KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 345-358

Comparison Of The C4.5 And Naïve Bayes Algorithm For Recommendations For Aid Recipients For The Smart Indonesian Program

Rahmat Hidayat[1], Miftahul Huda[2], Nanang Pradita[3]
[1,2,3]Ilmu Komputer, Universitas Putra Bangsa, Jawa Tengah, Indonesia
E-mail: [1]ayat150190@gmail.com, [2]hudablue11@gmail.com, [3]nanangpradita885@gmail.com

Abstract

Indonesian Smart Program (Program Indonesia Pintar "PIP") is a government backing program aimed at scholars who come from poor or vulnerable families to finance education. Scholars entering PIP correspond of two orders, scholars who have a Indonesian Smart Card (Kartu Indonesia Pintar) Indonesian Health Card (Kartu Indonesia Sehat) and scholars who are recommended by the academy to get an education. The conventional way carried out by seminaries in furnishing education recommendations takes a long time and is prone to mismatch recommendations so that it can affect in the distribution of PIP backing that isn't on target. From these problems the authors are interested in conducting exploration by exercising data mining ways. This exploration compares the C4.5 algorithm and the Naive Bayes algorithm. Testing was carried out on 139 SDN 03 Karanganyar pupil data. The results of the test set up that the C4.5 algorithm is better than the naive bayes algorithm. So that the Rule generated by the C4.5 algorithm can be used to make a decision-making system at SDN 03 Karanganyar.

Keywords: *Data Mining, Classification, C4.5, Naïve Bayes, PIP*

Abstrak

Program Indonesia Pintar (PIP) adalah program pemerintah yang ditujukan bagi para siswa yang berasal dari keluarga miskin atau rentan untuk membiayai pendidikan. Penerima beasiswa yang masuk PIP terdiri dari dua kategori,yaitu penerima beasiswa yang memiliki Kartu Indonesia Pintar (Kartu Indonesia Pintar), Kartu Indonesia Sehat (Kartu Indonesia Sehat) dan penerima beasiswa yang direkomendasikan oleh sekolah untuk mendapatkan beasiswa. Cara konvensional yang dilakukan sekolah dalam memberikan rekomendasi pendidikan memakan waktu yang lama dan rentan terhadap ketidaksesuaian rekomendasi sehingga berdampak pada penyaluran bantuan PIP yang tidak tepat sasaran. Dari permasalahan tersebut penulis tertarik untuk melakukan eksplorasi dengan melakukan cara data mining. Eksplorasi ini membandingkan algoritma C4.5 dan algoritma Naive Bayes. Pengujian dilakukan terhadap 139 data siswa SDN 03 Karanganyar. Hasil pengujian menunjukkan bahwa algoritma C4.5 lebih baik dibandingkan dengan algoritma Naive Bayes. Sehingga Rule yang dihasilkan oleh algoritma C4.5 dapat digunakan untuk membuat sebuah sistem pengambilan keputusan di SDN 03 Karanganyar.

Kata Kunci: *Data Mining, Klasifikasi, C4.5, Naïve Bayes, PIP*

1. Introduction

Education is commodity that every Indonesian citizen must have. Educating the nation is commanded in the preamble of the UUD 1945 Constitution as well as Composition 31 paragraph 1 of the 1945 Constitution which states that" Every citizen has the right to education". And it must be a priority in terms of education with at least 20 percent of the APBN and APBD for education according to the 1945 Constitution Article 31 paragraph

ISSN: 2720-992X

4 [1]. Despite the current data, it's still delicate for the government to balance the obligation to equate education throughout Indonesia to ameliorate the quality of education[2]. Until now, the powerhouse rate for abecedarian, inferior high, high academy, and public and private vocational seminaries in 2021 alone in Indonesia has reached 83.7 thousand [3]. And the loftiest powerhouse rate passed in 2017 reaching up to 187,828 [4].

Economics is one of the factors causing the uneven education that occurs in Indonesia, which clearly affects the low participation rate of the community in education[5]. The high cost of education presently results in the incapability of people with low husbandry to pierce education, performing in a gap [6]. One form of trouble made by the government in making guaranteed access to education by furnishing education backing aimed at children with low profitable families is the Indonesian Smart Program (*Program Indonesia Pintar*). With the end of minimizing or minimizing the possibility of dropping out of academy due to difficulty in paying for education so that they can continue at advanced education position[7] and at least adding academy participation of children progressed(6- 21 times) to get 12 times of natural education.

The Indonesian Smart Program (*Program Indonesia Pintar*) is backing issued by the Ministry of Education and Culture in the form of cash, expanded access, or learning openings handed to scholars from poor families with the end of helping to get an education. The Ministry of Education and Culture will match pupil data in Integrated Social Welfare Data (Data Terpadu Kesejahteraan Sosial) managed by the social service office with dapodik to determine the eligibility of PIP donors. seminaries manage to input pupil data in dapodik according to the PIP philanthropist criteria [8]. The criteria that are considered by the recommendation include several aspects, videlicet casing conditions, achievements, average report card scores, parents' income, and dependents. This is why Indonesian Smart Program (Program Indonesia Pintar) donors aren't only scholars who have Indonesian Smart Card (Kartu Indonesia Pintar) but also scholars who get recommendations from seminaries grounded on data contained in Basic Education Data (Data Pokok Pendidikan “dapodik”) grounded on Indonesian Smart Program (Program Indonesia Pintar) philanthropist criteria.

Seminaries registering scholars in Basic Education Data (Data Pokok Pendidikan “dapodik”) in opting Indonesian Smart Program (Program Indonesia Pintar) Donors still use conventional styles so that it takes a long time. As well as the possibility of incompatibility of recommendations with the criteria for Indonesian Smart Program (Program Indonesia Pintar) donors which can affect in the distribution of Indonesian Smart Program (Program Indonesia Pintar) programs that aren't on target [9], that the data set up by BPK in the 2018- 2020 Ministry of Education and Culture, the perpetration of the Indonesian Smart Program (Program Indonesia Pintar) program isn't acceptable, this is because the data used as a source of proposing prospective donors isn't dependable. This results in the distribution of backing for Indonesian Smart Program (Program Indonesia Pintar) not being right on target and there are still numerous children who should admit backing who don't admit it [8].

A analogous problem still occurs at SDN 03 Karanganyar, where the determination of recommendations made on numerous scholars is quite a burdensome workload for seminaries. With 139 scholars with different conditions and circumstances, of course, determining recommendations is delicate. Still, with this data, a fashion can be employed, videlicet data mining. Where the data mining fashion itself is the process of chancing knowledge or information hidden from data [10]. Grounded on the problems that do, bracket ways are the most applicable in opting donors of Indonesian Smart Program (Program Indonesia Pintar) backing.

The use of data mining ways is frequently used with the end of clusterization, vaccination, bracket [11]. As the base of the data analysis process, bracket is a fashion that can be used to gain new useful information from data by forming a rule or rule

KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 345-358

grounded on relating objects into classes, groups, or orders grounded on procedures, characteristics and delineations [12]. Student data at SDN 03 Karanganyar clearly has characteristics so that identification of groups or classes can be made grounded on the closeness of the attributes of the data. Algorithms in bracket styles that are popular moment are theC.45 algorithm and naive bayes [13].

Classification methods in data mining can be implemented on various data, with the C4.5 algorithm there is a decision tree and can process numerical and discrete data to produce rules or rules that are fast and easy to interpret[14], such as research conducted by (Astuti et al., 2023) which obtained a high accuracy value of up to 100% on classifying and grouping internet quota sales data using the C4.5 algorithm in 2021-2022. While the naïve bayes algorithm provides a practical way with a combination of data and is easy to apply ([15]. In recent research conducted by [16] using the naïve bayes algorithm, the resulting classification has a high enough accuracy of 89% in selecting data on non-cash food aid recipients. So, this research aims to compare the use of the C4.5 and naive bayes algorithms based on accuracy, precision and recall in Classifying Indonesian Smart Program (Program Indonesia Pintar) beneficiary recommendation data. The comparison carried out on the two algorithms is useful for knowing the best algorithm that can be used in classifying student data recommendations for Indonesian Smart Program (Program Indonesia Pintar) beneficiaries and as a supporting aspect in decision-making.

2. Reseach Methodology

This research was conducted on 139 SDN 03 Karanganyar student data which will be processed using data mining methods with classification techniques applying the c4.5 and naïve bayes algorithms. The following is a framework for the research conducted.

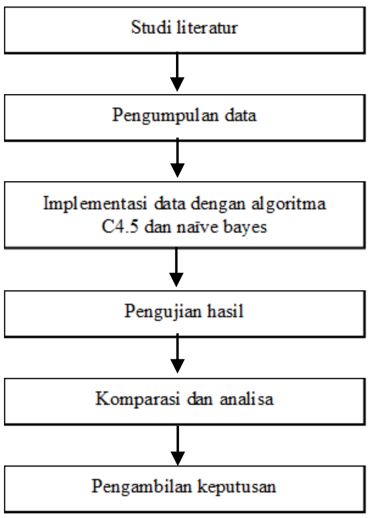

Figure 1. Research Framework

2.1. Literature study

To achieve the goal, this stage is carried out by searching for literature that can be used regarding data mining, classification, C4.5 algorithm, naïve bayes algorithm taken from various reading materials that can support research. The literature search is based on previous research to its development and the latest research

2.2. Data Collection

The data used uses student data contained in SDN 03 Karanganyar. Data registered into dapodik as a recommendation for Indonesian Smart Program (Program Indonesia

KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 345-358

Pintar) beneficiaries in the form of achievement data, average report card scores, parents' income, dependents, and housing conditions. The school plays an important role in entering student data for Indonesian Smart Program (Program Indonesia Pintar) beneficiary recommendations into dapodik [8]. All student data is used in this research.

2.3. Implementation C4.5 Algorithm

The C4.5 algorithm is a development of the ID3 algorithm discovered by Ross Quinlan, the C4.5 algorithm is used to classify data that has attributes in the form of numeric or categorical. The C4.5 algorithm is an algorithm for data classification. In the C4.5 algorithm before building a decision tree the most important thing to do is determine the attribute as the root [8]. Where the model rules or rules and decision trees formed in the C4.5 algorithm are easy to interpret and change [17].

In the C4.5 algorithm there are several stages in making a decision tree, namely:

1. Preparing training data. Training data is usually taken from historical data that has occurred before or called past data and has been grouped into certain classes
2. Calculating the root of the tree. The root will be taken from the attribute that will be selected, by calculating the gain value of each attribute, the highest gain value will be the first root. Before calculating the gain value of the attribute, first calculate the entropy value. To calculate the entropy value, the formula is used.

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

   Description:
   S: case set
   n: number of partitions
   pi: proportion of si to S
3. Calculate the gain value using the following equation

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

   Description:
   S: Case Set
   A: Fitur
   n: Number of partitions
   |Si|: Proportion of Si to S
   |S|: Number of a case
4. Repeat Step 2 and Step 3 until all records are partitioned.
5. The decision tree partitioning process will stop when:
   a. All records in node N get the same class.
   b. No attribute in the record is partitioned anymore
   c. There are no records in the same branch

2.4. Implementation Naïve Bayes Algorithm

The naive bayes algorithm is a simple probability-based prediction technique based on the application of the bayes rule with the assumption of strong dependence or independence [18]. This algorithm is also included in the classification algorithm as well as C4.5. Easy to use for machine learning data is an advantage of the naive bayes algorithm, then the naive bayes algorithm only requires one scan of the training data, and is used for handling missing attribute values and continuous data. The naïve bayes algorithm has another advantage, namely because it is easy to build, so it does not require complicated iterative estimation schemes and parameters which can be directly applied to large or large data [15]. In addition, the naïve bayes algorithm is easy to interpret and has consistent performance on a high-dimensional scale (Basuki, 2023). And this method can determine classification estimation parameters based on only relatively small training data

[19]. To classify data using the naïve bayes algorithm, you can use the following equation:

$$P(R|S) = \frac{P(R)P(S|R)}{P(S)} \quad (3)$$

Description:

R : Data for which the class is sought
S : Hypothesis on special class data
P(R|S) : Probability value based on hypothesis R that is based on condition S
P(R) : Probability value under hypothesis R
P(S|R) : Probability value of S based with hypothesis R
P(S) : Probability value of S

2.5. Result Testing

Testing of the results of data implementation carried out on both algorithms is carried out using rapid miner software. The testing mechanism is carried out using existing variables in the form of data from SDN 03 Karanganyar students, namely achievement, activeness, parents' income, number of dependents and student housing conditions. Testing is carried out on the C4.5 and naïve bayes algorithms with the same measurement standards so that a better algorithm can be found. The test results are presented in a confusion matrix or commonly referred to as a confusion table which is a method that can be used to calculate accuracy in data mining which can then also be used to find the precision and recall values of the calculation results. This table has a function to record the results obtained so that a match is found and is used to measure the performance of the classification method[20]. Performance measurement in the confusion matrix can be interpreted as follows.

Table 1. Confusion matrix

Class	**Classified Positive**	**Classified Negative**
Positive	TP (True Positive) correctly detected data	FN (False Negative) positive data that is detected incorrectly
Negative	FP (False Positive) negative data but detected as positive data	TN (True Negative) negative data detected correctly

To be able to know the performance of each algorithm tested, the next step is to calculate the accuracy, precision and recall values of the test results [21].

Accuracy is a depiction of how accurately an algorithm can classify data correctly. This means that the number of correct predictions divided by the total amount of data is formulated in the following equation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Precision is a description of the number of data with positive categories that can be classified correctly divided by the total data classified as positive with the following equation.

$$\text{Precision} = \frac{FP}{FP + TP} \times 100\ \% \quad (5)$$

Recall shows the percentage of correctly classified positive category data according to the following equation.

$$\text{Recall} = \frac{TP}{FN + TP} \times 100\ \% \quad (6)$$

2.6. Comparison and analysis

Comparisons are made on the results of calculations and tests based on the results of calculations carried out using rapid miner software, the amount of accuracy, precision,

KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 345-358

and recall being the benchmark for the performance of algorithms in calculating data. Comparisons made between the C4.5 algorithm and the naïve bayes algorithm can help in drawing conclusions. The analysis is carried out to be able to find out the information obtained from the comparison of the test results so that the algorithm with the best or efficient performance can be determined.

2.7. Conclusion Drawing

The end of this research is drawing conclusions based on calculations, comparisons, and tests carried out on results based on data owned by applying the c4.5 algorithm and the naïve bayes algorithm. Conclusions are drawn as a result of the research carried out as a red thread to achieve the objectives of this research.

3. Result and Discussion

The following are the results and discussion of this research:

3.1. Implementation of C4.5 Algorithm

The application of the C4.5 algorithm to SDN 03 Karanganyar student data will form rules and a decision tree [22]. Entropy calculations are performed on the total of all data and each attribute in the data. Gain calculation is also done on all attributes in the data. The largest attribute gain value is set as the root node in the first iteration or node 1. Total entropy in this study is sought by minus recommendations per amount of data multiplied by $\log_2$ of recommendations per amount of data plus minus no recommendations per amount of data multiplied by log2 of no recommendations per amount of data. When looking for the entropy value per attribute, the amount of data is replaced with the number of recommendations and not recommendations on the attribute value. The calculation of the gain value is sought by the total entropy minus the number of attribute values per amount of data multiplied by the entropy of the attribute minus the results of the calculation in the same way on each attribute. The calculation results are presented in the following table

Table 2. Node 1

Attribute	Total	Recommendations	Not	Entropy	Gain
	139	51	87	1.271	
Average score					
≥ 90	44	26	18	1.624	
≥ 80	83	29	59	1.096	
≥ 70	12	1	11	0.441	
					0.064
achievements					
Many	48	19	29	1.315	
Enough	50	14	36	1.055	
Less	41	18	23	1.397	
					0.025
Income					
≥ 4 million	8	0	8	0.000	
≥ 3 million	12	3	9	0.978	
≥ 2 million	27	8	19	1.096	
< 2 million	92	40	52	1.390	
					0.054
Dependents					
>5	30	11	19	1.255	
<4	109	40	69	1.256	
					0.015
House Condition					
Less feasible	41	34	7	1.747	
Feasible	98	17	81	0.758	
					0.221

KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 345-358

From the calculation of the gain value on each of these attributes, a comparison is made and it can be concluded that the gain value of the house condition is the highest gain value, so the condition of the house becomes the root node [23]. The house condition attribute has 2 values, namely less and feasible, to determine the next node, it is necessary to do calculations on less and feasible values, because the classification of both values is not absolute recommendation or not. It can be interpreted that even though the condition of the house It is a priority to determine beneficiaries but it is not absolute that students with decent home conditions cannot receive assistance because consideration is also based on several other supporting attributes. In the next node, the values of the amount of data and total entropy are replaced with the number and entropy of the attribute with the greatest gain in the previous calculation. The attribute of the condition of the house is not recalculated because it has become a root note and so on every iteration of the highest gain value will be taken and leave another attribute, calculated until the last attribute remains.

Table 3. Node 2

Attribute	Total	Recommendations	Not	Entropy	Gain
	41	34	7	1.747	
Average score					
≥ 90	13	13	0	0	
≥ 80	24	20	4	1.745	
≥ 70	4	1	3	0.977	
					1.417
achievements					
Many	12	10	2	1.745	
Enough	12	10	2	1.745	
Less	17	14	3	1.748	
					1.231
Income					
≥ 4 million	0	0	0	0	
≥ 3 million	0	0	0	0	
≥ 2 million	1	1	0	0	
< 2 million	33	26	7	1.752	
					1.331
Dependents					
>5	8	8	0	0	
<4	33	26	7	1.752	
					1.331

The value attribute becomes the gain with the highest value, in the value attribute a comparison is made between the amount of entropy and the attribute value with the highest amount of entropy is taken.

Table 4. Node 3

Attribute	Total	Recommendations	Not	Entropy	Gain
	24	20	4	1.745	
achievements					
Many	12	10	2	1.745	
Enough	12	10	2	1.745	
Less	17	14	3	1.748	
					1.230
Income					
≥ 4 million	0	0	0	0	
≥ 3 million	0	0	0	0	
≥ 2 million	1	1	0	0	
< 2 million	33	26	7	1.748	
					1.242

Attribute	Total	Recommendations	Not	Entropy	Gain
Dependents					
>5	8	8	0	0	
<4	32	25	7	1.751	
					189.1

The dependent attribute of parents has the highest gain value at iteration or node 3, so the remaining two attributes, namely income and achievement.

Table 5. Node 4

Attribute	Total	Recommendations	Not	Entropy	Gain
	32	25	7	1.751	
achievements					
Many	6	4	2	1.698	
Enough	5	4	1	1.752	
Less	6	5	1	1.745	
					1.572
Income					
≥ 4 million	0	0	0	0	
≥ 3 million	0	0	0	0	
≥ 2 million	1	1	0	0	
< 2 million	16	12	4	1.745	
					1.550

The last remaining attribute is parental income, which became the last node in the calculation using the application of the C4.5 algorithm in this study.

Table 6. Node 5

Attribute	Total	Recommendations	Not	Entropy	Gain
	5	4	1	1.752	
Income					
≥ 4 million	0	0	0	0	
≥ 3 million	0	0	0	0	
≥ 2 million	1	1	0	0	
< 2 million	16	12	4	1.745	
					1.551

From the calculations made, a decision tree will be formed. Based on the decision tree, it can be used in decision making by interpreting solutions to problems that arise based on existing rules [24]. Decision trees play a role in exploring data, finding hidden relationships between variables in the data. The decision tree in this study is of course about recommendations based on the priority level of entropy and gain calculations on SDN 03 Karanganyar student data. Similarly, in making a decision will be influenced by many factors making considerations before making decisions.

KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 345-358

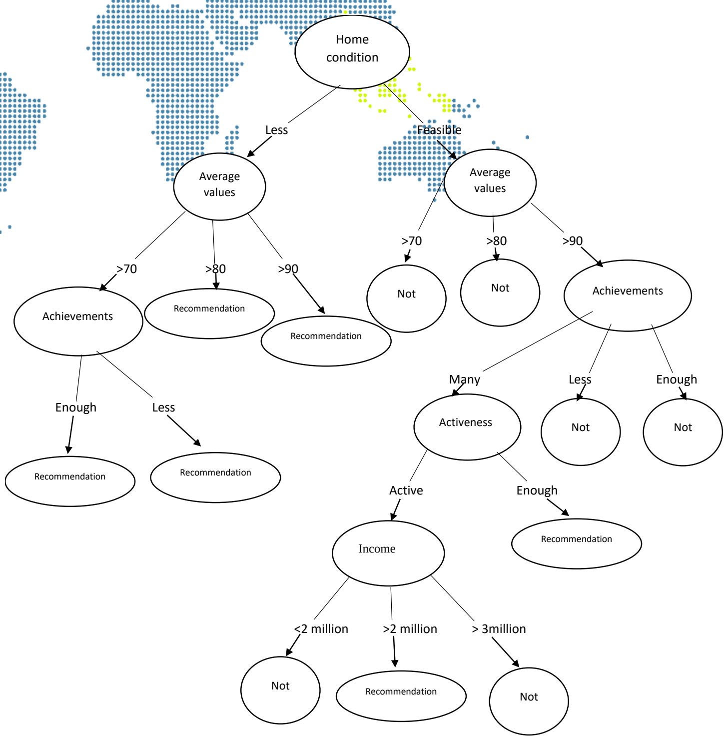

Figure 2. Decision Tree

3.2. Implementation of Naïve Nayes Algorithm

This algorithm is an algorithm that performs simple probabilistic classification and calculates the probability of a set of data by summing the frequencies or attributes of existing data. The method contained in the naïve bayes algorithm needs to be known that in the classification process requires instructions to determine the class or category for the data analyzed, therefore data testing and training data are needed. This naïve bayes algorithm is widely used in text classification that has problems with various classes, this is because it takes relatively little time to manage data. Starting with calculating the total probability value or prior probability in the training data, namely in this study with recommendation attributes or not recommendations divided by the amount of data. Probability calculation is the focus of this algorithm [25]so it is carried out on all attributes owned.

Table 7. Probability Prior

(P(Ci))	**Recommendations**	**Not**
	0.366	0.633

In Naïve Bayes, this method is also used to perform probability predictions of different classes based on attributes. In this study, the determination of attribute probability is carried out on each value on the attribute divided by the number of attributes included in the recommendation or not recommended.

Table 8. Probability Attribute

(P(Ci))	**Recommendations**	**Not**
	0.366	0.633
Average Score		
75	0.020	0.125
80	0.314	0.420
85	0.157	0.250
90	0.431	0.170
95	0.078	0.034

(P(Ci))	Recommendations	Not
Achievements		
Less	0.353	0.261
Enough	0.275	0.409
Many	0.373	0.330
Parents Income		
800000	0.157	0.068
850000	0.078	0.011
900000	0.235	0.080
950000	0	0.023
1000000	0.196	0.148
1100000	0.039	0.023
1200000	0.039	0.080
1300000	0	0.045
1400000	0.020	0.023
1500000	0.020	0.091
2000000	0.059	0.114
2500000	0.098	0.102
3000000	0.059	0.034
3500000	0	0.068
4000000	0	0.023
4200000	0	0.034
4300000	0	0.011
4500000	0	0.023
Dependents		
2	0.019	0.011
3	0.313	0.295
4	0.450	0.477
5	0.156	0.204
7	0.058	0.011
Home Condition		
Less Feasible	0.666	0.079
Feasible	0.333	0.920

Calculation with the naïve bayes algorithm is followed by calculating the probability of testing data based on the results of the probability of each attribute of the training data. Where in this study each data is calculated to the possibility of being recommended or not recommended, if one of the two classes is higher in grade, it is included in the class with the highest score. The calculation of testing data is carried out for the recommendation prediction class as follows.

P(Student Name $_{Recommendation}$) = P(Average Report Score $_{Recommendation}$) x P(Achievements $_{Recommendation}$) x P(parents Income $_{Recommendation}$) x P(Dependents $_{Recommendation}$) x P(House Condition $_{Recommendations}$)

Meanwhile, the calculation on the testing data for the prediction class is not recommended as follows.

P(Student Name $_{Not\ Recommendation}$) = P(Average Report Score $_{Not\ Recommendation}$) x P(Achievements $_{Not\ Recommendation}$) x P(parents Income $_{Not\ Recommendation}$) x P(Dependents $_{Not\ Recommendation}$) x P(House Condition $_{Not\ Recommendations}$)

The result of the calculation is a prediction obtained in the application of the naïve bayes algorithm. Of course, there can be found disapproval of the resulting class because of the accuracy value possessed by the naïve bayes algorithm

itself. The naïve bayes algorithm is classified as an algorithm that is easy and fast in predicting data.

3.3. Testing Results

In conducting testing, of course, it must be done with uniform assessment standards in order to find out the best algorithm through comparisons made. Testing conducted in this study by doing calculations manually with Microsoft Excel. And using rapid miner software that applies the C4.5 algorithm with the naive bayes algorithm with the features owned by the rapid miner software as needed to determine the prediction results of the C4.5 algorithm and naïve bayes. Based on calculations made from manual methods and using rapid miner software shows the same results.

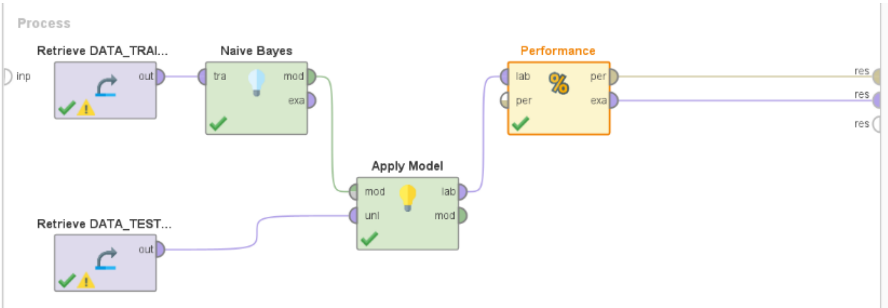

Figure 3. Calculation of riped miner algorithm Naïve Bayes

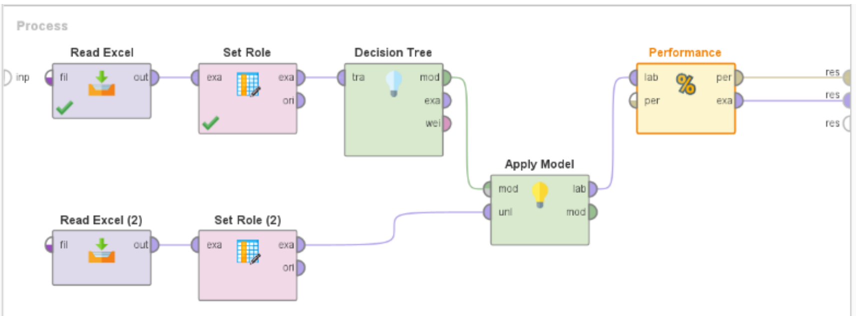

Figure 4. Calculation of RapidMiner algorithm C4.5

Below are the results of calculations performed using RapidMiner by applying C4.5 and Naïve Bayes algorithms to identical training data and test datasets.

Table 9. Confusion table of test results on rapid miner with C4.5 algorithm

Predicted	**Recommendations**	**Not**
Recommendation	48	8
Not	3	80

Table 10. Confusion table of test results on rapid miner with Naïve Bayes algorithm

Predicted	**Recommendations**	**Not**
Recommendation	35	9
Not	16	79

To evaluate the performance of each algorithm tested, the next step is to calculate the accuracy result value, the precision result value, and the recall value of the calculation results. The test results on the data are as follows.

Table 11. Confusion table

Algorithm		
Testing	C4.5	Naïve Nyes
Accuracy	92%	82%
Precision	94%	68%
Recall	85%	79%

From tests conducted on data applying the C4.5 algorithm and naïve bayes based on accuracy, it can be analyzed that the accuracy obtained by the C4.5 algorithm is 92% higher than the accuracy obtained from the naïve bayes algorithm by 82%. Accuracy is obtained from the right number of predictions based on testing data and predictions from training data divided by the total amount of data can be concluded that the accuracy of data calculations using the C4.5 algorithm is superior, because it is able to classify data with more correct [26].
Based on the results of tests conducted by calculating the precision of the results of calculating the prediction of Indonesian Smart Program (Program Indonesia Pintar) beneficiaries with the application of the C4.5 algorithm and naïve bayes, it can be analyzed that the precision value produced by the C4.5 algorithm is 94% and the precision value produced by the Naïve bayes algorithm is 68%. The precision produced by the C4.5 algorithm is higher than that produced by the Naïve Bayes algorithm. In testing by calculating the precision value, it can be concluded that the c4.5 algorithm is better than the naïve bayes algorithm, this is because the greater the precision results produced, the possibility of changes in each calculation will be smaller [27], so that the results of the calculation will have the opportunity to have the same or consistent results even though several repetitions or experiments are calculated.

Tests conducted by calculating the recall value, the calculation of data using the C4.5 algorithm gets a value of 85% and the results of calculating recall using the naïve bayes algorithm of 79%. From these results, it can be interpreted that the C4.5 algorithm is better than the naïve bayes algorithm. This is because the amount of recall value shows that the amount of data with a positive category [28] if in this study is data that is classified as recommended can be classified correctly using the C4.5 algorithm.

From the tests carried out by calculating the accuracy, precision, and recall of the results of the two algorithms used, namely C4.5 and naïve bayes, the C4.5 algorithm is better at testing accuracy, precision and recall, this is because the value of the test results obtained by the C4.5 algorithm is higher than the naïve bayes algorithm. This test was conducted to support the decision-making process and determine a better algorithm between C4.5 and Naïve Bayes in calculating the prediction of Indonesian Smart Program (Program Indonesia Pintar) beneficiary recommendations.

4. Conclusion

By referring to the results of calculations and trials carried out, it can be concluded that. Implementation can be carried out effectively on student data at SDN 03 Karanganyar to determine recommendations for Indonesian Smart Program (Program Indonesia Pintar) beneficiaries by utilizing the C4.5 algorithm and the Naïve Bayes algorithm. Based on tests carried out on the results of recommendations for Indonesian Smart Program (Program Indonesia Pintar) beneficiaries, the C4.5 algorithm was better in testing with an accuracy value of 92%, a precision value of 94%, and a recall value of 85%. From this comparison it can be concluded that the C4.5 algorithm is more effective in producing recommendations for Indonesian Smart Program (Program Indonesia Pintar) beneficiaries based on student data at SDN 03 Karanganyar. This can be seen from the results of higher accuracy measurements, showing the ability to classify data more accurately and with greater volume. In addition, the algorithm is able to classify positive data better, which in the context of this research refers to recommended data. Apart from that, the C4.5 algorithm is also quite good and fast in classifying data and has a higher precision value compared to the Naïve Bayes algorithm where the results obtained tend to be more consistent because the results obtained have relatively small differences in each calculation. even though it has been done many times. general. By referring to the calculations, standard results, data trials, and calculations carried out in this research, it can be concluded that the C4.5 algorithm and the Naïve Bayes algorithm can be implemented in designing the development of a decision-making system for decision making recommendations to students. The rules generated from these two algorithms can be a basic reference for decision making program rules that can be applied at SD N Karanganyar.

References

[1] Syafrida, "No" vol. 105, no. 3, pp. 129–133, 1945.

[2] F. P. Sianipar and S. T. Maulia, "Kebijakan Pendidikan : Pemerataan dan Perluasan Akses (Rehabilitasi Ruang Kelas Yang Rusak)," *J. Pendidik. Sastra Ingg.*, vol. 3, no. 1, pp. 191–198, 2023.

[3] D. I. Desa, T. Seteko, K. Ogan, I. N. Tanjung, S. Village, and O. I. District, "JOLL 6 (1) (2023) Journal of Lifelong Learning," vol. 6, no. 1, 2023.

[4] F. S. Octavianti, "Implementasi Konvensi Hak Anak Di Kota Jakarta Utara (Studi Kasus: Pemenuhan Hak Asasi Manusia Perspektif Pendidikan Bagi Pelajar SMA)," *Pedagog. J. Ilmu Pendidik.*, vol. 23, no. 1, pp. 30–37, 2023, doi: 10.24036/pedagogi.v23i1.1476.

[5] N. Nurokhmah, "Program Indonesia Pintar (PIP): Implementasi Kebijakan Kesejahteraan dalam Upaya Meningkatkan Angka Partisipasi Pendidikan," *J. Paradig. J. Multidisipliner Mhs. Pascasarj. Indones.*, vol. 2, no. 1, pp. 37–48, 2021, [Online]. Available: https://jurnal.ugm.ac.id/paradigma/article/view/66256

[6] N. E. Rohaeni and O. Saryono, "Implementasi Kebijakan Program Indonesia Pintar (PIP) Melalui Kartu Indonesia Pintar (KIP) dalam Upaya Pemerataan Pendidikan," *J. Educ. Manag. Adm. Rev.*, vol. 2, no. 1, pp. 193–204, 2018.

[7] R. Purnama and I. Izzatusholekha, "Analisis Program Indonesia Pintar dalam Mengurangi Putus Sekolah di Kabupaten Bandung," *J. Manaj. dan Ilmu Adm. Publik*, vol. 5, no. 1, pp. 66–73, 2023, doi: 10.24036/jmiap.v5i1.600.

[8] S. B. Arminy, N. Mutiah, and R. P. Sari, "Penentuan Penerimaan Beasiswa PIP Menggunakan Metode Moora Pada SD Negeri 11 Sandai," vol. 11, no. 1, pp. 43–50, 2023.

[9] Lidia Lusiana, "Implementasi Program Indonesia Pintar (PIP) Di Desa Sukomulyo Kecamatan Sepaku Kabupaten Penajam Paser Utara (Studi Kasus di SDN 011 dan SDN 013)," *eJournal Adm. Negara*, vol. 6, no. 1, pp. 6991–7005, 2017, [Online]. Available: https://ejournal.ap.fisip-unmul.ac.id/site/wp-content/uploads/2018/01/Jurnal LIDYA (01-17-18-07-33-51).pdf

[10] J. S. Parapat and A. S. Sinaga, "Data Mining Algoritma C4.5 Pada Klasifikasi Kredit Koperasi Simpan Pinjam," *J. Ilmu Tek. Elektro Komput. dan Inform.*, vol. 4, no. 2, pp. 144–154, 2018.

[11] A. Primajaya, B. N. Sari, and A. Khusaeri, "Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4.5 Studi Kasus Provinsi Kalimantan Barat," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 188, 2020, doi: 10.26418/jp.v6i2.37834.

[12] D. Jollyta, G. Gusrianty, and D. Sukrianto, "Analysis of Slow Moving Goods Classification Technique: Random Forest and Naïve Bayes," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 5, no. 2, pp. 134–139, 2019, doi: 10.23917/khif.v5i2.8263.

[13] Y. I. Kurniawan, U. M. Surakarta, and N. Bayes, "Comparison of Naive Bayes and C . 45 Algorithm in Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 455–464, 2018, doi: 10.25126/jtiik.

[14] R. Haqmanullah Pambudi, B. Darma Setiawan, and Indriati, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 7, pp. 2637–2643, 2018, [Online]. Available: http://j-ptiik.ub.ac.id

[15] A. H. Wicaksono *et al.*, "Klasifikasi Siswa Slow Learner Untuk Mendukung Sekolah Algoritma Naïve Bayes Classification of Slow Learner Students To Support Schools in Improving Student Understanding Using the Naïve Bayes," vol. 9, no. 3, pp. 589–596, 2022, doi: 10.25126/jtiik.202295509.

[16] M. M. Alfitri, N. Nurahman, and ..., "Evaluasi Performa Algoritma Naïve Bayes Dalam Mengklasifikasi Penerima Bantuan Pangan Non Tunai," *J. Media ...*, vol. 7, pp. 1433–1445, 2023, doi: 10.30865/mib.v7i3.6151.

[17] S. Anggraini, S. Defit, and G. W. Nurcahyo, "Analisis Data Mining Penjualan Ban Menggunakan Algoritma C4. 5," *J. Ilmu Tek. Elektro ...*, vol. 5, no. 2, pp. 0–7, 2018, [Online]. Available: https://core.ac.uk/download/pdf/295348196.pdf

[18] P. Subarkah, E. P. Pambudi, and S. O. N. Hidayah, "Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 1, pp. 139–148, 2020, doi: 10.30812/matrik.v20i1.826.

[19] M. K. Anam, B. N. Pikir, and M. B. Firdaus, "Penerapan Na¨ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 139–150, 2021, doi: 10.30812/matrik.v21i1.1092.

[20] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.

[21] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.

[22] K. Halim, D. Erny Herwindiati, and T. Sutrisno, "Penerapan Metode Decision Tree Untuk Prakiraan Cuaca Kota Bekasi," *J. Ilmu Komput. dan Sist. Inf.*, vol. 11, no. 2, pp. 1–5, 2023, doi: 10.24912/jiksi.v11i2.26026.

[23] Z. Abidin, "Eka Nurhana 2) , Permata 3) , Faruq Ulum 4) 1,2,3 Sistem Informasi, Universitas Teknokrat Indonesia 4 Informatika, Universitas Teknokrat Indonesia 1,2,3,4 JL," *ZA. Pagar Alam*, vol. 17, no. 2, pp. 567–582, 2023, [Online]. Available: www.kaggle.com

[24] P. R. Adinda, "Pohon Keputusan C4. 5 Algoritma Untuk Klasifikasi Program Bantuan Belajar," *J. Portal Data*, vol. 2, no. 9, pp. 1–13, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/215%0Ahttp://portaldata.org/index.php/portaldata/article/download/215/204

[25] A. Yudhana, I. Riadi, and M. R. Djou, "Pengembangan Layanan Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1062, 2022, doi: 10.30865/jurikom.v9i4.4515.

[26] F. Y. Rahman, I. I. Purnomo, and N. Hijriana, "Penerapan Algoritma Data Mining Untuk Klasifikasi Kualitas Air," *Technol. J. Ilm.*, vol. 13, no. 3, p. 228, 2022, doi: 10.31602/tji.v13i3.7070.

[27] I. Solihat, A. P. Tirta, A. P. Ramdani, and A. Nandang Roziafanto, "Verifikasi Metode Pengujian Kadar Nitrit dalam Air Limbah Secara Spektrofotometri UV-Visibel," *KOVALEN J. Ris. Kim.*, vol. 8, no. 1, pp. 53–59, 2022, doi: 10.22487/kovalen.2022.v8.i1.15756.

[28] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.