors of post-Islamic Senior High School students Misbahul Ulum. This system was developed using the PHP programming language.

(2) The C4.5 algorithm has been successfully implemented in the system to determine students' study majors. This algorithm is able to process numeric and categorical data so that it is suitable for analyzing student grades.

(3) This system makes it easy for users to determine students' study majors, so that it can support decision making in planning for further study.

Suggestion

Here are some suggestions from researchers that are expected to be input for the improvement and development of the system in the future.

(1) Add a broader category of college majors so that students can be more diverse in determining college majors.

(2) Conduct trials on various different data sets to ensure the model can adapt to various conditions and scenarios.

Acknowledgments

In the process of this research, the author sometimes experienced several obstacles and difficulties, but thanks to the guidance, assistance, direction and encouragement from several parties, the author would like to express his deepest gratitude to:

1. Prof. Dr. H. Herman Fithra, S.T., M.T., IPM., ASEAN.Eng. as the Chancellor of Malikussaleh University

2. Mr. Bustami, S.Si., M.Si., M.Kom as the main supervisor and Mr. Sujacka Retno, S.T., M.Kom as the assistant supervisor who have provided a lot of guidance, provided criticism, suggestions and very useful motivation.

References

[1] Badrul, M. (2021). Penerapan Metode Profile Matching Untuk Menunjang Keputusan Seleksi Pegawai Baru. PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer, 8(1), 75–82. https://doi.org/10.30656/prosisko.v8i1.2815.

[2] Darnila, E., Al-Kautsar, H., & Iksan, Y. (2020). Analysis of Patient Disease Trends Based on Medical Record Data Using the C4. 5 Algorithm. Login: Jurnal Teknologi …, 14(1), 7–12. http://www.login.seaninstitute.org/index.php/Login/article/view/11%0Ahttp://www.login.seaninstitute.org/index.php/Login/article/download/11/18.

[3] Devia, E. (2023). Penerapan Decision Tree Dengan Algoritma C4.5 Untuk Menentukan Rekomendasi Kenaikan Jabatan Karyawan. Jurnal Information System, 3(1), 28–37. https://doi.org/10.61488/jis.v3i1.257.

[4] Effendy, E., Siregar, E. A., Fitri, P. C., & Damanik, I. A. S. (2023). Mengenal Sistem Informasi Manajemen Dakwah (Pengertian Sistem, Karakteristik Sistem). Jurnal Pendidikan Dan Konseling, 5(2), 4343–4349.

[5] Fikry, M., Rizal, Fadlisyah, Nurdin, Bustami, Hamdhana, D., & Qamal, M. (2020). Data Mining for Processing of Research and Community Service by Lecturer Using Decision Tree Method. International Journal of Psychosocial Rehabilitation, Conference Special Issue, 24(1475–7192), 367–371.

[6] Kadir, A. M., Yahya, M., & Mappalotteng, A. M. (2021). Pengembangan Sistem Rekomendasi Pemilihan Program Studi di Universitas Negeri Makassar Menggunakan Metode Decision Tree. Seminar Nasional LP2M UNM, 2750–2581.

[7] Tamir, M. A. (2019). Perancangan Aplikasi Data Mining Menggunakan Metode Decision Tree Untuk Analisa Pemberian Kredit Pada Bri Unit LalabataRilau. Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika, 2(1), 1–10.

[8] Yahya, V. Y., Yudhistira, F., & Manurung, L. (2023). Perancangan Sistem Informasi Pemesanan Sparepart Pengolahan Kelapasawit Pada Cv Ltsu Berbasis Java. Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi), 7(1), 711–717. https://doi.org/10.30998/semnasristek.v7i1.6404.