Implementation of the C45 Algorithm in Classifying Classes

Nellysa Putri Denila Nasution, Fristi Riandari
Informatics Engineering, STMIK Pelita Nusantara, Sumatera Utara, Indonesia

ARTICLE INFO

ABSTRACT

Mts Swasta YPII Kotarih is one of the educational institutions which in its implementation classifies several classes for students, one of which is the superior class. However, in practice, each person is still selected based on the rank in each class, making the classification process less accurate and efficient, a computerized classification technique is needed in its implementation, the classification technique used by Data Mining is the C4.5 algorithm that works with the Decision tree to find the highest gain value from each criterion existing criteria, the criteria used in this study are the subjects of Q. Hadith, Akidah, Fiqh, SKI, Arabic, Tahfizh, General research results show that from there are 3 criteria that are most influential in the classification process, namely General, Arabic, Fiqh, implemented in the form of a computerized system. Based on the research results, the highest gain value is the highest gain is General with a value of 0.3629, so that the general attribute becomes node 1 or the root of the decision tree, second highest gain B. Arabic attribute, with a gain value of 1.7189, and the third highest gain is Fiqh with a value of 0.3486.

This is an open access article under the CC BY-NC license.

Corresponding Author:
Fristi Riandari,
Informatics Engineering,
STMIK Pelita Nusantara, Sumatera Utara, Indonesia
Jl. Iskandar Muda No.1 Medan, Sumatera Utara, Indonesia
Email: fristy.rianda@ymail.com

1. INTRODUCTION

Mts Swasta YPII Kotarih is one of the educational institutions which in its implementation classifies several classes for students, one of which is the superior class. The industrial revolution 4.0 is the latest advancement in the world of technology that changes the way human resources work, putting aside the role of humans and replacing them with digital technology (Arta et al., 2017)(Asroni et al., 2018). Data mining is often called knowledge discovery in database (KDD), an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets (Suntoro, 2019). C4.5 algorithm is a tree structure in which nodes describe the attributes, each branch describes the results of the tested attributes, and each leaf describes the class(Saputra et al., 2016).

The qualifications of each school are different in terms of the things they do, so the YPII Kotarih private MTS must specify which students will be classified based on class. This research was conducted to classify students based on the scores achieved from each student, using the C45 algorithm classification method, which uses a decision tree (SAWITRI, 2019), (Sutoyo, 2018).

Based on previous research (Septiani, 2017)(Handoko, 2016)[15]. Analysis and Application of the C4.5 Method for Predicting Customer Loyalty. Essay. shows that the accuracy reaches 97.5%, which indicates that the C4.5 algorithm is suitable for measuring cellular data customer loyalty (Putra & Wadisman, 2018)(Renhoran et al., 2018). Meanwhile, Algorithm classification method C4.5 for
Credit Eligibility Prediction at Bank Mayapada Jakarta Industry from this research, the model formed with the C4.5 algorithm itself already has a good accuracy of 83.67% with the selection process attributes by the C4.5 algorithm (Karsito & Sari, 2019)(Nasrullah, 2018).

The goal is to classify excellent students who will enter the superior class at the Private MTS YPII Kotarih, improve the accuracy of the selection of superior students according to existing criteria (ISWATI, n.d.). of this research As an alternative in classifying superior classes at YPII Kotarih Private MTS and can be used as a reference source for other researchers (Al Syahdan & Sindar, 2018).

2. METHOD
2.1 Research Framework

The framework discusses the research implementation model in classifying superior classes, the following stages:

- **Identifying Problems**
- **Data Collection**
- **Data Analysis**
- **Data Mining**
- **System Planning**
- **System Testing**

![Fig 1. Research Framework](image)

a. **Problem Identification**

At this stage, identification of the problem will be carried out, which is found, namely how to make the right decisions in classifying students into superior classes.

b. **Data Collection**

In this study the authors get data directly from the object of research through observations made, these observations are made frankly to the data source, that the author is conducting research, the author also uses passive particion observations, namely the researcher comes to the researcher's place but is not involved in the activity. at the research site to obtain complete data, the results of the data obtained are in the form of student names, NISN, value of subjects.

c. **Data Analysis**

At this stage, the inspection and cleaning process is carried out, with the aim of finding useful information, using qualitative data analysis, namely after the data obtained through interviews, the data needed are grouped, at this stage the data obtained is the data of grade 7 students in the odd semester of the 2020 academic year. /2021.

d. **Data Mining**

The use of large data to find patterns, in this study using the C4.5 algorithm Identification of problems. Includes stages : selection, preprocessing, transformation, data mining, interpretation (Karsito & Sari, 2019). There are several stages in making a decision tree in the C4.5 algorithm (Handoko, 2016), (Iriadi & Nuraeni, 2016):

1) Prepare training data
2) Calculating the root of the tree, first calculate the entropy value, to calculate the entropy value is used the formula:

\[
\text{Entropy} (S) = \sum_{i}^{n} = 1 - p_{i} \log_{2} (p_{i})
\]

Information:

S = Case set
\( n \) = Number of partitions  
\( p_i \) = Proportion \( p_i \) to \( S \)

3) Calculating the Gain value using Equation 2

\[
Gain(S, A) = entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \text{Entropy} (S_i)
\]

Information:
- \( S \) = Case set  
- \( A \) = feature  
- \( n \) = Number of attribute partitions \( A \)  
- \(|S_i|\) = Proportion \( S_i \) to \( S \)  
- \(|S|\) = number of case in \( S \)

4) Repeat step 2 and step 3 until all records are partitioned

5) The decision tree partitioning process will stop when:  
- Records in node \( N \) have the same class and there are no attributes in the partitioned record anymore.

3. RESULT And DISCUSSION

There are 7 criteria used for the classification of the superior class students, these criteria are set by MTS Swasta YPII Kotarih, namely: Q. Hadith, Akidah, Fiqih, SKI, Tahfizh, B. Arabic, General. After the data criteria are available, the existing values will be changed into categories that have been set by MTS Swasta YPII Kotarih:

<table>
<thead>
<tr>
<th>Number</th>
<th>Value</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9 – 100</td>
<td>Sangat Baik</td>
</tr>
<tr>
<td>2</td>
<td>83 – 90</td>
<td>Baik</td>
</tr>
<tr>
<td>3</td>
<td>70-82</td>
<td>Cukup</td>
</tr>
<tr>
<td>4</td>
<td>≤ 69</td>
<td>Buruk</td>
</tr>
</tbody>
</table>

This research was conducted at MTS Swasta YPII Kotarih, this study used passive partition observations. The data obtained from the data of class VII students in the odd semester of the 2020/2021 academic year, the number of student data taken as a sample is 142 people. The output in this study is divided into two categories, namely Excellent Students and Not. Based on the output to be generated, the authors use classification techniques.

3.1 Data Selection

At this stage the data that has been obtained will be selected according to the needs of the research.

The data used is grade 7 odd semester for the 2020/2021 academic year, the data obtained is 142 data.

<table>
<thead>
<tr>
<th>Num</th>
<th>Name</th>
<th>NISN</th>
<th>Q.Hadith</th>
<th>Akidah</th>
<th>Fiqih</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aldyansah Ramad</td>
<td>0087072669</td>
<td>87</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>Alfiath Fatwa</td>
<td>0086122120</td>
<td>85</td>
<td>86</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>Andran Syahputra</td>
<td>0072392934</td>
<td>85</td>
<td>87</td>
<td>85</td>
</tr>
<tr>
<td>4</td>
<td>Andri Iriyanto</td>
<td>0075216135</td>
<td>85</td>
<td>86</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>Andrian sentosa</td>
<td>008233829</td>
<td>86</td>
<td>88</td>
<td>85</td>
</tr>
<tr>
<td>6</td>
<td>Deniel ma’ruf</td>
<td>0085198547</td>
<td>87</td>
<td>87</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>Deswita Indah</td>
<td>0073094706</td>
<td>88</td>
<td>88</td>
<td>87</td>
</tr>
<tr>
<td>8</td>
<td>Diki irwandi</td>
<td>0089128545</td>
<td>81</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>9</td>
<td>Dinda priska dewi</td>
<td>0063241235</td>
<td>85</td>
<td>87</td>
<td>86</td>
</tr>
<tr>
<td>10</td>
<td>Efranda gunawan</td>
<td>0085240512</td>
<td>85</td>
<td>86</td>
<td>83</td>
</tr>
</tbody>
</table>
3.3 Pre-processing
The process of data preparation by cleaning inconsistent data and noise. At this stage all student
data is used because inconsistent data and noise are not found.

3.4 Transformation
The data will be converted into an assessment according to predetermined criteria. The results
of the research carried out are superior and not superior students based on the information
Furthermore, after the data has been cleaned, it enters the transformation stage, where the data will
be changed or combined into the assessment according to the predetermined value of the
assessment indicator.

<table>
<thead>
<tr>
<th>TABLE 6. VALUE ASSESSMENT CRITERIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranged</td>
</tr>
<tr>
<td>1-10</td>
</tr>
<tr>
<td>&gt;10</td>
</tr>
</tbody>
</table>

For the attributes of price, taste and service have the same assessment criteria. After the
questionnaire data is cleaned, then the data is then transformed from each criterion which can be
seen in the following table.

<table>
<thead>
<tr>
<th>TABLE 7. TRANSFORMATION DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

3.5 Data Mining
Processing the data Abundance becomes important information [2], for each attribute To
calculate the gain using the formula :

\[
Gain(S, A) = \text{entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \text{entropy}(S_i)
\]

While to calculate entropy :

\[
\text{Entropy}(S) = \sum_{i=1}^{n} - p_i \log_2(p_i)
\]

Entropy (Total) with the following formula :

\[
\text{Entropy (total)} = \sum_{i=1}^{n} - P_i \times \log_2 P_i
\]

Calculating the Entropy Value of each attribute :

\[
\text{Entropy (SB)} = \left( \frac{27}{43} \times \log_2 \left( \frac{27}{43} \right) \right) + \left( \frac{16}{43} \times \log_2 \left( \frac{16}{43} \right) \right) = 0.9522
\]

The entropy (total) calculates the total value of the leading decisions of 27 and 16 unseeded, while
43 is the total number.

a. Q.Hadist

\[
\text{Entropy (SB)} = \left( \frac{12}{28} \times \log_2 \left( \frac{12}{28} \right) \right) + \left( \frac{20}{28} \times \log_2 \left( \frac{20}{28} \right) \right) = 0.9666
\]

b. Akidah

\[
\text{Entropy (C)} = \left( \frac{7}{12} \times \log_2 \left( \frac{7}{12} \right) \right) + \left( \frac{5}{12} \times \log_2 \left( \frac{5}{12} \right) \right) = 0.9798.
\]
Entropy(SB) = \( \frac{1}{2} \times \log_2(\frac{26}{40}) + \frac{1}{2} \times \log_2(\frac{14}{40}) = 0 \)

Entropy(B) = \( \frac{26}{40} \times \log_2(\frac{26}{40}) + \frac{14}{40} \times \log_2(\frac{14}{40}) = 0.9340 \)

Entropy(C) = \( \frac{9}{31} \times \log_2(\frac{9}{31}) + \frac{22}{31} \times \log_2(\frac{22}{31}) = 0 \)

c. Fiqih
Entropy(SB) = \( \frac{5}{6} \times \log_2(\frac{5}{6}) + \frac{1}{6} \times \log_2(\frac{1}{6}) = 0.6500 \)

Entropy(B) = \( \frac{22}{31} \times \log_2(\frac{22}{31}) + \frac{9}{31} \times \log_2(\frac{9}{31}) = 0.8691 \)

Entropy(C) = \( \frac{5}{12} \times \log_2(\frac{5}{12}) + \frac{7}{12} \times \log_2(\frac{7}{12}) = 0.9871 \)

d. SKI
Entropy(SB) = \( \frac{1}{2} \times \log_2(\frac{1}{2}) + \frac{1}{2} \times \log_2(\frac{1}{2}) = 1 \)

Entropy(B) = \( \frac{9}{24} \times \log_2(\frac{9}{24}) + \frac{5}{24} \times \log_2(\frac{5}{24}) = 0.7382 \)

Entropy(C) = \( \frac{5}{19} \times \log_2(\frac{5}{19}) + \frac{14}{19} \times \log_2(\frac{14}{19}) = 0.9819 \)

e. B. Arab
Entropy(SB) = \( \frac{0}{6} \times \log_2(\frac{0}{6}) + \frac{1}{6} \times \log_2(\frac{1}{6}) = 0 \)

Entropy(B) = \( \frac{19}{24} \times \log_2(\frac{19}{24}) + \frac{8}{24} \times \log_2(\frac{8}{24}) = 0.9871 \)

Entropy(C) = \( \frac{19}{24} \times \log_2(\frac{19}{24}) + \frac{15}{24} \times \log_2(\frac{15}{24}) = 0.9871 \)

f. Tahfizh
Entropy(SB) = \( \frac{2}{3} \times \log_2(\frac{2}{3}) + \frac{1}{3} \times \log_2(\frac{1}{3}) = 0 \)

Entropy(B) = \( \frac{7}{10} \times \log_2(\frac{7}{10}) + \frac{3}{10} \times \log_2(\frac{3}{10}) = 0.8812 \)

Entropy(C) = \( \frac{17}{30} \times \log_2(\frac{17}{30}) + \frac{13}{30} \times \log_2(\frac{13}{30}) = 0.9871 \)

g. Umum
Entropy(SB) = \( \frac{4}{4} \times \log_2(\frac{4}{4}) + \frac{0}{4} \times \log_2(\frac{0}{4}) = 0 \)

Entropy(B) = \( \frac{16}{17} \times \log_2(\frac{16}{17}) + \frac{1}{17} \times \log_2(\frac{1}{17}) = 0.3227 \)

Entropy(C) = \( \frac{7}{22} \times \log_2(\frac{7}{22}) + \frac{15}{22} \times \log_2(\frac{15}{22}) = 0.9456 \)

After that calculate the gain of each attribute:

a. Gain(Total, Q.Hadist)
\[ = \text{Entropy}(\text{Total}) - \sum_{i=1}^{n} \frac{[\text{Q.Hadist}_i]}{\text{Total}} \times \text{Entropy}(\text{Q.Hadist}_i) \]
\[ = 0.9522 - (\frac{3}{43} \times 0) + (\frac{28}{43} \times 0.9661) + (\frac{29}{43} \times 0.9798) = 0.0493 \]

b. Gain(Total, Akidah)
\[ = \text{Entropy}(\text{Total}) - \sum_{i=1}^{n} \frac{[\text{Akidah}_i]}{\text{Total}} \times \text{Entropy}(\text{Akidah}_i) \]
\[ = 0.9522 - (\frac{1}{43} \times 0) + (\frac{40}{43} \times 0.9340) + (\frac{2}{43} \times 0) = 0.0833 \]

c. Gain(Total, Fiqih)
\[ = \text{Entropy}(\text{Total}) - \sum_{i=1}^{n} \frac{[\text{Fiqih}_i]}{\text{Total}} \times \text{Entropy}(\text{Fiqih}_i) \]
\[ = 0.9522 - (\frac{6}{43} \times 0.6500) + (\frac{31}{43} \times 0.8691) + (\frac{9}{43} \times 0) = 0.2349 \]

d. Gain(Total, SKI)
\[ = \text{Entropy}(\text{Total}) - \sum_{i=1}^{n} \frac{[\text{SKI}_i]}{\text{Total}} \times \text{Entropy}(\text{SKI}_i) \]
\[ = 0.9522 - (\frac{2}{43} \times 1) + (\frac{11}{43} \times 0.6840) + (\frac{30}{43} \times 0.9871) = 0.0420 \]
e. Gain(Total, B.Arab)  

\[ \text{Gain}(\text{Total, B.Arab}) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{[\text{B.Arab}_i] * \text{Entropy}(\text{B.Arab}_i)}{[\text{Total}]} \]

\[ = 0.9522 - (\frac{0}{43} * 0) + (\frac{24}{43} * 0.7382) + (\frac{10}{43} * 0) = 0.9819 \]

f. Gain(Total, Tahfizh)

\[ \text{Gain}(\text{Total, Tahfizh}) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{[\text{Tahfizh}_i] * \text{Entropy}(\text{Tahfizh}_i)}{[\text{Total}]} \]

\[ = 0.9522 - (\frac{4}{43} * 0) + (\frac{17}{43} * 0.3227) + (\frac{22}{43} * 0.9456) = 0.3408 \]

From these calculations, the highest gain is General with a value of 0.3629, then node 1 is General, following Decision Tree node 1:

![Decision Tree node 1.2](image)

Fig 2. Decision Tree node 1.2

Next, node 1.2 is calculated as the root for the Entropy, Enough value of all Q. Hadith, Akidah, Fiqh, SKI, B.Arabic, and Tahfiz attributes, the same as the previous calculation, then calculates Entropy (Sufficient General Value), with the formula:

\[ \text{Entropy (total)} = - \sum_{i=1}^{n} p_i \text{log}_2(p_i) \]

\[ = (\frac{22}{31} \text{log}_2(\frac{22}{31})) + (\frac{9}{31} \text{log}_2(\frac{9}{31})) = 0.8691 \]

a. Atribut Q.Hadist

\[ \text{Entropy (SB)} = (\frac{2}{2} * \text{log}_2(\frac{2}{2})) + (\frac{1}{2} * \text{log}_2(\frac{1}{2})) = 0 \]

\[ \text{Entropy (B)} = (\frac{18}{25} * \text{log}_2(\frac{18}{25})) + (\frac{2}{25} * \text{log}_2(\frac{2}{25})) = 0.855 \]

\[ \text{Entropy (C)} = (\frac{5}{2} * \text{log}_2(\frac{5}{2})) + (\frac{1}{2} * \text{log}_2(\frac{1}{2})) = 1 \]

b. Atribut Akidah

\[ \text{Entropy (SB)} = (\frac{0}{6} * \text{log}_2(\frac{0}{6})) + (\frac{5}{6} * \text{log}_2(\frac{5}{6})) = 0 \]

\[ \text{Entropy (B)} = (\frac{20}{29} * \text{log}_2(\frac{20}{29})) + (\frac{9}{29} * \text{log}_2(\frac{9}{29})) = 0.7973 \]

\[ \text{Entropy (C)} = (\frac{0}{2} * \text{log}_2(\frac{0}{2})) + (\frac{2}{2} * \text{log}_2(\frac{2}{2})) = 0 \]

c. Atribut Fiqh

\[ \text{Entropy (SB)} = (\frac{5}{6} * \text{log}_2(\frac{5}{6})) + (\frac{1}{6} * \text{log}_2(\frac{1}{6})) = 0.6500 \]

\[ \text{Entropy (B)} = (\frac{10}{19} * \text{log}_2(\frac{10}{19})) + (\frac{9}{19} * \text{log}_2(\frac{9}{19})) = 0.4854 \]

\[ \text{Entropy (C)} = (\frac{6}{6} * \text{log}_2(\frac{6}{6})) = 0 \]

d. Atribut SKI

\[ \text{Entropy (SB)} = (\frac{0}{11} * \text{log}_2(\frac{0}{11})) + (\frac{1}{11} * \text{log}_2(\frac{1}{11})) = 0 \]

\[ \text{Entropy (B)} = (\frac{8}{11} * \text{log}_2(\frac{8}{11})) + (\frac{3}{11} * \text{log}_2(\frac{3}{11})) = 0.8453 \]
Entropy (C)=\(\frac{14}{19} \times \log_2(\frac{14}{19}) + \frac{5}{19} \times \log_2(\frac{5}{19})\) = 0.8314

e. Atribut B.Arab

Entropy (SB) = \(\frac{6}{14} \times \log_2(\frac{6}{14}) + \frac{8}{14} \times \log_2(\frac{8}{14})\) = 0

Entropy (B) = \(\frac{16}{17} \times \log_2(\frac{16}{17}) + \frac{1}{17} \times \log_2(\frac{1}{17})\) = 0.7382

Entropy (C) = \(\frac{6}{14} \times \log_2(\frac{6}{14}) + \frac{8}{14} \times \log_2(\frac{8}{14})\) = 0.9852

f. Atribut Tahfizh

Entropy (SB) = \(\frac{1}{3} \times \log_2(\frac{1}{3}) + \frac{2}{3} \times \log_2(\frac{2}{3})\) = 0

Entropy (B) = \(\frac{4}{7} \times \log_2(\frac{4}{7}) + \frac{3}{7} \times \log_2(\frac{3}{7})\) = 0.7219

Entropy (C) = \(\frac{13}{25} \times \log_2(\frac{13}{25}) + \frac{6}{25} \times \log_2(\frac{6}{25})\) = 0.904

Calculating the gain value for each attribute:

a. Gain(Total, Q.Hadist)

\[ Gain(Total, Q.Hadist) = \frac{\text{Entropy}(\text{Total}) - \sum_{i=1}^{\text{Q.Hadist}} \text{Entropy}(\text{Q.Hadist}_i)}{n} \]

\[ = \frac{0.8691 - (\frac{6}{31} \times 0 + \frac{25}{31} \times 0.855 + \frac{4}{31} \times 1)}{31} = 0.0522 \]

b. Gain(Total, Akidah)

\[ Gain(Total, Akidah) = \frac{\text{Entropy}(\text{Total}) - \sum_{i=1}^{\text{Akidah}} \text{Entropy}(\text{Akidah}_i)}{n} \]

\[ = \frac{0.8691 - (\frac{5}{31} \times 0 + \frac{29}{31} \times 0.7973 + \frac{2}{31} \times 0)}{31} = 0.12325 \]

c. Gain(Total, Fiqih)

\[ Gain(Total, Fiqih) = \frac{\text{Entropy}(\text{Total}) - \sum_{i=1}^{\text{Fiqih}} \text{Entropy}(\text{Fiqih}_i)}{n} \]

\[ = \frac{0.8691 - (\frac{6}{31} \times 0 + \frac{20}{31} \times 0.8453 + \frac{5}{31} \times 0)}{31} = 0.04457 \]

d. Gain(Total, SKL)

\[ Gain(Total, SKL) = \frac{\text{Entropy}(\text{Total}) - \sum_{i=1}^{\text{SKL}} \text{Entropy}(\text{SKL}_i)}{n} \]

\[ = \frac{0.8691 - (\frac{1}{31} \times 0 + \frac{11}{31} \times 0.8453 + \frac{19}{31} \times 0.8314)}{31} = 0.0595 \]

e. Gain(Total, B.Arab)

\[ Gain(Total, B.Arab) = \frac{\text{Entropy}(\text{Total}) - \sum_{i=1}^{\text{B.Arab}} \text{Entropy}(\text{B.Arab}_i)}{n} \]

\[ = \frac{0.8691 - (\frac{6}{31} \times 0 + \frac{17}{31} \times 0.7382 + \frac{14}{31} \times 0.9852)}{31} = 1.71894 \]

f. Gain(Total, Tahfizh)

\[ Gain(Total, Tahfizh) = \frac{\text{Entropy}(\text{Total}) - \sum_{i=1}^{\text{Tahfizh}} \text{Entropy}(\text{Tahfizh}_i)}{n} \]

\[ = \frac{0.8691 - (\frac{1}{3} \times 0 + \frac{12}{31} \times 0.7219 + \frac{17}{31} \times 0.9043)}{31} = 0.0233 \]

The highest gain attribute is used as node 1.2, namely the B.Arab attribute, with a gain value of 1.7189. Therefore B.Arabic becomes node 1.2. The decision tree can be seen in the following: After successfully determining Node 1.2 then look for Node 1.3, as follows:

\[ Entropy (total) = (\frac{25}{44} \times \log_2(\frac{25}{44}) + (\frac{19}{44} \times \log_2(\frac{19}{44})) = 0.986 \]

The entropy (total) calculates the total value of the leading decisions of 25 and 19 unseeded, while 44 is the total number.

a. Q.Hadist

\[ Entropy(SB) = \frac{2}{3} \times \log_2(\frac{2}{3}) + \frac{1}{3} \times \log_2(\frac{1}{3}) = 0 \]

\[ Entropy(B) = \frac{14}{27} \times \log_2(\frac{14}{27}) + \frac{13}{27} \times \log_2(\frac{13}{27}) = 0.9990 \]
After that calculate the gain of each attribute:

a. **Gain(Total, Q.Hadist)**

\[
\text{Gain(Total, Q.Hadist)} = \sum_{i=1}^{n} \frac{|Q.Hadist_i| \times \text{Entropy}(Q.Hadist_i)}{\text{Total}}
\]

= 0.9865 - ((\frac{3}{44} \times 0) + (\frac{22}{44} \times 0.9990) + (\frac{14}{44} \times 0.9852)) = 0.0600

b. **Gain(Total, Akidah)**

\[
\text{Gain(Total, Akidah)} = \sum_{i=1}^{n} \frac{|Akidah_i| \times \text{Entropy}(Akidah_i)}{\text{Total}}
\]

= 0.9865 - ((\frac{1}{44} \times 0) + (\frac{41}{44} \times 0.9788) + (\frac{2}{44} \times 0)) = 0.0744

c. **Gain(Total, Fiqih)**

\[
\text{Gain(Total, Fiqih)} = \sum_{i=1}^{n} \frac{|Fiqih_i| \times \text{Entropy}(Fiqih_i)}{\text{Total}}
\]

= 0.9865 - ((\frac{6}{44} \times 0.6500) + (\frac{28}{44} \times 0.8631) + (\frac{10}{44} \times 0)) = 0.3486

d. **Gain(Total, SKI)**

\[
\text{Gain(Total, SKI)} = \sum_{i=1}^{n} \frac{|SKI_i| \times \text{Entropy}(SKI_i)}{\text{Total}}
\]

= 0.9865 - ((\frac{2}{44} \times 0) + (\frac{11}{44} \times 0.9456) + (\frac{31}{44} \times 0.9811)) = 0.0588

e. **Gain(Total, B.Arab)**
\[ \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|B_{\text{Arab}}|}{|\text{Total}|} \ast \text{Entropy}(B_{\text{Arab}}) \]
\[ = 0.9865 - ((\frac{0.5}{44} \ast 0) + (\frac{2.7}{44} \ast 0.8256) + (\frac{1.7}{44} \ast 0.8739)) = 0.1422. \]
f. \[ \text{Gain}(\text{Total, Tahfizh}) \]
\[ = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|Tahfizh|}{|\text{Total}|} \ast \text{Entropy}(Tahfizh) \]
\[ = 0.9865 - ((\frac{3}{44} \ast 0) + (\frac{12}{44} \ast 1.02626) + (\frac{29}{44} \ast 0.9705)) = 0.0669 \]
g. \[ \text{Gain}(\text{Total, Umum}) \]
\[ = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|Umum|}{|\text{Total}|} \ast \text{Entropy}(Umum) \]
\[ = 0.9865 - ((\frac{4}{44} \ast 0) + (\frac{18}{44} \ast 0.6500) + (\frac{22}{44} \ast 0.8453)) = 0.2979. \]
From these calculations, it can be seen that the attribute with the highest gain is Fiqh with a value of 0.3486, then node 1.3 is Fiqh, as follows:

Fig 3. Decion tree node 1.3

Evaluation : correction of pattern obtained, so that it can be identified Is the pattern already representative, knowledge to be achieved [9].

The rules or rules formed based on the last decision tree as shown in Figure, 19 rules:
1) IF General = Very Good THEN decision = Excellent
2) General IF = Enough AND Fiqh = Very Good AND Arabic = Good THEN decision = Excellent
3) General IF = Enough AND Fiqh = Very Good AND Arabic = Enough THEN decision = Not Excellent
4) General IF = Enough AND Fiqh = Good AND Aqidah = Good AND Arabic = Good AND SKI = Good THEN decision = Excellent
5) IF General = Enough AND Fiqh = Good AND Aqidah = Good AND B.Arabic = Good AND SKI = Enough AND Q. Hadith = Good AND Tahfizh = Good THEN Decision = Excellent
6) IF General = Enough AND Fiqh = Good AND Aqidah = Good AND B.Arabic = Good AND SKI
= Enough
AND Q. Hadith = Good AND Tahfizh = Enough THEN decision = Not Excellent
7) General IF = Enough AND Fiqh = Good AND Akidah = Good AND Arabic = Good AND SKI = Enough AND Q. Hadith = Enough THEN decision = Not Excellent
8) IF General = Enough AND Fiqh = Good AND Akidah = Good AND Arabic = Enough THEN decision = Not Excellent
9) General IF = Enough AND Fiqh = Good AND Aqeedah = Enough THEN decision = Not superior
10) General IF = Enough AND Fiqh = Enough THEN decision = Not Superior
11) IF General = Good AND Arabic = Enough AND SKI = Very Good THEN decision = Excellent
12) General IF = Good AND Arabic = Enough AND SKI = Good THEN decision = Superior
13) General IF = Good AND B. Arabic = Enough AND SKI = Enough AND Q. Hadith = Good AND Akidah = Good AND Fiqh = Good AND Tahfizh = Good THEN decision = Superior
14) IF General = Good AND Fiqh = Very Good THEN decision = Excellent
15) IF General = Good AND Fiqh = Good THEN decision = Excellent
16) IF General = Good AND Fiqh = Enough THEN decision = Not Superior
17) IF General = Fair AND SKI = Very Good THEN decision = Excellent
18) IF General = Enough AND SKI = Good THEN decision = Excellent
19) General IF = Enough AND SKI = Enough THEN decision = Not Excellent
IF Umum = Very Good, THEN Decision = superior

4 CONCLUSION

Based on the results of the decision tree above, there are 19 rules that were tested, the data testing was based on the 19 rules that had been set, from 142 data tested 73.94% : 26.06%, 105 data were appropriate and 37 data did not match. Application of the C4.5 Algorithm in classifying using a decision tree. The C45 algorithm is used to analyze student assessments in semester 7 of 2020/2021, the student data is applied to the KDD stages starting from data selection, Pre-Processing, Transformation, Data Mining, and Evaluation, by classifying which students are superior students, so that placement students who enter the superior class according to the achievements achieved.

References


SAWITRI, D. (2019). Revolusi Industri 4.0: Big Data Menjawab Tantangan Revolusi Industri 4.0. *Jurnal Ilmiah MAKSITEK, 4*(3).

