KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
VOL. 21 (1) (2024), 01-14 p-ISSN: 693-7554, e-ISSN:2654-3990

Comparative Analysis Accuracy ID3 Algorithm and C4.5 Algorithm in S election of Candidates Basic Physics Laboratory Assistant

Wahyu Supriyatin[1], **Yasman Rianto**[2]

[1]Information Systems Study Program, Faculty of Computer Science and Information Technology, Gunadarma University, West Java, 16424, Indonesia
[2]Computer Systems Study Program, Faculty of Computer Science and Information Technology, Gunadarma University, West Java, 16424, Indonesia

Abstract

Basic Physics Laboratory is one of the supporting laboratories at Gunadarma University. Each practical activity in the laboratory is supervised by respective assistants. Therefore, a support system is needed as a basis for decision-making in determining assistant candidates. This decision-making process is processed using data mining techniques, specifically classification algorithms. The criteria or attributes used in the decision-making process include written test scores, practical test scores, presentation scores, equipment usage abilities, and interviews. The classification algorithms used in this research are ID3 and C4.5 algorithms. The tools used to implement these algorithms are RapidMiner Studio 9.10. These algorithms will generate decision trees that can be used as decision support. The aim of this research is to conduct an accuracy comparison analysis for the ID3 and C4.5 algorithms. The highest accuracy obtained will be used as a reference for determining whether assistant candidates are accepted or not. The accuracy results show that the C4.5 algorithm has the highest accuracy, precision, and recall compared to the ID3 algorithm. The determination of the highest value is done using the k-fold cross-validation model for values 2, 4, 6, 8, and 10. The C4.5 algorithm has the highest accuracy of 96.67% at k-fold value = 2.

Keywords: *Analysis, Classification, C4.5 Algorithm, Data Mining, ID3 Algorithm*

1. Introduction

Basic Physics Laboratory is one of the laboratories at Gunadarma University. The Basic Physics Laboratory is used as a supporting laboratory for the Basic Physics Practicum course for students from the Faculty of Computer Science & Information Technology and the Faculty of Industrial Technology. Every semester, the Basic Physics Laboratory always conducts practical activities where each module is always supervised by one assistant. The assistants assigned to the Basic Physics Laboratory are active students who have completed at least 4 semesters. The accepted assistants have gone through a series of recruitment processes. The assistant recruitment process is carried out by the laboratory every year. The assistant recruitment process consists of two stages. The first stage is the selection process based on academic records and GPA. The second

*Corresponding author: ***E-mail address:*** *ayu_ ws@staff.gunadarma.ac.id*
Received: *12 Nov 2023*, Accepted: *05 Dec 2023* and available online *30 Jan 2024*
//DOI: https://doi.org/10.33751/komputasi.v21i1.9198

stage consists of a written test, practical test, module presentation, equipment usage skills, and interview. The second stage is the most important stage in the assistant recruitment process, so a reference is needed in the decision-making process. One way of making decisions is by using a decision tree in the form of a decision tree. The decision tree is created using several attributes used as input variables with one output variable as the final result. The decision tree is part of the data mining process carried out using classification algorithms.

KDD (Knowledge Discovery in Database) is one of the methods used to acquire knowledge from a provided database [1]. KDD serves as a technique to form patterns or rules within information [2]. The resulting information is obtained from large data and stored in a database that was initially unknown, resulting in useful data [2]. The knowledge obtained will be utilized as a knowledge base for decision making [1]. The concept used to generate rules in knowledge discovery is known as data mining [2]. Data mining or data extraction is a process that involves artificial intelligence, statistical methods, mathematics, artificial intelligence, and machine learning to identify and extract useful information and knowledge related to large databases [1] [2]. Data mining is also a discipline aimed at unearthing, discovering, and adding knowledge from owned data or information [3]. The dataset is a collection of data to be used as a sample in data training and data testing.

Classification is a data mining technique used to map data into predetermined groups or classes with significant development [4]. Classification is the most common method of data mining used to discover data from large databases [5]. One of the classification methods that is easily interpreted by humans is the decision tree [5] [**?**]6. The concept of the decision tree algorithm is to convert data into a decision tree and decision rules [6]. Some of the decision tree algorithm models include IDS, ID3, C4.5, CHAID, and CART [5] [6]. Data mining is a series of processes to extract added value from a collection of data in the form of knowledge that has not been known manually [5] [5]. Data mining analyzes a large amount of observational data sets and discovers unexpected relationships and summarizes data so that it is easily understood by users [1] [8]. Data mining is a field that consists of various disciplines that integrate machine learning, pattern recognition, statistics, databases, and visualization [7] [9]. This can be seen from the numerous studies that use classification algorithms in data mining in various fields [4].

The ID3 algorithm is an algorithm used to generate decision trees using the concept of information entropy [3] [6]. Attribute selection in the ID3 algorithm is done using Information Gain, where the attribute with the highest gain is chosen. On the other hand, entropy is used as a requirement in the class, where a low entropy value is good for extracting classes [3]. The C4.5 algorithm is a development of the ID3 algorithm, with attribute selection using gain ratio [6]. The C4.5 algorithm is widely used in research for performing classification functions, where the output in the algorithm is a decision tree [3].

Cross Validation is a model validation technique used to test and evaluate the effectiveness of machine learning models with validation techniques by dividing all data into training sets and test sets [4]. K-fold cross validation model is one of the cross validation techniques used to eliminate bias in the data used to test the accuracy level of the classification algorithm models used [4] [10].

The purpose of this research is to analyze the accuracy comparison between the two classification algorithms using RapidMiner Studio. The classification algorithms to be compared are ID3 and C4.5. Accuracy is performed using cross validation technique for multiple tests, resulting in the highest accuracy value as the solution. The attributes to be analyzed are written test, practical test, presentation, skill, and interview. Both written test, practical test, presentation, skill, and interview are evaluated as good, fair, and sufficient. This research will produce a solution in the form of the highest accuracy value, precision value, and recall value, which will be used as a reference in decision tree creation. The decision tree will serve as a basis to help generate decisions.

Research by Dahri and Fujiati [4], related to the analysis of the ID3 and C4.5 algorithms for ATM money filling data. In the study conducted, it was found that both algorithms used showed that the cross-validation of the C4.5 algorithm performed the best with an accuracy rate of 96.17%. This indicates that the C4.5 algorithm is an effective and efficient classification algorithm. The C4.5 algorithm can create the best decision tree with rules that produce predictions for ATM filling in the future.

Research conducted by Dana and Wijanarto [11] regarding the comparison analysis between ID3 and C4.5 for the classification of grant recipients for PDAM drinking water installation to improve drinking water services. The study conducted using RapidMiner showed that the ID3 algorithm had an accuracy rate of 98.91% and C4.5 had a higher accuracy rate of 99.14%. Therefore, the C4.5 algorithm has a higher accuracy level compared to the ID3 algorithm.

Research by Maingi, Lukandu, and Mwau [12] compares the use of decision tree algorithms C4.5 and ID3 for the analysis of disease symptoms and diagnosis. The C4.5 and ID3 algorithms can provide methods for solving existing data and obtaining entropy and information from the variables determined in disease outbreak data. The decision tree was successfully built and used to determine disease classification

Research [5] by Hssina, Merbouha, Ezzikouri, and Erritali compares the decision tree ID3 and C4.5, stating that decision trees can solve problems presented in data processing. The research compares ID3 and C4.5, C4.5 and C5.0, and C5.0 and CART. The test results show that C4.5 is the most preferred and used method in machine learning.

Ramesh, Swathi, Babu, and Padmavathamma [13] in their research on the comparative analysis of ID3 and C4.5 algorithms in B2B marketing. The study was conducted to assess the effectiveness of distributors in B2B marketing related to the mismatch between manufacturers and distributors. Where the factory will play a role in selecting suitable distributors. ID3 and C4.5 algorithms analysis is used for classification on various parameters related to the pattern of process for safe decision making. Decision trees are built to select secure distributors from B2B marketing data sets. The results of the study decide that C4.5 is capable of identifying users and is more satisfying compared to ID3.

The research conducted by Sathyadevan and Nair [14] on the comparison of ID3, C4.5, and Random Forest decision tree algorithms. The research explains that the three algorithms have differences in their accuracy. A comparison of algorithms is done using a data set and comparing the results. Apparently, Random Forest provides better prediction results and is an accurate algorithm for classification problems.

Research [15] by Sudrajat, Irianingsih, and Krisnawan relates to the analysis of data mining classification using a comparison of the C4.5 and ID3 algorithms. The analysis results show that when both algorithms are tested with the same dataset, they produce different models and accuracies. For discrete data, the ID3 algorithm has a higher accuracy of 99.83% compared to C4.5. On the other hand, for numeric data, the C4.5 algorithm has a higher accuracy of 89.69% compared to ID3. The results indicate that both algorithms can achieve good accuracy with minimal errors.

2. Methods

The data mining classification algorithm that will be used in the comparison in this study is the ID3 algorithm with the C4.5 algorithm. The ID3 and C4.5 algorithms are used to observe the accuracy, precision, and recall values generated between the two algorithms. These values are observed using cross-validation models. Based on the results obtained, the highest accuracy, precision, and recall values will be used as the basis and reference for decision making. Furthermore, the best algorithm with the highest accuracy value will be used to create a decision tree, which will show the attributes that will influence the decision support for the acceptance or rejection of assistant candidates. The research flow in the comparison of data mining classification algorithms is shown in Figure 1.

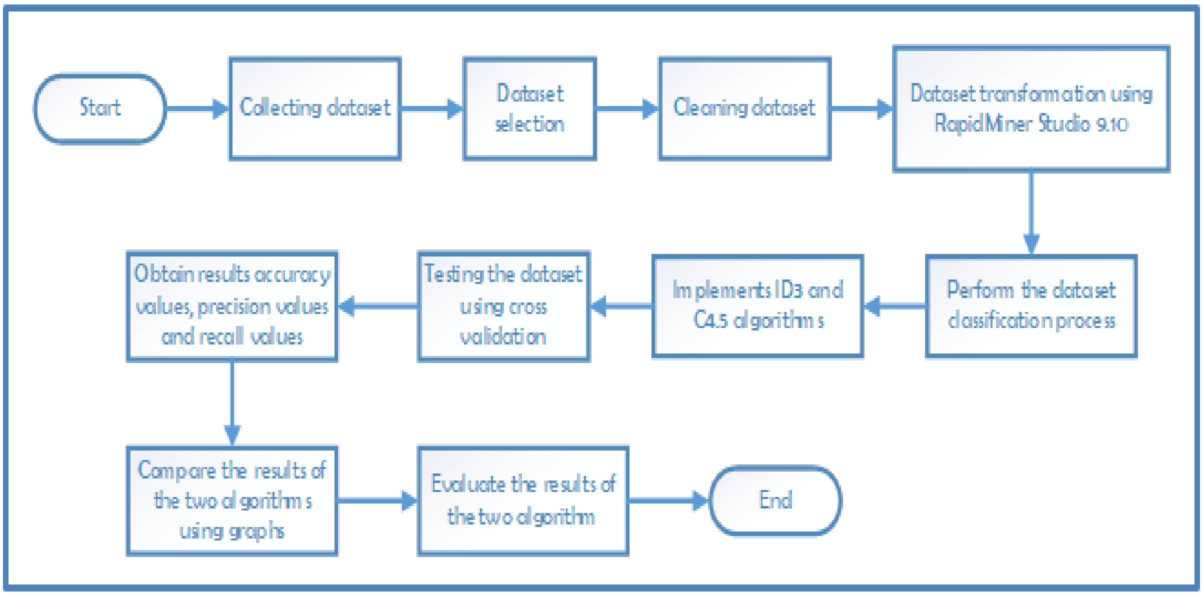


Figure 1. Research Flow

The pseudocode of the ID3 algorithm used in the study to obtain the attributes that influence the acceptance of assistant candidates is [3] [4]:

a) Start

b) Select attribute as the root by calculating its entropy value. The formula used to calculate the entropy value can be seen in equation 1

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

c) Creating branches for each attribute that has the highest gain value. The formula used to calculate the gain ratio value using equation 2.

$$Gain(S, A) = Entropy(S) - \sum_{v} Enilai(A) \left|\frac{a}{b}\right| Entropy(S_v) \tag{2}$$

d) Repeat steps b and c for each branch until all cases have a decision.

e) Creating rules based on the decision tree.

f) Finished.

Pseudocode Algorithm C4.5 used in this research to obtain the attributes that influence in determining the candidate assistant to be accepted are [4]:

a) Start.

b) Select the attribute to be used as the root.

c) Calculate the entropy and gain value. The formula to calculate the entropy value is using equation 3.

$$Entropy(S) \sum_{i=1}^{n} -pi * log_2 pi \tag{3}$$

d) The result of the entropy value is then used to calculate the gain value using equation 4

$$Gain(S, A) = entropy(S_i) - \sum_{i=1}^{n} \frac{|s_i|}{|s|} * Entropy(S_i) \tag{4}$$

e) Calculate the split info value using equation 5.

$$SplitInformation(S, A) = -\sum_{i=1}^{n} \frac{|s_t|}{|s|} + Log_2 \frac{|s_t|}{|s|} \tag{5}$$

f) The result of the gain value and split info is used to find the value of the gain ratio using equation 6.

$$GainRatio(S, A) = \frac{GainInformation(S, A)}{SplitInformation(S, A)} \tag{6}$$

g) The highest gain ratio is taken as the root.

h) Repeat steps b and c for each branch until all cases have a decision.

i) Create rules based on the decision tree.

j) Finished.

Steps in data mining classification using ID3 and C4.5 algorithms are:

a. Data Collection
   Process Data collection process is done by taking samples from the Laboratory of Basic Physics at Gunadarma University. The sample data used consists of 60 samples taken during the assistant intake in the odd semester of 2023/2024.

b. Data Processing Process
   Data processing in the data mining stage includes data selection, data cleaning, and data transformation. Data selection process involves selecting relevant and irrelevant data, so that only the necessary data for analysis is used. Data cleaning process involves removing irrelevant, inconsistent, and duplicate data. Data transformation involves changing the data according to the needs and inputting the dataset into the necessary tools.

c. Data Classification Process
   Data classification process is part of data mining, which involves comparing classification algorithms such as ID3 and C4.5.

d. Data Testing Process Data
   The Testing process is the final step in the classification process, which involves validating the classification process performed in the previous stages. Validation process is done using cross-validation method.

3. Result and Discussion

Implementation of data mining classification algorithms for the analysis of ID3 and C4.5 algorithms processed using RapidMiner Studio 9.10 tools. This will result in a decision tree that contains attributes that influence the selection of assistant candidates. The data will be tested and analyzed using cross-validation for both algorithms. The data used in the study is taken from 60 samples of assistant candidate acceptance data. The steps performed from data collection to data testing are as follows:

a. Data Collagtion Phase

This is the phase where the process of collecting the dataset to be used as sample data takes place. The dataset used in the study is taken from the acceptance data of new assistants in the Basic Physics Laboratory for the odd semester of 2023/2024, which took place over three weeks. The data collected and used as samples are 60 data points. The attributes of the assistant candidates include name, class, major, phone number, gender, address, written test score, practical test score, equipment presentation ability, equipment usage ability, and interview result.

b. Data Processing

After the data collection stage, the data processing stage is carried out. The data processing stage in data mining consists of three parts: data selection, data cleaning, and data transformation. Data selection involves the process of selecting data from 60 datasets that will be used. The selection process is done for the attributes needed in the formation of decision trees and those that are not used. After the data selection process is done, the data cleaning process is carried out by removing or eliminating irrelevant and unnecessary attributes in the next stages. As a result, the number of attributes is reduced after the data cleaning process, resulting in a decision tree as required. The attributes generated after the data cleaning process are five attributes: written test score, practical test score, presentation tool ability, tool usage ability, and interview result.

The final stage in data processing is the data transformation process. The data transformation process involves grouping the data into simpler parts. It will be divided into two variables, namely input variables and output variables. Input variables are the attributes used after the data cleaning process. Output variables are the conclusions or outcomes of what is being targeted. The input variables in the study are:

1) Attribute "Written Test" containing values "Good", "Sufficient", and "Poor".
2) Attribute "Practical Test" containing values "Good", "Sufficient", and "Poor".
3) Attribute "Presentation" containing values "Good", "Sufficient", and "Poor".
4) Attribute "Skills" containing values "Good", "Sufficient", and "Poor".
5) Attribute "Interview" containing values "Good", "Sufficient", and "Poor".

Variable output is the variable result in research that becomes the goal in the study. The target in the study is the attribute "Outcome" which contains "Pass" and "Fail". Table 1 is a dataset table for the process of data transformation that has gone through the data cleaning stage.

Table 1. Dataset

Written Test	Practical Test	Presentassion	Skills	Interview	Outcome
Poor	Poor	Sufficient	Poor	Poor	Fail
Good	Sufficient	Sufficient	Sufficient	Good	Pass
Sufficient	Good	Good	Sufficient	Good	Pass
Sufficient	Poor	Good	Good	Good	Pass
Poor	Poor	Poor	Sufficient	Sufficient	Fail
Good	Good	Good	Sufficient	Good	Pass
Good	Poor	Poor	Sufficient	Sufficient	Fail
Poor	Poor	Poor	Poor	Sufficient	Fail
Poor	Poor	Poor	Poor	Poor	Fail
Sufficient	Sufficient	Good	Good	Good	Pass
Poor	Sufficient	Poor	Poor	Sufficient	Fail
Sufficient	Sufficient	Good	Poor	Poor	Pass
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
Sufficient	Sufficient	Poor	Poor	Poor	Fail
Good	Sufficient	Good	Sufficient	Good	Pass
Poor	Poor	Poor	Poor	Poor	Fail
Sufficient	Good	Sufficient	Good	Sufficient	Pass
Sufficient	Sufficient	Good	Good	Good	Lulus

Figure 2 shows the process of importing the dataset from Table 1 into RapidMiner Studio 9.10. Figure 2 will undergo preprocessing to check for any problematic data. Figure 3 shows the result of the dataset from Image 2. Variables that have more than two values are created in polynomial data type, while variables that only have two values are created in binomial form. For the output variable "Hasil", it is also created in binomial data type, and in "Label" form as well, as it serves as the final goal or output in the decision-making process.

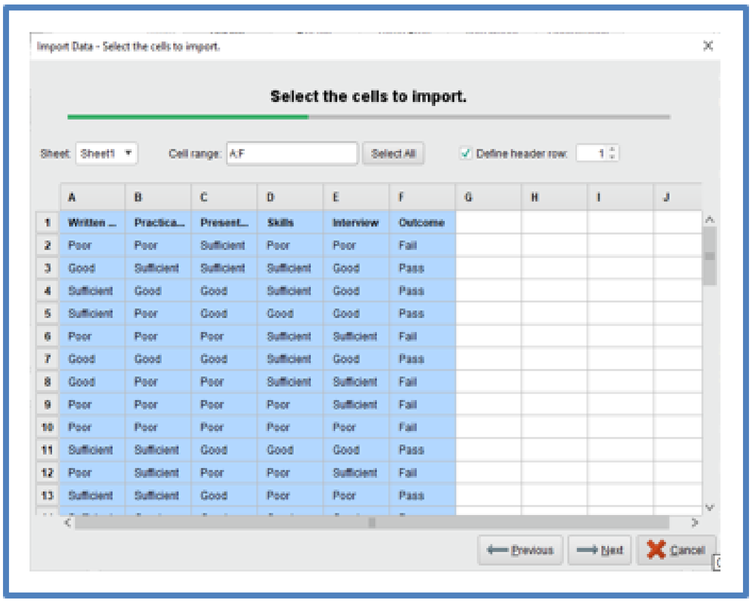
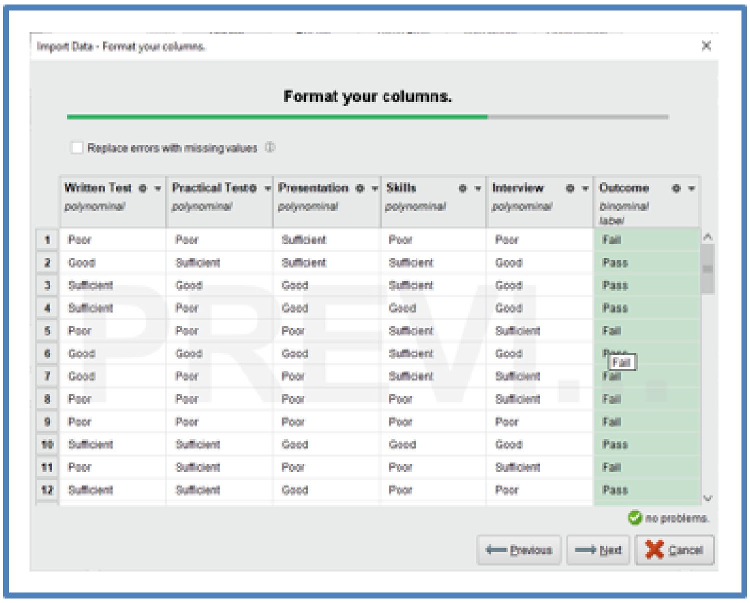


Figure 2. Import Dataset



	Written Test polynominal	Practical Test polynominal	Presentation polynominal	Skills polynominal	Interview polynominal	Outcome binominal label
1	Poor	Poor	Sufficient	Poor	Poor	Fail
2	Good	Sufficient	Sufficient	Sufficient	Good	Pass
3	Sufficient	Good	Good	Sufficient	Good	Pass
4	Sufficient	Poor	Good	Good	Good	Pass
5	Poor	Poor	Poor	Sufficient	Sufficient	Fail
6	Good	Good	Good	Sufficient	Good	[illegible]
7	Good	Poor	Poor	Sufficient	Sufficient	Fail
8	Poor	Poor	Poor	Poor	Sufficient	Fail
9	Poor	Poor	Poor	Poor	Poor	Fail
10	Sufficient	Sufficient	Good	Good	Good	Pass
11	Poor	Sufficient	Poor	Poor	Sufficient	Fail
12	Sufficient	Sufficient	Good	Poor	Poor	Pass

Figure 3. Dataset Transformation Process

c. Data Clasification Phase

After the import transformation process of the dataset, as shown in Figure 3, the next step is to perform data classification by comparing the two classification algorithms used, namely ID3 and C4.5. Figure 4 represents the cross-validation model for both algorithms, where the cross-validation model for the ID3 algorithm is shown in the image with cross-validation and dataset (top), while for the C4.5 algorithm, it is shown with cross-validation (2) and dataset (2) (bottom).

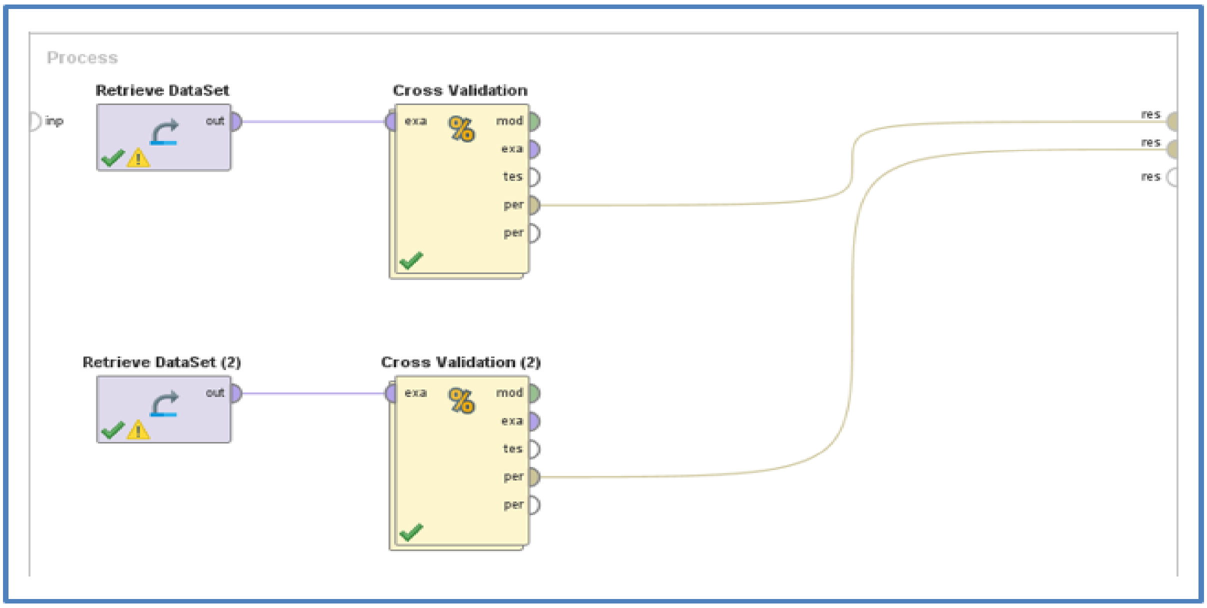


Figure 4. Cross Validation Model of ID3 Algorithm and C4.5 Algorithm

The cross validation testing method for the ID3 algorithm is seen in Figure 5, and the cross validation testing for the C4.5 algorithm is seen in Figure 6. In the ID3 algorithm, the parameter used in the criteria is information gain, while in the C4.5 algorithm, the parameter used in the criteria is gain ratio. Both the ID3 and C4.5 algorithms will be evaluated based on their accuracy, precision, and recall values, in order to determine which algorithm has the highest value and will be used as a reference in decision tree creation.

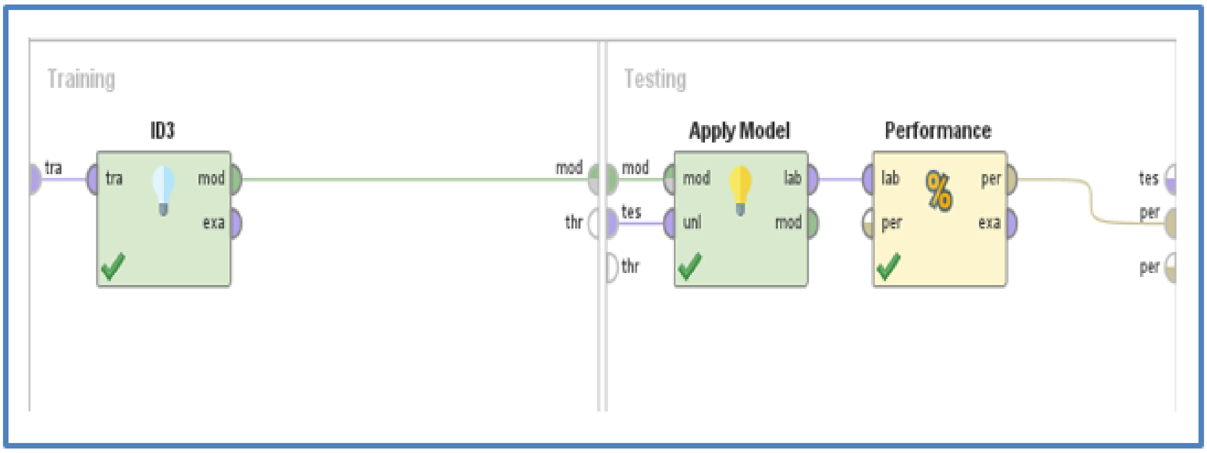


Figure 5. Cross Validation Testing for ID3 Algorithm

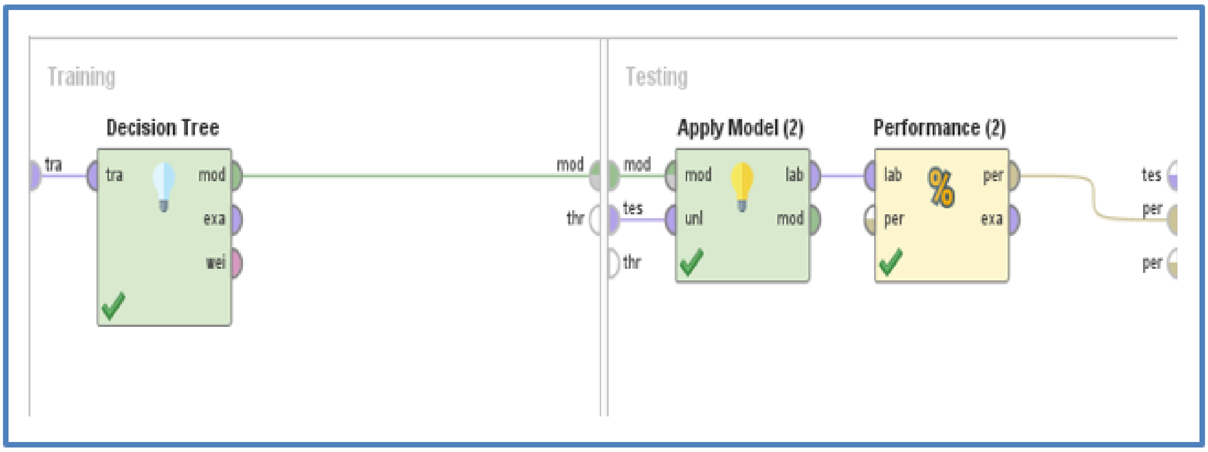


Figure 6. C4.5 Algorithm Cross Validation Testing

d. Data Testing Phase

The testing phase becomes the final stage in the data classification process, where validation is performed for each previously completed stage. Data validation is done using the cross-validation method. The cross-validation method used in the testing phase is the K-Fold Cross Validation model. This model is used to assess the accuracy level of both classification algorithm models used. K-Fold Cross Validation is one of the cross-validation techniques used to eliminate bias in the dataset [4]. Testing and training will be conducted k times, in this case, the k-fold cross-validation values used for both algorithms are 2, 4, 6, 8, and 10 folds. The validation results for the testing data performed on both algorithms against the k-fold values can be seen in Table 2, Table 3 and Table 4. Meanwhile, the comparison between the two algorithms can be seen in the form of graphs in Figure 7, Figure 8 and Figure 9.

Table 2 shows the validation results for comparing the precision values against the testing data using both classification algorithms. The comparison of precision values from both algorithms can be seen in the graph in Figure 7.

Table 2. Precision Value Comparison of Algorithms with Cross Validation

Classification Algorithm	Number of Fold	Precision
ID3	2	94,44%
	4	94,44%
	6	93,94%
	8	94,44%
	10	94,44%
C4.5	2	100%
	4	96,88%
	6	94,29%
	8	93,94%
	10	96,97%

The results of testing using cross validation on the ID3 algorithm showed the highest precision value of 94.44%. This highest value of 94.44% was obtained for all k-fold values, except for k-fold = 6. Meanwhile, the C4.5 algorithm's cross validation value showed the highest value of 100% at k-fold = 2. From the comparison of the precision values obtained, it is evident that the C4.5 algorithm has a higher precision value compared to the ID3 algorithm, as depicted in Figure 7.

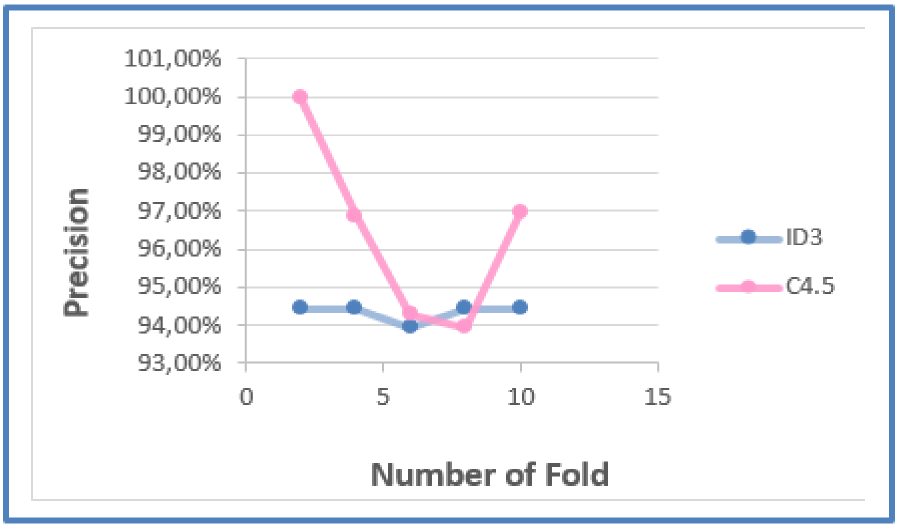


Figure 7. Precision Graph

Table 3 is a table comparing the validation test results for recall values between the ID3 and C4.5 algorithms. The comparison of recall values between the two algorithms can be seen in the form of a graph in Figure 8.

Table 3. Recall Value Comparison of Algorithms with Cross Validation

Classification Algorithm	Number of Fold	Recall
ID3	2	97,14%
	4	97,14%
	6	88,57%
	8	97,14%
	10	97,14%
C4.5	2	100%
	4	88,57%
	6	94,29%
	8	88,57%
	10	91,43%

The test results conducted on Table 3 show that the highest recall value for the ID3 algorithm is 97.14%, which applies to all k-fold values except for k-fold = 6. Meanwhile, the highest recall value for the C4.5 algorithm is 100% at k-fold = 2. The comparison of the obtained recall values indicates that the C4.5 algorithm has a higher value compared to the ID3 algorithm, which is 100% at k-fold = 2 as depicted in Figure 8.

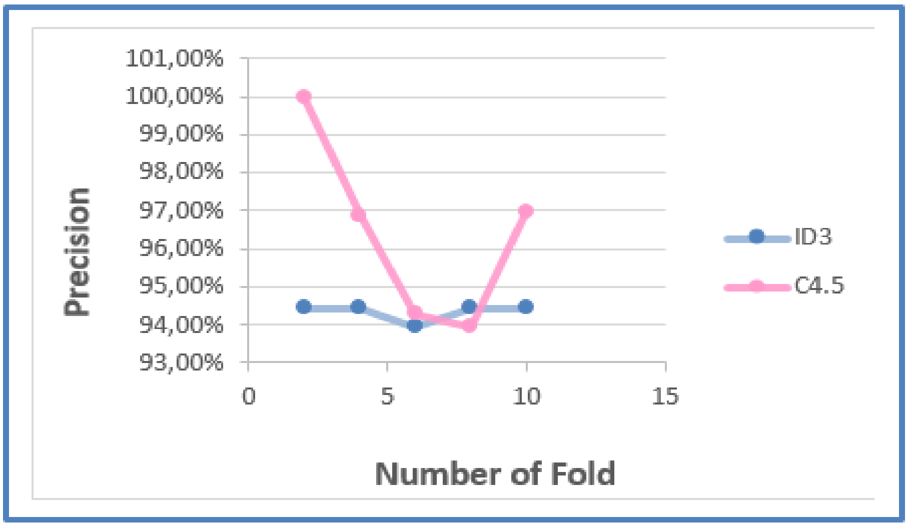


Figure 8. Recall Chart

The comparison of accuracy values generated between the two algorithms using cross validation can be seen in Table 4. For a visual comparison of both in graph form, refer to Figure 9.

Table 4. Accuracy Values Comparison of Algorithms with Cross Validation

Classification Algorithm	Number of Fold	Accuracy
ID3	2	95%
	4	95%
	6	90%
	8	94,87%
	10	95%
C4.5	2	96,67%
	4	91,67%
	6	93,33%
	8	89,96%
	10	93,33%

The comparison of accuracy values in Table 4 shows that the highest accuracy value for the ID3 algorithm is at k-fold = 2, 4, and 10, which is 95%. Meanwhile, for the C4.5 algorithm, the highest accuracy value is at k-fold = 2, which is 96.67%. From the generated accuracy values, it can be seen that the C4.5 algorithm is higher than the ID3 algorithm with a k-fold value of 2, which is 96.67%, as shown in the graph in Figure 9.

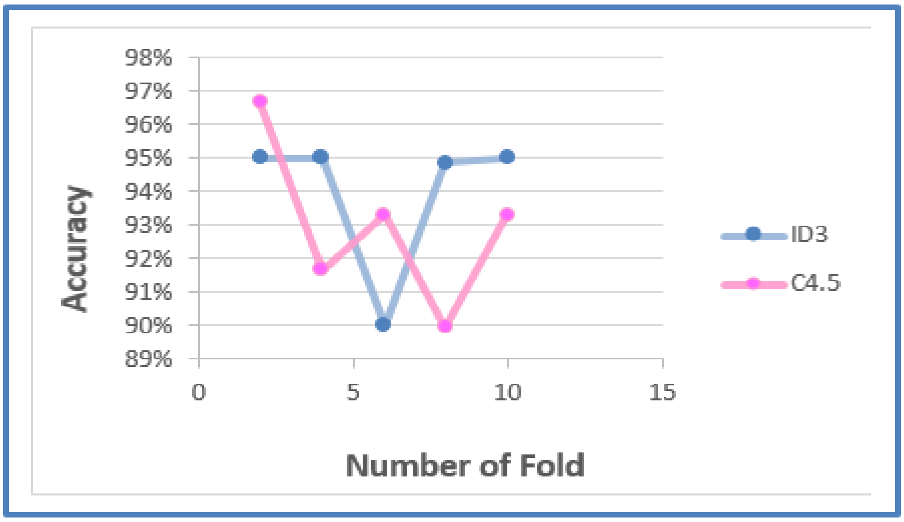


Figure 9. Accuracy Graph

Therefore, it can be concluded from the cross-validation testing conducted between the two classification algorithms that the C4.5 algorithm has a higher accuracy value compared to the ID3 algorithm. It can be seen that for all values ranging from accuracy, precision, and recall, the C4.5 algorithm always has higher values than the ID3 algorithm. Thus, from these validation values, it can be stated that the decision tree formation is carried out using the C4.5 algorithm as a reference in decision making. The decision tree generated from the C4.5 algorithm can be used as a basis in the process of determining candidates for assistants in the Basic Physics Laboratory. The decision tree from the C4.5 algorithm is created using RapidMiner Studio 9.10 tools and can be seen in Figure 10.

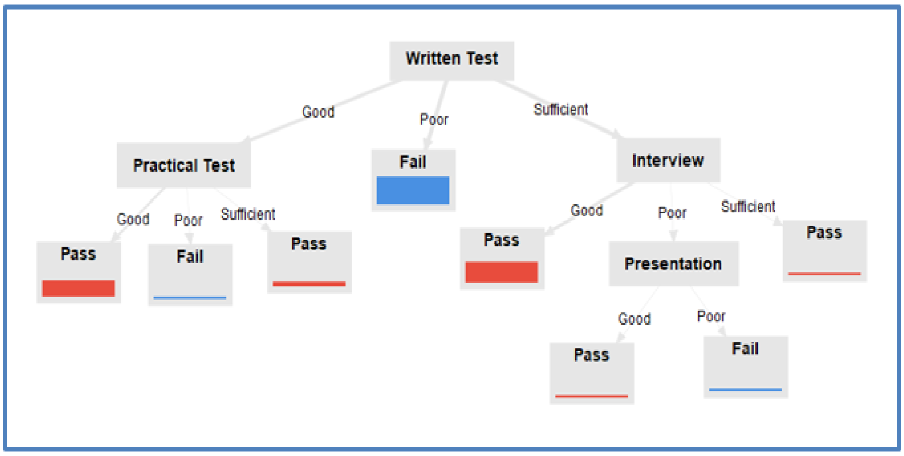


Figure 10. C4.5 Algorithm Decision Tree

Figure 11 shows the rules generated from the decision tree in Image 10 using five attributes as input variables. From the decision tree generated in Figure 11, it can be seen that out of the 60 datasets, the candidates declared "Passed" have five rules as results, while the candidates declared "Failed" have three rules as results, all of which refer to the attribute "written test."

Tree

```
Written Test = Good
|   Practical Test = Good: Pass {Fail=0, Pass=13}
|   Practical Test = Poor: Fail {Fail=1, Pass=0}
|   Practical Test = Sufficient: Pass {Fail=0, Pass=3}
Written Test = Poor: Fail {Fail=23, Pass=0}
Written Test = Sufficient
|   Interview = Good: Pass {Fail=0, Pass=17}
|   Interview = Poor
|   |   Presentation = Good: Pass {Fail=0, Pass=1}
|   |   Presentation = Poor: Fail {Fail=1, Pass=0}
|   Interview = Sufficient: Pass {Fail=0, Pass=1}
```

Figure 11. C4.5 Decision Tree Rule

4. Conclusion

Comparison analysis of accuracy between two classification algorithms, namely ID3 and C4.5 algorithms, resulted in the decision that the C4.5 algorithm has a higher accuracy compared to the ID3 algorithm. The comparison of test results using cross validation on the C4.5 algorithm is not only higher in accuracy but also in precision and recall values. These testing results were obtained using the K-Fold Cross Validation model with variation values of 2, 4, 6, 8, and 10 folds. For accuracy value, the C4.5 algorithm has a value of 96.67% at k-fold = 2. Therefore, the C4.5 algorithm can be used as a reference in making decisions for determining the acceptance of candidates for the Basic Physics Laboratory assistant position using rules on attributes such as written tests, practical tests, interviews, and presentations. Further development related to this research can utilize other data mining classification algorithms to make decisions that are appropriate for the attributes used. Another development related to this research includes determining decision-supporting attributes that have a greater impact on the accuracy gain ratio.

References

[1] Gustientiedina, M. Adiya, Desnelita and Yenny, "Penerapan Algoritma K-Means Untuk Clustering Daya Obat-Obatan Pada RSUD Pekanbaru," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 05, no. 01, pp. 17-24, 2019.

[2] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," Jurnal SIMETRIS, vol. 11, no. 2, pp. 1-13, 2020.

[3] Ardiyansyah, P. A. Rahayuningsih and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Maning Untuk Dataset Blogger Dengan Rapid Miner," Jurnal Khatulistiwa Informatika, vol. VI, no. 1, pp. 20-28, 2018.

[4] D. Y. H. Tanjung and Fujiati, "Comparison Analysis of ID3 and C4.5 Algorithm on ATM Money Charging Data," CSRID Journal, vol. 13, no. 3A, pp. 231-242, 2021.

[5] B. Hssina, A. Merbouha, H. Ezzikouri and M. Erritali, "A comparative study of decision tree ID3 and C4.5," (IJACSA) International Journal of Advanced Computer Science and Applications, pp. 13-19.

[6] S. Defiyanti and D. L. C. Pardede, "Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi Spam-Mail," Gunadarma University Repository

[7] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," Innovation in Research of Informatics (INNOVATICS), vol. 1, no. 2, pp. 63-69, 2019.

[8] W. Supriyatin and Y. Rianto, "Application of Naive Bayes Algoritm to Analysis of Free Fatty Acid (FFA) Production Based on Fruit Freshness Level," KOMPUTASI: Jurnal Ilmiah Ilmu Komputer dan Matematika, vol. 20, no. 1, pp. 24-34, 2023.

[9] W. Supriyatin, "Palm oil extraction rate prediction based on the fruit ripeness levels using C4.5 algoritm," ILKOM Jurnal Ilmiah, vol. 13, no. 2, pp. 92-100, 2021.

[10] M. M. Fajar, A. R. Putri and K. F. H. Holle, "Perbandingan Metode Klasifikasi Data Mining Untuk Deteksi Keaslian Lowongan Pekerjaan di Medsos," Jurnal Ilmiah Informatika, vol. 7, no. 1, pp. 41-48, 2022.

[11] D. M. Agustina and Wijanarto, "Comparative Analysis of ID3 and C4.5 Algotihm for Classification of Grant Reciipients of Drinking Water Installation at PDAM Kendal," Journal of Applied Intelligent System, vol. 1, no. 3, pp. 234-244, 2016.

[12] N. N. Maingi, I. A. Lukandu and M. Mwau, "Comparative Analysis of the C4.5 and ID3 Decision Tree Algorithms for Disease Symptom Classification and Diagnosis," International Research Journal of Advanced Engineering and Science, vol. 4, no. 2, pp. 188-194, 2019.

[13] C. P. Ramesh, R. Swathi, G. V. R. Babu and M. Padmavathamma, "Comparative Analysis of ID3 and C4.5 Decision Making Algorithm in B2B Marketing," International Journal of Engineering Research & Technology (IJERT), pp. 208-212, 2014.

[14] S. Sathyadevan and R. R. Nair, "Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest," Computational Intelligence in Data Mining, vol. 1, pp. 549-562, 2015.

[15] R. Sudrajat, I. Irianingsih and D. Krisnawan, "Analysis of data mining classification by comparison of C4.5 and ID algorithms," IOP Conference Series: Materials Science and Engineering, pp. 1-8, 2017.