Journal of Lorem
HTTPS://EJOURNALEJOURNAL.MEDIAKUNKUN.COM/INDEX.PHP/LOREM
LOREM : Computational Engineering and Computer Information Systems
E-ISSN : 3032-3924, P-ISSN : 3032-3894

Classification of Student Readiness for Educational Unit Exams: Decision Tree Approach C4.5 Based on Try Out Scores at MTs Nahdlatul Arifin

Mas'ud Hermansyah[1], Nur Andita Prasetyo[2], Agung Muliawan[3], M. Faiz Firdausi[4], Ferry Wiranto[5]

Sistem dan Teknologi Informasi, Institut Teknologi dan Sains Mandala[1,2]
Rekayasa Perangkat Lunak, Institut Teknologi dan Sains Mandala[3,4,5]

ABSTRACT

The National Examination, which has now changed to an Educational Unit Examination which is held every year by the Government, has become the school's concentration in preparation for it. In this case, each school has its own way of preparing students to be ready to face the Education Unit Examination. Classifying students in readiness for the Educational Unit Examination is one way for the school to do so. MTs Nahdlatul Arifin is one of the schools that implements this. Classifying students based on the scores from the Try Out results held by the school is a method carried out by MTs Nahdlatul Arifin. Along with advances in technology, classifying student grades can be done using Data Mining with several algorithms. However, in this research a comparison of several algorithms has been carried out. Compared with the Weka tools, the C4.5 algorithm was finally chosen for this research. The number of classifications carried out with correct results on a total of 100 student data is 20 + 33 + 39 = 92 and the number of classifications carried out with incorrect results is 2 + 2 + 4 = 8. So the accuracy of this model is (92/100) = 0.92 or 92%.

Keywords: *Data Mining, Classification, Try Out, C4.5 Algorithm*

Corresponding Author:
Mas'ud Hermansyah
- masudhermansyah@gmail.com

Received: September 11, 2023
Revised: October 11, 2023
Accepted: January 02, 2024
Published: January 11, 2024

1. INTRODUCTION

The education sector in Indonesia faces various challenges in efforts to improve the quality and quality of learning. One of the most crucial moments in education is educational unit examinations, such as the National Examination (UN) at the secondary level, which functions as an assessment of student competency achievement. The results of these exams not only have an impact on student achievement, but also reflect the effectiveness of educational institutions in equipping students with the knowledge and skills necessary for success in the future (Rahmatullah, 2016).

Currently, research on data mining and education systems is increasingly being carried out. Research on data mining in the world of education has been around for a long time (since the 1990s) and was only grouped into an Educational Data Mining research field in 2005 when a group of data mining researchers created a data mining research organization in the world of education starting in 2008, this organization held EDM annual conference which discusses data mining research in education throughout the world (Rahmayuni, 2014).

In the world of education, abundant and continuous data regarding the students being coached continues to be produced. Abundant data opens up opportunities for applying data mining for better educational management and data mining for implementing more effective

computer-assisted learning. By applying classification techniques in data mining, strategic information can be extracted. Information from the analysis can be used to find new opportunities and find strategic plans in the process of classifying students (Darmawan et al., 2022).

Learning that is given in the right way will provide optimal results. Based on this idea, MTs Nahdlatul Arifin is present in the midst of the increasingly developing world of education to create pleasant learning conditions by optimizing children's abilities and directing them. MTs Nahdlatul Arifin is a private school that is integrated with an Islamic boarding school in Sumberejo Village, Ambulu District, Jember Regency by implementing active learning, fun learning, and IQRO (Inquiry, Question, Repeat, & Action) so that children become efficient and efficient learners. noble character. In preparing students for the Education Unit Exam they will face, MTs Nahdlatul Arifin held Try Outs for grade 3 students 5 times. To be able to see students' abilities in preparation for the Education Unit Examination. MTs Nahdlatul Arifin plans to classify students according to their abilities based on the Try Out results.

Based on the data from the try out results, an object will be classified into a set of categories based on the object in question, such as researchers created a website-based New Student Admission System (Muliawan et al., 2023). The algorithm that can be used for classification is the C4.5 algorithm. After going through classification and data processing, patterns will be obtained where the resulting output is a classification of students' readiness to face the Educational Unit Examination. The expected result of this research is to help MTs Nahdlatul Arifin determine what steps to take to help students face the Education Unit Examination.

This research also has the potential to provide broader insight into the use of data mining techniques in increasing the efficiency of education at the secondary level, which can become a reference for other educational institutions in Indonesia. Thus, it is hoped that this research can make a positive contribution to efforts to improve the quality of education in Indonesia, especially at the secondary level.

2. METHODS

2.1 Data Mining

Data mining is also called Knowledge Discovery in Database (KDD), namely extracting information from piles of data. The process of searching for information will find trend patterns in the data which then results from mining which can become knowledge and information that is easy to understand (Yulia & Azwanti, 2018). The results of the knowledge obtained can be used as a knowledge base that is used to make decisions. In more detail, the KDD process as in the following figure was adopted from (Adiya & Desnelita, 2019):

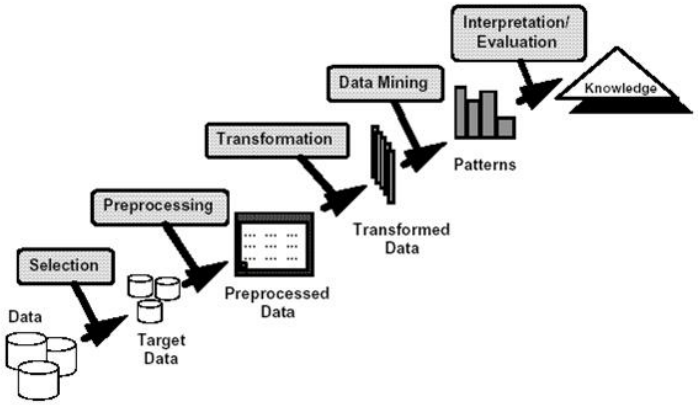


Figure 1. Knowledge Discovery Database (KDD) Process
Source: (Lumbantoruan & Kennedy, 2015)

Based on the image above, it is defined as one part of the Knowledge Discovery in Database (KDD) process which is tasked with extracting patterns or models from data using a specific algorithm. The KDD process is as follows (Masrizal & Munandar, 2019):

1. Selection
   Selection of data from a set of operational data needs to be done before the information mining stage in KDD begins. The selected data that will be used for the data mining process is stored in a file, separate from the operational database (Mahartika & Wibowo, 2019).
2. Pre-processing
   In preprocessing there are two stages, namely as follows :
   a. Data Cleaning
      Eliminate unnecessary data such as handling missing values, noisy data and handling inconsistent and relevant data.
   b. Data Integration
      Performed on attributes that identify unique entities (Adiya & Desnelita, 2019).
3. Transformation
   Transformation of the selected data, so that the data is suitable for the data mining process. The coding process in KDD is a creative process and really depends on the type or pattern of information to be searched in the database (Mahartika & Wibowo, 2019).
4. Data Mining
   Data mining is a technology that automates the process of finding interesting and sensitive patterns from large data sets. Data mining is useful for providing solutions to decision makers in business to improve the company's business (Siregar, 2018).
5. Interpretation / Evaluation
   At this stage, an interpretation/evaluation of the patterns resulting from the results of data classification is carried out using the algorithm used. At this stage also the use and overall feedback on the patterns and findings obtained by data mining (Budiman et al., 2015).
6. Knowledge

The resulting patterns will be presented to the user. At this stage, the new knowledge produced can be understood by everyone and will be used as a reference for decision making (Adiya & Desnelita, 2019).

2.2 Classification

Classification is a loan word from the Dutch language, classificatie, which itself comes from the French classification. This term refers to a method for compiling data systematically or according to several predetermined rules or conventions. Literally it can also be said that classification is the division of something according to classes. According to science, classification is the process of grouping objects based on similarities and differences.

Classification is one method of data mining. Classification is a predictive method that uses learning from existing data to produce a model that is used to predict new data. Classification is the process of finding a group of models that explain data classes, so that these models can be used to predict the value of an unknown class in an object. To get a model, we have to analyze the training data. Meanwhile, test data is used to determine the level of accuracy and model that has been produced. Classification can be used to predict the name or value of a data object (Astuti, 2018).

2.3 C4.5 Algorithm

The C4.5 Decision Tree Algorithm is an algorithm used to form decision trees. Decision trees are a very powerful and well-known classification and prediction method. Decision tree methods convert very large facts into decision trees that represent rules. Rules can be easily understood with natural language. And they can also be expressed in the form of a database language such as Structured Query Language to search for records in certain categories. Decision Trees are also useful for exploring data, finding hidden relationships between a number of candidate input variables and a target variable (Rusito & Firmansyah, 2016).

There are several stages in creating a decision tree in the C4.5 algorithm, namely:

1. Prepare customer data. Customer data is taken from previously existing records or called past data and has been grouped into certain classes.
2. Count the roots of the tree. The roots will be taken from the attributes that will be selected, by calculating the gain value for each attribute, the highest gain value will be the first root. Before calculating the gain value of the attribute, first calculate the entropy value.

Stages of Decision Tree Algorithm C4.5:

a. Prepare training data.
b. Determine the root of the tree.
c. Calculate the Entropy value:

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

where :

S : case set
n : number of partitions s
pi : proportion of Si to S

d. Calculates Gain information :

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{S}\ x\ Entropy\ (Si) \quad (2)$$

where :

S : case set
A : attribute
n : number of attribute A partitions

|Si| : number of cases in partition i
|S| : number of cases in S

e. Repeat the 3rd process until all branches have the same class. The branching process will stop if (Setio et al., 2020) :
   1) All records in node N get the same class.
   2) There are no attributes in the partitioned record anymore.
   3) There are no records in the empty branch (Pambudi et al., 2018).

f. The formation of decision trees in C4.5 has rules, namely (Iskandar et al., 2019) :
   1) Select the attribute that acts as the root by looking at the largest gain value.
   2) Create a branch with a rule value with a branching rule with a condition value on the root attribute.
   3) Split cases back into branches.
   4) Repeat cases until branches have the same class.

To deal with attributes with continuous or descriptive data, namely by sorting the values from the training data first (Julianto et al., 2014). Another technique, normalization, can also be done by looking for the average value and standard deviation in a continuous attribute range, as in equations 5 and 6.

2.4 Confusion Matrix

Measuring the performance of a classification system is important. Classification system performance describes how well the system is at classifying data. Wu's inside (Dewi, 2016) explained that Confusion Matrix is a method that is usually used to calculate accuracy in data mining concepts. This formula performs calculations with 4 outputs, namely: recall, precision, accuracy and error rate. Evaluation of classification models is based on tests to estimate true and false objects.

Based on the number of class outputs, classification systems can be divided into 4 (four) types, namely binary, multi-class, multi-label and hierarchical classification (Sokolova & Lapalme, 2009). In his research (Sokolova & Lapalme, 2009) also says that in binary classification, input data is grouped into one of two classes. This type of classification is the simplest and most widely used form of classification. Examples of its use include systems that detect people or not, vehicle detection systems or not, and movement detection systems or not.

Meanwhile, in the multi-class classification form, the input data is classified into several classes. For example, a system that can classify types of vehicles such as bicycles, motorbikes, cars, buses, trucks, and so on. The form of multi-label classification is basically the same as multi-class where data is grouped into several classes, but in multi-label classification, data can be included in several classes at once. The final form of classification is hierarchical. Input data is grouped into several classes, but these classes can be regrouped into simpler classes hierarchically. For example, in this research, movement directions are grouped into 12 movement directions which of course can be simplified into 4 directions. Example of confusion matrix from Olsen in (Indriani, 2014) shown in the table.

Tabel 1 Confusion Matrix Table

Classification	Predicted Class	
	Class = Yes	Class = No
Class = Yes	TP	FN
Class = No	FP	TN

Source: (Indriani, 2014)

In measuring performance using the confusion matrix, there are 4 (four) terms as representations of the results of the classification process. The four terms are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The True Negative (TN) value is the amount of negative data that is detected correctly, while the False Positive (FP) is negative data but is detected as positive data. Meanwhile, True Positive (TP) is positive data that is detected correctly. False Negative (FN) is the opposite of True Positive, so the data is positive, but is detected as negative data (Yang & Berdine, 2017).

Accuracy is the proportion of correct predictions. The accuracy calculation formula can be seen from equation (3) (Indriani, 2014):

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (3)$$

Precision is defined as the ratio of selected relevant items to all selected items. Precision can be interpreted as the match between a request for information and the answer to that request (Rosandy, 2016). In this case, precision defines the level of "noise" in the information displayed. The calculation formula can be seen in the equation (4).

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

Recall is the proportion of the number of controlled relevant text documents among all relevant text documents in the collection (Andika et al., 2019). The recall formula can be expressed as:

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

Confusion matrix only presents information in the form of numbers. For the case of 2 class classification (binary classification), if you want to display information on the performance of the classification algorithm in graphical form, you can use the Receiver Operating Characteristic (ROC) or Precision-Recall Curve.

3. RESULTS AND DISCUSSION

In this research, the data obtained from the research object will be analyzed and tested. Before being processed using data mining, the data undergoes a validation process to find and convert the data so that it can be used in data mining algorithms and obtain good accuracy and performance. Data validation used is by deleting incomplete or empty data that has no value (null). By using the Quota Sampling Sample Selection Method, the data is used as training data and testing data. The comparison used to determine training data and testing data was validated using k-fold cross-validation.

The data obtained from MTs Nahdlatul Arifin is data from the results of the 3rd grade junior high school students who tried out 5 times. The data that has been collected consists of data from the 2020-2021 and 2021-2022 student Try Out results which will be used as samples in the research which will be used as model validation, totaling 136 samples.

Figure 2 is a sample of the data used for this research. Because the data cannot be displayed as a whole, only a sample of 30 students with their Try Out results and classification is displayed.

NO	NIS	NAMA	MAT1	B. IND1	IPA1	B ING1	RATA21	KET1	...	MAT5	B. IND5	IPA5	B ING5	RATA25	KET5	KLASIFIKASI
1	1314.8.053	Adam Milando	35.00	72.00	45.00	74.00	56.50	Belum Lulus	...	35.00	76.00	25.00	76.00	53.00	Belum Lulus	Belum Siap
2	1213.7.003	Andi Vallian Superani	37.50	72.00	45.00	76.00	57.63	Belum Lulus	...	37.50	90.00	50.00	86.00	65.88	Belum Lulus	Belum Siap
3	1213.7.029	Aslam Akyas	52.50	66.00	62.50	66.00	61.75	Lulus	...	50.00	82.00	60.00	70.00	65.50	Lulus	Siap
4	1314.8.056	Bintang Hendrawan Rahmanto	30.00	50.00	35.00	48.00	40.75	Belum Lulus	...	30.00	66.00	50.00	64.00	52.50	Belum Lulus	Butuh Bantuan
5	1213.7.020	Dimas Zidane Wardana	42.50	64.00	60.00	56.00	55.63	Lulus	...	25.00	86.00	0.00	0.00	27.75	Belum Lulus	Belum Siap
6	1213.7.028	Fajar Bintang Sadewa	20.00	54.00	45.00	38.00	39.25	Belum Lulus	...	0.00	0.00	50.00	34.00	21.00	Belum Lulus	Butuh Bantuan
7	1213.7.076	Hanif Akmal Arsyi	57.50	68.00	60.00	70.00	63.88	Lulus	...	52.50	86.00	77.50	88.00	76.00	Lulus	Siap
8	1213.7.007	Hilmy Abdul Hadi	67.50	62.00	57.50	70.00	64.25	Lulus	...	37.50	36.00	57.50	34.00	41.25	Belum Lulus	Butuh Bantuan
9	1213.7.009	Ja' Far Tsabit Rabbani	55.00	74.00	0.00	0.00	32.25	Belum Lulus	...	52.50	86.00	72.50	82.00	73.25	Lulus	Siap
10	1213.7.027	Moch. Zida Fabial Yustisianif	42.50	60.00	55.00	52.00	52.38	Belum Lulus	...	47.50	82.00	67.50	68.00	66.25	Lulus	Siap
11	1213.7.011	Muhammad Bilal Sholih	42.50	66.00	60.00	52.00	55.13	Lulus	...	50.00	70.00	50.00	76.00	61.50	Lulus	Belum Siap
12	1213.7.022	Muhammad Ibrahim Kholil	45.00	72.00	0.00	0.00	29.25	Belum Lulus	...	40.00	80.00	67.50	82.00	67.38	Lulus	Siap
13	1213.7.014	Muhammad Nur Elfikri	0.00	0.00	60.00	76.00	34.00	Belum Lulus	...	45.00	82.00	67.50	80.00	68.63	Lulus	Siap
14	1314.8.058	Muhammad Pashya Islami	47.50	48.00	0.00	0.00	23.88	Belum Lulus	...	37.50	74.00	65.00	80.00	64.13	Belum Lulus	Butuh Bantuan
15	1213.7.016	Muhammad Tifal Nawwafin	42.50	74.00	50.00	72.00	59.63	Lulus	...	45.00	82.00	70.00	82.00	69.75	Lulus	Siap
16	1213.7.018	Ridho Muhammad Adzka	42.50	70.00	47.50	78.00	59.50	Lulus	...	50.00	78.00	67.50	68.00	65.88	Lulus	Siap
17	1213.7.019	Saifulhaq Asaduddin	32.50	72.00	45.00	68.00	54.38	Belum Lulus	...	30.00	74.00	72.50	72.00	62.13	Belum Lulus	Belum Siap
18	1213.7.025	Yusuf Romadhon Munif	35.00	70.00	0.00	0.00	26.25	Belum Lulus	...	45.00	86.00	67.50	78.00	69.13	Lulus	Belum Siap
19	1213.7.001	Adnan Yasir	37.50	70.00	42.50	72.00	55.50	Belum Lulus	...	40.00	86.00	60.00	76.00	65.50	Lulus	Belum Siap
20	1213.7.002	Ahmad Fadlur Rahman	32.50	72.00	42.50	42.00	47.25	Belum Lulus	...	42.50	78.00	52.50	56.00	57.25	Lulus	Butuh Bantuan
21	1213.7.004	Bima Alana Putra	20.00	54.00	52.50	66.00	48.13	Belum Lulus	...	20.00	70.00	65.00	84.00	59.75	Belum Lulus	Butuh Bantuan
22	1213.7.026	Dzikri Muhammad Isthafa	27.50	64.00	57.50	76.00	56.25	Belum Lulus	...	40.00	78.00	52.50	86.00	64.13	Lulus	Belum Siap
23	1213.7.005	Fatah Saiful Haq	35.00	70.00	45.00	72.00	55.50	Belum Lulus	...	50.00	90.00	72.50	86.00	74.63	Lulus	Belum Siap
24	1213.7.006	Fathurahman	65.00	74.00	40.00	54.00	58.25	Lulus	...	42.50	50.00	45.00	74.00	52.88	Belum Lulus	Butuh Bantuan
25	1213.7.008	Ja' Far Abdurrahman	27.50	64.00	45.00	50.00	46.63	Belum Lulus	...	30.00	64.00	62.50	60.00	54.13	Belum Lulus	Butuh Bantuan
26	1213.7.021	Mohammad Alfatih	22.50	70.00	57.50	72.00	55.50	Belum Lulus	...	30.00	62.00	0.00	0.00	23.00	Belum Lulus	Butuh Bantuan
27	1213.7.031	Mohammad Febrian Zulkarnain	25.00	64.00	50.00	56.00	48.75	Belum Lulus	...	35.00	58.00	0.00	0.00	23.25	Belum Lulus	Butuh Bantuan
28	1213.7.010	Muhammad	65.00	78.00	60.00	78.00	70.25	Lulus	...	70.00	84.00	67.50	74.00	73.88	Lulus	Siap
29	1213.7.012	Muhammad Farhan Rambe	15.00	74.00	62.50	74.00	56.38	Belum Lulus	...	55.00	82.00	42.50	88.00	66.88	Lulus	Belum Siap
30	1213.7.073	Muhammad Fauzan Fadhillah	45.00	76.00	42.50	54.00	54.38	Belum Lulus	...	45.00	78.00	62.50	52.00	59.38	Lulus	Siap

Figure 2. Sample Student Data and Try Out Values
Source: author's research data, 2023

The results that can be concluded from the Student Classification process based on the Try Out results are that the classification process can be carried out after students have undergone the Try Out 5 times, so that classification can be carried out according to the student's abilities from the Try Out results. In this way, MTs Nahdlatul Arifin can make decisions about preparing students for the National Examination. There are 3 classification categories carried out by MTs Nahdlatul Arifin, namely: Ready, Not Ready and Need Help. Where in each existing classification category the school will provide tutoring according to each student's abilities so that they are ready to face the Education Unit Examination.

Steps in calculations using the C4.5 algorithm on data from the try out results of 136 MTs Nahdlatul Arifin students. Divided into 100 training data and 36 test data. After determining the attribute as the root, the entropy calculation is then carried out. After obtaining the entropy value, then calculate the gain value for each attribute, then select the highest gain value. The selected attribute gain is the one with the highest gain value compared to the other gain values, which is then used as the root of the tree. Calculations to produce initial entropy values are shown in table 2:

$$Entropy(S) = \sum_{i=1}^{n} - pi.log_2 Pi \quad (6)$$

Table 2. Initial Entropy Calculation

Siswa	Siap	Belum Siap	Butuh Bantuan	Entropi
100	37	24	39	1.55466057896

Source: author's data processing, 2023

Entropy
= (-37/100*log2(37/100))+(-24/100*log2(24/100))+(-39/100*log2(39/100))
= 1.55466057896

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy\,(S_i) \quad (7)$$

The highest gain is in Try Out 2 with value 0.661989585
= 1.55466057896 - ((45/100)*0.846470282 + (55/100)*0.930471576)

Table 3. Four (4) calculations to determine Entropy and Gain

	Ket.	Jum.	Siap	Belum Siap	Butuh Bantuan	Entropi	Gain
Try Out 1	Passed	32	25	5	2	1.0229611	
	Not yet passed	68	12	19	37	1.3851658	0.2854002
Try Out 2	Passed	45	37	8	0	0.8464702	
	Not yet passed	55	0	16	39	0.9304715	0.6619895
Try Out 3	Passed	72	35	21	16	1.4023273	
	Not yet passed	28	2	3	23	0.8865374	0.2967543
Try Out 4	Passed	68	36	20	12	1.3807870	
	Not yet passed	32	1	4	27	0.8551146	0.3420886
Try Out 5	Passed	58	35	17	6	1.2445103	
	Not yet passed	42	2	7	33	0.95376329	0.4322639

Source: author's data processing, 2023

The data used to determine training data and testing data is validated using 10-Fold Cross-Validation. This data will be processed to produce a data mining model using WEKA tools as seen in Figure 3.

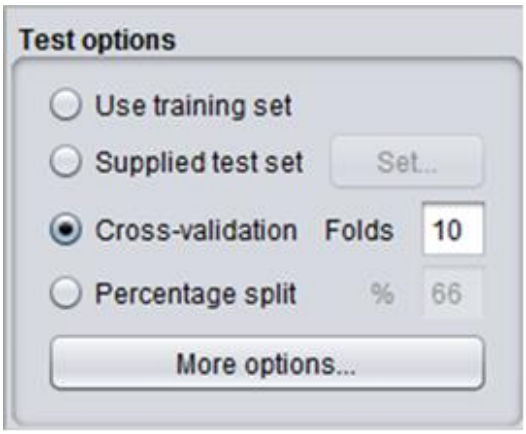


Figure 3. Ten (10)-Fold Cross-Validation with WEKA
Source: author's data processing, 2023

In Figure 4 is the calculation of the confusion matrix value for the C4.5 algorithm using 33 attributes and 136 data records of student Try Out results, as follows:

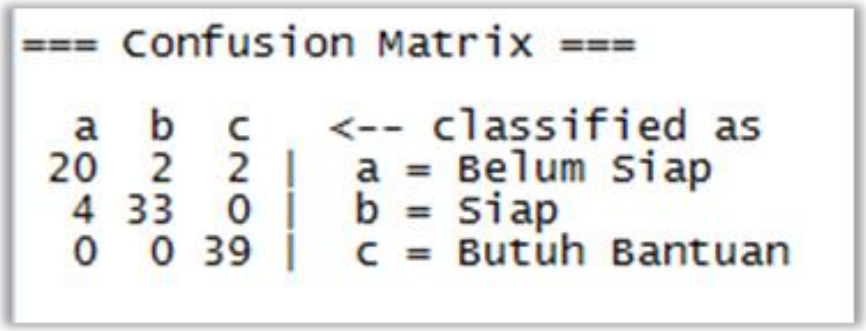
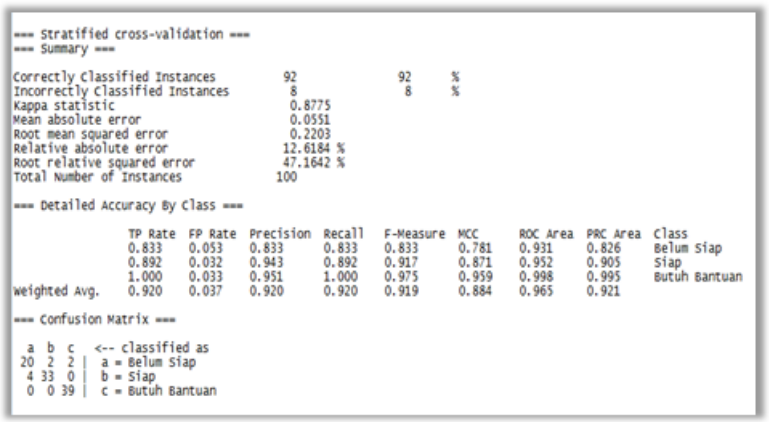


Figure 4. Confusion Matrix with WEKA
Source: author's data processing, 2023

From the confusion matrix in Figure 4.8, the accuracy value can be calculated as follows:

1. In the Not Ready criteria "20 2 2" shows that a total of 100 student data, there are 20 students who get the Not Ready criteria, there are 2 students who get the Ready criteria but are categorized into the Not Ready criteria, and there are 2 students with the criteria. Need help but categorized as Not Ready.
2. The Ready criteria "4 33 0" shows that a total of 100 student data, there are 33 students who get the Ready criteria, there are 4 students who get the Not Ready criteria but are categorized into the Ready criteria, there are 0 students who get the Need Help criteria but are categorized into the Ready criteria.
3. The Need Help criteria "0 0 39" shows that a total of 100 student data, there are 39 students who get the Need Help criteria, there are 0 students who get the Not Ready criteria but are categorized into the Need Help criteria, and there are 0 students who get the Ready criteria but are categorized into the Need Help criteria.
4. Thus, the number of classifications carried out with correct results on a total of 100 student data is 20 + 33 + 39 = 92 and the number of classifications carried out with incorrect results is 2 + 2 + 4 = 8. So the accuracy of this model is (92/100) = 0.92 or 92%.
5. The following are the results of trials using WEKA tools on students' try out scores using the C4.5 algorithm. From 100 training data records used for testing with 33 attributes, accuracy, precision and recall values were obtained as in Figure 5.



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          92               92      %
Incorrectly Classified Instances         8                8      %
Kappa statistic                          0.8775
Mean absolute error                      0.0551
Root mean squared error                  0.2203
Relative absolute error                 12.6184 %
Root relative squared error             47.1642 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.833    0.053    0.833      0.833    0.833      0.781    0.931     0.826     Belum Siap
                 0.892    0.032    0.943      0.892    0.917      0.871    0.952     0.905     Siap
                 1.000    0.033    0.951      1.000    0.975      0.959    0.998     0.995     Butuh Bantuan
Weighted Avg.    0.920    0.037    0.920      0.920    0.919      0.884    0.965     0.921

=== Confusion Matrix ===

  a  b  c   <-- classified as
 20  2  2 |  a = Belum Siap
  4 33  0 |  b = Siap
  0  0 39 |  c = Butuh Bantuan
```

Figure 5. Details of Confusion Matrix Results Data
Source: author's data processing, 2023

4. CONCLUSION

After conducting research at MTs Nahdlatul Arifin, the author succeeded in concluding that the application of data mining through the application developed can help MTs Nahdlatul Arifin in classifying students based on readiness to face the Education Unit Examination. This process is carried out by taking the Try Out result value 5 times, so it is more efficient. Apart from that, the use of data mining can also reduce the accumulation of data that was previously underutilized. The number of classifications carried out with correct results on a total of 100 student data is 20 + 33 + 39 = 92 and the number of classifications carried out with incorrect results is 2 + 2 + 4 = 8. So the accuracy of this model is (92/100) = 0.92 or 92%.

REFERENCES

Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *01*, 17–24.

Andika, L. A., Azizah, P. A. N., & Respatiwulan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, *2*(1), 34. https://doi.org/10.13057/ijas.v2i1.29998

Astuti, R. (2018). Data Mining untuk Klasifikasi dengan Algoritma Cart (Classification and Regression Trees). *Media Informatika*, *17*(3), 114–124. https://doi.org/10.37595/mediainfo.v17i3.15

Budiman, I., Muliadi, & Ramadina, R. (2015). Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. *Jurnal Jupiter*, *7*(1).

Darmawan, I., Rachmatullah, S., Ariyanto, S. V., Hari, N. H., & Rohman, A. F. (2022). Implementasi Metode Clustering Pada Data Penerimaan Siswa Baru Menggunakan Algoritma K – Mean (Studi Kasus SMK Darul Falah Kertagena Tengah). *Insand Comtech : Information Science and Computer Technology Journal*, *7*(2). https://doi.org/10.53712/jic.v7i2.1679

Dewi, S. (2016). Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *Jurnal Techno Nusa Mandiri*, *XIII*(1), 60–66.

Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 5–10.

Iskandar, M., Rochman, A., Ratnawati, D. E., & Anam, S. (2019). Penerapan Algoritme C4 . 5 untuk Klasifikasi Fungsi Senyawa Aktif Menggunakan Kode Simplified Molecular Input Line System (SMILES). *Jurnal Pengetahuan Teknologi Informasi Dan Ilmu Komputer*, *3*(1), 761–769.

Julianto, W., Yunitarini, R., & Sophan, M. K. (2014). Algoritma C4.5 untuk Penilaian Kinerja Karyawan. *SCAN*, *IX*, 33–39.

Lumbantoruan, R., & Kennedy, P. S. J. (2015). Analisis Data Mining dan Warehousing. *Jurnal Ilmiah Buletin Ekonomi*, *19*(1), 66–76.

Mahartika, I. R., & Wibowo, A. (2019). Data Mining Klasterisasi dengan Algoritme K-Means

untuk Pengelompokkan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional. *Prosiding Seminar Nasional Sisfotek*.

Muliawan, A., Fauziyah, D. A., Hermansyah, M., & Wiranto, F. (2023). *Implementation and Testing of a Website Based New Student Admission System at Vocational High School Visi Global Jember*. *6*(1), 14–19.

Pambudi, R. H., Setiawan, B. D., & Indriati. (2018). Penerapan Algoritma C4 . 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(7), 2637–2643.

Rahmatullah, M. (2016). Kemampuan Mengajar Guru dalam Meningkatkan Kinerja Guru dan Hasil Belajar Siswa. *Jurnal Penelitian Manajemen Pendidikan*, *1*(2), 119–126.

Rahmayuni, I. (2014). Perbandingan Performansi Algoritma C4.5 dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang. *Teknoif*, *2*(1), 40–46.

Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembiayaan. *Jurnal Teknologi Informasi Magister Darmajaya*, *2*(01), 52–62.

Rusito, & Firmansyah, M. T. (2016). Implementasi Metode Decision Tree dan Algoritma C4.5 untuk Klasifikasi Data Nasabah Bank. *INFOKOM*, *1*(12).

Setio, P. B. N., Saputro, D. R. S., & Winarno, B. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.4. *Prosiding Seminar Nasional Matematika*, *3*, 64–71.

Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, *1*(2), 83–91. https://doi.org/10.36378/jtos.v1i2.24

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Yang, S., & Berdine, G. (2017). The Receiver Operating Characteristic (ROC) Curve. *The Southwest Respiratory and Critical Care Chronicles*, *5*(19), 34–36. https://doi.org/10.12746/swrccc.v5i19.391