Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison Of Naïve Bayes and Decision Trees in Determining the Best Manager of Nurul Jadid Islamic Boarding School Based on Forward Selection

Farhan Dardiri*
Univesrsitas Dian Nuswantoro, Indonesia
Farhan.f4124n@gmail.com

Submitted :Feb 16, 2024 | **Accepted** : Mar 1, 2024 | **Published** : Apr 1, 2024

Abstract: : In an effort to find a solution for determining the best administrators, Islamic boarding school administrators try to determine the nominations for the best administrators using existing service data and knowledge. The process of determining nominations for the best administrators is less accurate, requiring computational methods to classify which administrators fall into the best category. In the context of data mining, classification is an important aspect. One of the classification models used is Naïve Bayes which focuses on class probability, and Decision Tree C4.5 which produces a decision tree to determine the priority of indicators that are most influential in predicting the best management. Both of these algorithms have their respective advantages. This research aims to analyze and compare the performance of the Naïve Bayes and Decision Tree classification algorithms. The comparative results of testing the Naïve Bayes and C4.5 algorithms in determining the nominations for the best administrators at the Nurul Jadid Paiton Probolinggo Islamic Boarding School on 455 administrator data tested in this study show that there is a fairly large comparison of accuracy. Naïve Bayes with Forward Selection has an accuracy rate of 91.21%, higher than Naïve Bayes itself whose accuracy results are only 87.64%. there is a difference of 3.57%. Likewise, the accuracy of C4.5 with Forward Selection has an accuracy rate of 90.99%, higher than C4.5 alone which has an accuracy rate of 90.11%. there is a difference of 0.88%. So in the comparison between 4 algorithm model trials, Naïve Bayes and Forward Selection had the most dominant accuracy with an accuracy result of 91.21%.

Keywords: Classification, Naïve Bayes, C4.5 Decision Trees, Forward Selection.

INTRODUCTION

Education in Islamic boarding schools is not much different from education outside Islamic boarding schools. There are many similarities between subjects, learning curricula and school/college regulations. Because the curriculum at Islamic boarding schools also follows the government curriculum. The implementation of this curriculum cannot be separated from the important role of the Islamic boarding school community, including administrators, teachers, employees and regional administrators who have dedicated themselves to Islamic boarding schools. The participation of Islamic boarding school servants must of course be appreciated with an award, the implementation of which is still confusing. After the Islamic boarding school administrators hold an internal meeting, one way to find a solution to the problem above, namely determining the nomination for the best administrator, is to use service data and explore the knowledge contained therein to obtain the main attributes that influence quality administrator performance can be known (Sahlan, 2023).

Several attributes that can be used to determine nominations for the best management are length of service and courtesy, totaling 17 attributes. Existing Islamic boarding school data can be utilized optimally according to needs and can be processed into useful information so that the relationship between its attributes can be known. data that can be analyzed and is expected to produce output in the form of management performance related to the period of service (Angga, 2020). Determining the best nominations for the best administrators will increase interest and provide motivation for other administrators. Therefore, knowledge is needed that can predict problems that occur (Zakaria et al., 2023). The process of determining nominations for the best administrators is less accurate, requiring computational methods to classify which administrators fall into the best category. In the context of data

* Corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

mining, classification is an important aspect. One of the classification models used is Naïve Bayes which focuses on class probability, and Decision Tree C4.5 which produces a decision tree to determine the priority of indicators that are most influential in predicting the best management. Both of these algorithms have their respective advantages (Suhendar et al., 2024).

Research related to prediction problems has been carried out using various classification methods, such as research conducted by Mochamad Subhan (Muhammad & Zakaria, 2024). This research discusses Forward Selection to improve prediction performance (Naïve Bayesin) using registration datasets at SMK Buduran Sidoarjo. The application of the Forward Selection selection feature aims to identify influential attributes and increase classification accuracy in the Naïve Bayes algorithm. Therefore, it is hoped that the Naïve Bayes method will become the standard method used by the management of Vocational High Schools (SMK) in the student affairs sector and it is hoped that it can also become a reference for using the Naïve Bayes Algorithm model for others. practitioners or researchers to apply to other cases. After carrying out the testing process using the Naïve Bayes method and also testing using the combined Naïve Bayes method with Forward Selection, different accuracy results were obtained. If you only use the Naïve Bayes method without the selection feature, the accuracy is 86.22% and after testing Naïve Bayes with the genetic algorithm selection feature, the accuracy is 99.49%, resulting in a significant increase, namely 13.27%.

Khairunnazri also explained in his research that he compared algorithms for classifying diabetes using the Pima Indian Diabetes Dataset (PIDD) dataset (Folorunso et al., 2023). This research creates a classification model using the classification algorithm, namely Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Neural Network (NN) using a public dataset obtained from the Pima Indian Diabetes Dataset (PIDD), totaling 786 records, testing All records use the Rapid Miner tool, the results of which are used to measure algorithm performance. The Neural Network algorithm has the highest accuracy between the two algorithms, with an accuracy value of 76.3%, in second place is the Naïve Bayes (NB) algorithm with an accuracy value of 75.5%, and in third place is the K-Nearst Nightbor algorithm. algorithm with an accuracy value of 73.4%.

From the explanation above, not all algorithms can work well because each algorithm has its own advantages and disadvantages in certain cases or research. So the researcher aims to analyze and compare the Naïve Bayes method, Decision Tree based on Forward Selection to find out which method has a higher level of accuracy in calculating nominations to determine the best administrator and applying the Forward Selection selection feature to determine influential attribute models. and can improve the second classification accuracy results. algorithm.

LITERATURE REVIEW

Hamdani (Hamdani et al., 2024). in his research discusses the difficulties that exist in determining health development priorities based on podes data in classifying podes data into 3 village statuses and also the absence of a classification model that is used to determine priority indicators for health services. So we will carry out a priority analysis of the main indicators of development in the health sector using the classification method with the Decision Tree C4.5 algorithm. The aim of this research is to provide input to determine which development in the health sector is more important, so that errors do not occur in implementing development and also to create a classification model to produce a decision tree that can be applied in development in the health sector. health sector. In this research, a model was tested using K-means Euclidean to group the 2014 IPD data into 3 groups. The resulting model was tested to obtain a Bouldin Index value from each algorithm so that the test results obtained using data after testing with the rapidminer tool produced a Bouldin Index value of 1.694 using the Eculidean K-means method. After using K-means to determine indicator priorities, the algorithm used is Decision Tree C4.5. The results of testing the C4.5 algorithm obtained an accuracy of 96.52% and class recall of 87.67% for class 2, 98.05% for class 1, 99.19% and class 0 so that it becomes a priority indicator for development in the health sector in IPD Data 2014 can be determined, namely indicator I_9.

Apart from that, Rahmayani (Rahmayani, 2023). in his research discusses how to determine the completion of a student's studies by correlating the student's initial entry data with the academic achievement of Community Academy students using the C4.5 Algorithm based on Forward Selection and how to increase prediction accuracy using the C4.5 Algorithm based on Forward Selection. The aim of this research is to predict the graduation of Community Academy students using the Forward Selection-based C4.5 algorithm method approach by conducting experiments on training data, observing the ability to consume a model in the form of a decision tree. And this research is expected to improve the performance accuracy of the C4.5 algorithm using the Forward Selection method. So this research can predict student study completion by tracing the relationship between students' initial entry data and academic achievement, where previously this data was only hidden in a warehouse where it could be processed into valuable and useful information for the Academy Community in this case. uses a decision tree approach in building a model using the C4.5 algorithm based on Forward Selection. The dataset used in this research consists of 175 data records.

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

Data collection is the first step in a research, where the data source in this research is taken from the Nurul Jadid Islamic Boarding School board exam dataset. The data obtained is in the form of univariate time series data. Some of these datasets are in softcopy (.xlsx) form and some are in hardcopy form. So data in hardcopy form must be retyped to be processed into a model. In determining the comparison of which prediction model is the best between 2 classification algorithms, namely Decision Tree C4.5 and Naïve Bayes based on preprocessing data on the best administrator dataset at the Nurul Jadid Paiton Probolinggo Islamic Boarding School. The proposed model design includes data preprocessing (Forward Selection) which can be depicted in the image below:

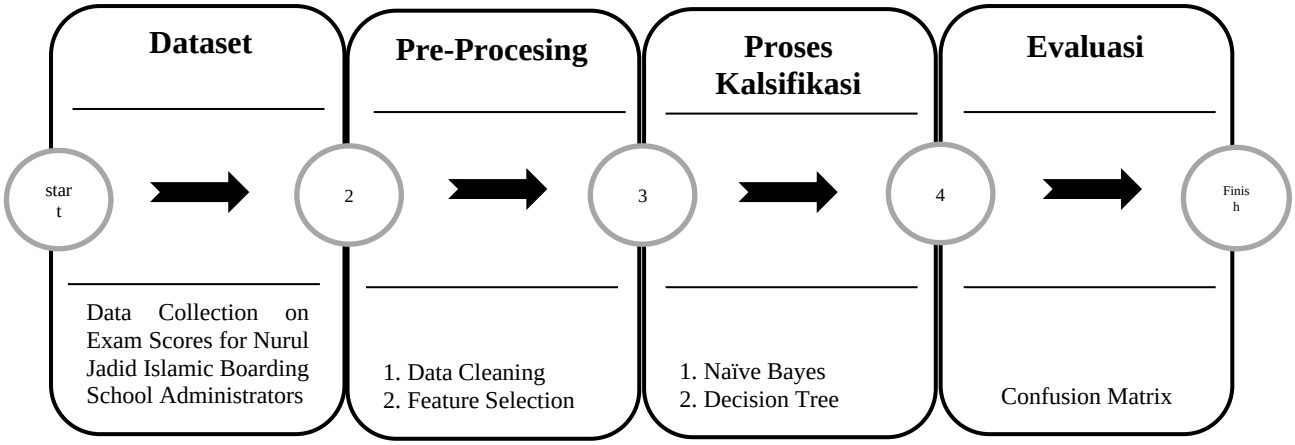

Figure 1 Proposed model

Evaluation with Confusion Matrix

Accuracy confusion matrix as a comparison of performance measurements in the two models later. In general, performance is measured through the accuracy obtained from the validation process using cross validation 10 times, so that the model formation can be directly tested by testing 10 times. Evaluation using the confusion matrix produces accuracy, precision and recall values. Accuracy in classification is the percentage of accuracy of data records that are classified correctly after testing the classification results. Precision or confidence is the proportion of cases that are predicted to be positive and are also true positive in the actual data. Recall or sensitivity is the proportion of actual positive cases that are correctly predicted to be positive (Dutra & Dawson, 2023).

Correct Classification	*Classfied as*	
	+	-
+	*True Positives*	*False Negatives*
-	*False Positives*	*True Negatives*

Table 1 Confusion Matrix

The following is the confusion matrix model equation:

$$akurasi = \frac{tp+tn}{tp+tn+fp+fn} \quad \text{“(1)”}$$

$$sensitivity = \frac{tp}{tp+fn} \quad \text{“(2)”}$$

$$specificity = \frac{tn}{tn+tp} \quad \text{“(3)”}$$

$$PPV = \frac{tp}{tp+fp} \quad \text{“(4)”}$$

$$NPV = \frac{tn}{tn+fn} \quad \text{“(5)”}$$

Validation with Cross Validation

The depiction of the model algorithm in Figure 3.2 will be a reference for comparison in this research. The existing dataset will be divided into X parts according to the X-Fold Cross Validation validation concept. Each part of the dataset will then be used as testing data, while the remaining dataset will be used to train the model that will be built. Each model will use training data that has been processed according to the selected data pre-

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

processing method. This is done as part of the training process for each learning algorithm that has been selected. The next step is to test the two algorithms that have been tested and trained with testing to carry out validation. The validation results were used to measure the performance of the two research models, and a comparison was carried out to find the model that had the best performance using accuracy confusion matrix analysis (Castro & Oliveira, 2023).

RESULT

Before the data is processed, preprocessing is the initial step where the dataset is cleaned from values that are still inconsistent, such as noisy data, missing data or so on.

No	*Tauhid integer*	Management period *integer*	Active Devotion *integer*	Long time at Islamic boarding school *integer*	Foster Guardian *polynominal*	Nomination *Binominal label*
1	88	4	2	10	Tidak	Tidak
2	85	4	2	7	Tidak	Tidak
3	83	4	1	10	Tidak	Ya
4	89	4	1	10	Tidak	Tidak
5	75	4	3	10	Tidak	Tidak
6	75	4	2	10	Tidak	Tidak
7	87	4	1	10	Tidak	Tidak
8	75	6	1	12	Tidak	Tidak
9	85	6	1	9	Ya	Ya
10	90	6	1	12	Tidak	Tidak
11	80	6	1	9	Ya	Tidak

Table 2 Entering the dataset in the RapidMinner tool

Prepare exam score data for the 2023 Nurul Jadid Islamic Boarding School regional administrators. Import the dataset in the tools then format the attributes and determine the data type according to the value, such as numeric, binominal, polynominal and label types. After the data has been successfully imported, check again to see if any data is missing (needs cleaning) or noisy (incorrect values), if not. Then we can pay attention to the data that has been imported.

Figure 2 Operator Selection Attributes in RapidMinner

Determine the attributes that influence and do not influence the output of the information to be searched, using the Select Attribute operator with subset selection

Naïve Bayes Algorithm with Forward Selection

The modeling stage aims to compare algorithms between Naïve Bayes and Naïve Bayes based on Forward Selection in determining classification to determine the best management nominations using RapidMiner software (Nurhalisha et al., 2023). Forward selection modeling uses RapidMiner tools as follows:

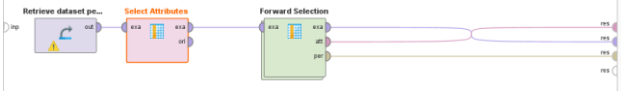

Figure 3 Modeling of Naïve Bayes and Forward Selection algorithms using RapidMiner tools

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The results of modeling using RapidMiner obtained six influential attributes. The following are the attributes of the results of applying Feature Forward Selection which are presented in Figure 5

Attribute Weight	BTQ	Fiqh	Akhlaq	Tauhid	Aktif Pengabdian	Wali Asuh	Masa Kepengurusan	Lama di Pesantren
	1	1	1	1	1	1	0	0

Figure 4 Most influential attributes

The result of combining Forward Selection with Naïve Bayes is accuracy: 91.20% +/- 3.85% (micro average: 91.21%) as seen in Figure 7

accuracy: 91.20% +/- 3.85% (micro average: 91.21%)

	true Tidak	true Ya	class precision
pred. Tidak	232	18	92.80%
pred. Ya	14	100	87.72%
class recall	94.31%	84.75%	

Figure 5 Results of the Naïve Bayes and Forward Selection Algorithms

After modeling with the Naïve Bayes and Forward Selection models, evaluation will be carried out with a confusion matrix which will produce a matrix consisting of true positive or positive tuples and true negative or negative tuples (Yulhendri et al., 2023). Then enter the test data provided in the confusion matrix so that results are obtained as in the following image:

accuracy: 91.20% +/- 3.85% (micro average: 91.21%)

	true Tidak	true Ya	class precision
pred. Tidak	232	18	92.80%
pred. Ya	14	100	87.72%
class recall	94.31%	84.75%	

Figure 6 Accuracy Values of the Naïve Bayes and Forward Selection Algorithms

precision: 88.63% +/- 9.46% (micro average: 87.72%) (positive class: Ya)

	true Tidak	true Ya	class precision
pred. Tidak	232	18	92.80%
pred. Ya	14	100	87.72%
class recall	94.31%	84.75%	

Figure 7 Precision Values of the Naïve Bayes and Forward Selection Algorithms

recall: 84.65% +/- 10.68% (micro average: 84.75%) (positive class: Ya)

	true Tidak	true Ya	class precision
pred. Tidak	232	18	92.80%
pred. Ya	14	100	87.72%
class recall	94.31%	84.75%	

Gambar 8 Nilai Recall dari Algoritma Naïve Bayes dan Forward Selection

The results of processing the confusion matrix evaluation of the Naïve Bayes algorithm and Forward Selection using the RapidMiner tool obtained results with accuracy values: 91.20% +/- 3.85% (micro average: 91.21%), precision: 88.63% +/- 9.26% (average micro: 87.72 %) (positive class: Yes), and recall: 84.65% +/- 10.68% (micro average: 84.75%) (positive class: Yes). Of the 364 data, there are 232 True Positive (TP) data, 18 False Negative (FN) data, 14 False Positive (FP) data, 100 True Negative (TP) data. Based on this data, evaluation using a confusion matrix can obtain accuracy, sensitivity, specificity, PPV and NPV values. The processed data is listed in the following table:

	Value
accuracy	0.912087912
sensitivity	0.928

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

specificity	0.877192982
PPV	0.943089431
NPV	0.847457627

Table 3 Accuracy, Sensitivity, Specificity, PPV, NPV and Forward Selection

C4.5 Algorithm with Forward Selection

The calculation results from the C4.5 algorithm will be added with the Forward Selection operator which is expected to get better accuracy results (Saputra et al., 2023), the steps are the same as in Figure 5 Namely, in modeling C4.5 and forward selection, we also use the help of the Rapidminer tool. However, there are differences in attributes that are influenced by the C4.5 forward selection model. These influential attributes can be seen as follows:

Attribute Weight	BTQ	Fiqh	Akhlaq	Tauhid	Aktif Pengabdian	Wali Asuh	Masa Kepengurusan	Lama di Pesantren
	1	1	1	1	0	0	1	0

Figure 9 The most influential attributes in the C4.5 and Forward Selection models

The result of combining Forward Selection with the C4.5 Algorithm is accuracy: 91.01% +/- 3.27 (micro average: 90.99%) as seen in the following image:

accuracy: 91.01% +/- 3.27% (micro average: 90.99%)

	true Tidak	true Ya	class precision
pred. Tidak	293	26	91.85%
pred. Ya	15	121	88.97%
class recall	95.13%	82.31%	

Figure 10 Results of the C4.5 Algorithm and Forward Selection

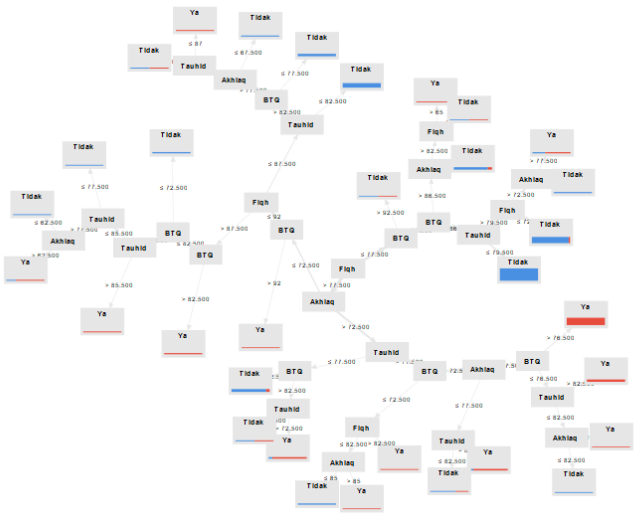

Figure 12 Description of the Role of the C4.5 Algorithm and Forward Selection

A decision tree was also found and an overview of the role of the C4.5 Algorithm using Forward Selection can be seen as follows:

Based on the resulting rules, 34 rules were formed from the C4.5 Algorithm decision tree and the forward selection amounted to 19 rules not being nominated and 15 rules being nominated in the prediction of regional administrators who were nominated for Best Management (Imaduddin et al., 2023). Nurul Jadid Islamic Boarding School. After modeling the C4.5 and Forward Selection algorithms, the Confusion Matrix Model will be evaluated which will produce a matrix consisting of true positive or positive tuples and true negative or negative tuples. Then enter the test data provided in the confusion matrix so that results are obtained as in the following image:

* Corresponding author

accuracy: 91.01% +/- 3.27% (micro average: 90.99%)

	true Tidak	true Ya	class precision
pred. Tidak	293	26	91.85%
pred. Ya	15	121	88.97%
class recall	95.13%	82.31%	

Figure 13 Accuracy Values of the C4.5 Algorithm and Forward Selection

precision: 89.95% +/- 7.09% (micro average: 88.97%) (positive class: Ya)

	true Tidak	true Ya	class precision
pred. Tidak	293	26	91.85%
pred. Ya	15	121	88.97%
class recall	95.13%	82.31%	

Figure 14 Pecision Values from the C4.5 Algorithm and Forward Selection

recall: 82.49% +/- 11.77% (micro average: 82.31%) (positive class: Ya)

	true Tidak	true Ya	class precision
pred. Tidak	293	26	91.85%
pred. Ya	15	121	88.97%
class recall	95.13%	82.31%	

Figure 15 Recall value from the C4.5 and Forward Selection algorithms

Based on the results above, in processing the evaluation of the fusion matrix of the C4.5 Algorithm and Forward Selection using the RapidMiner tool, the results obtained were accuracy values: 91.01% +/- 3.27% (micro average: 90.99%), precision: 89.95% +/ - 7.09% (micro average: 88.97%) (positive class: YES), recall: 82.49% +/- (micro average: 82.31%) (positive class: YES). Of the 455 data, there are 293 true positive (TP) data, 26 false negative (FN) data, 15 false positive (FP) data, 121 true negative (TP) data. Based on this data, evaluation using a confusion matrix can obtain accuracy, sensitivity, specificity, PPV and NPV values. The processed data is listed in the following table:

	Value
accuracy	0.90989011
sensitivity	0.918495298
specificity	0.889705882
PPV	0.951298701
NPV	0.823129252

Table 3 Accuracy, sensitivity, specificity, PPV and NPV values

DISCUSSIONS

Naïve Bayes algorithm

From the test results it can be concluded that there are 6 attributes that greatly influence the test, namely BTQ, Fiqh, Morals, Tauhid, Guardian and Active Service (Saleh et al., 2023). Comparison of the results of testing the Naïve Bayes Algorithm model without attribute selection with the Naïve Bayes Algorithm with Attribute Selection using Forward Selection is presented in the following table:

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 4 Naïve Bayes and Naïve Bayes Testing using Forward Selection

	Naïve Bayes	Naïve Bayes dan Forward Selection
Accuracy	0.876373626	0.912087912
Sensitivity	0.906882591	0.928
Specivicity	0.811965812	0.877192982
PPV	0.910569106	0.943089431
NPV	0.805084746	0.847457627

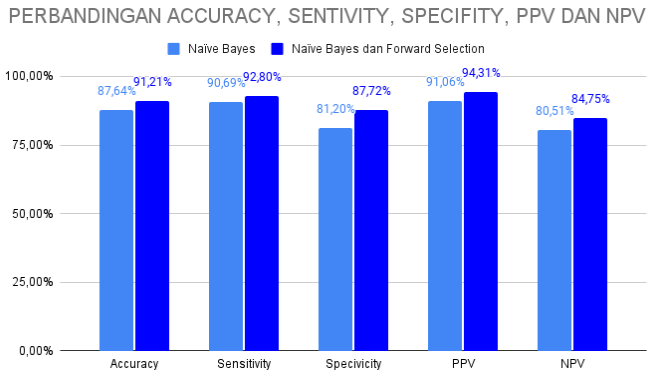

Figure 16 Comparison of Naïve Bayes and Naïve Bayes with Forward Selection

Based on the results of testing with the Evaluation Confusion Matrix, it is proven that testing carried out using the Naïve Bayes algorithm optimization with Forward Selection has a higher accuracy score compared to using the Naïve Bayes algorithm without the selection feature. The accuracy scores produced by the Naïve Bayes Algorithm model are accuracy: 87.64%, sensitivity: 90.69%, specificity: 81.20%, PPV: 91.06%, NPV: 80.51%. and the accuracy scores of the Naïve Bayes model and the Forward Selection method have accuracy values: 91.21%, sensitivity: 92.80%, specificity: 87.72%, PPV: 94.31%, NPV: 84.75%. From these scores, it can be seen that the difference in accuracy is 3.57%, sensitivity is 2.11%, specificity is 6.52%, PPV is 3.25%, and NPV is 4.25%.

Algorithm C4.5

From the test results it can be concluded that there are 5 attributes that influence the test, namely BTQ, Fiqh, Morals, Tauhid, Period of Management. A comparison of the test results of the C4.5 algorithm model without attribute selection and the C4.5 algorithm model with attribute selection using forward selection is presented in the following table:

Table 5 Testing C4.5 and C4.5 using Forward Selection

	Algotritma C4.5	Algoritma C4.5 dan Forward Selection
Accuracy	0.901098901	0.90989011
Sensitivity	0.920127796	0.918495298
Specificity	0.85915493	0.889705882
PPV	0.935064935	0.951298701
NPV	0.829931973	0.823129252

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13498
e-ISSN : 2541-2019
p-ISSN : 2541-044X

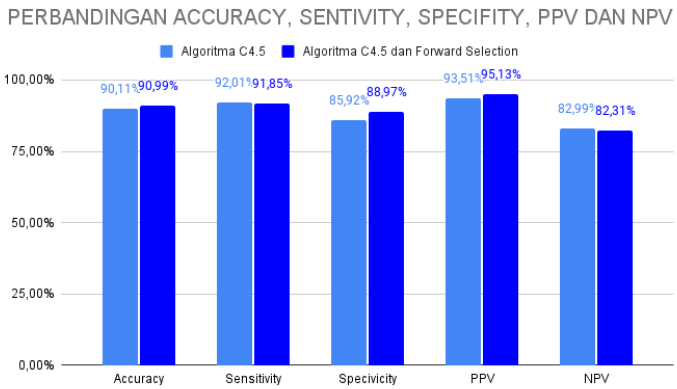

Figure 17 comparison of C4.5 and C4.5 with Forward Selection

Based on test results using the Confusion Matrix evaluation, it is proven that tests carried out using the C4.5 algorithm optimization with Forward Selection have a slightly higher accuracy score than using the C4.5 algorithm alone. The accuracy scores produced by the C4.5 algorithm model are accuracy: 90.11%, sensitivity: 92.01%, specificity: 85.92%, PPV: 93.51%, and NPV: 82.99%. and the accuracy score of the C4.5 and Forward Selection models has accuracy: 90.99%, sensitivity: 91.85%, specificity: 88.97%, PPV: 95.13%, and NPV: 82.31%. From these scores, we can see the difference in accuracy: 0.88%, sensitivity: -0.16%, specificity: 3.06%, PPV: 1.62%, and NPV: -0.68%. From the results obtained, it can be seen that the Naïve Bayes algorithm and the C4.5 algorithm itself have a lower level of accuracy compared to the Naïve Bayes algorithm and the C4.5 algorithm with feature selection using Forward Selection. As will be seen from the image below:

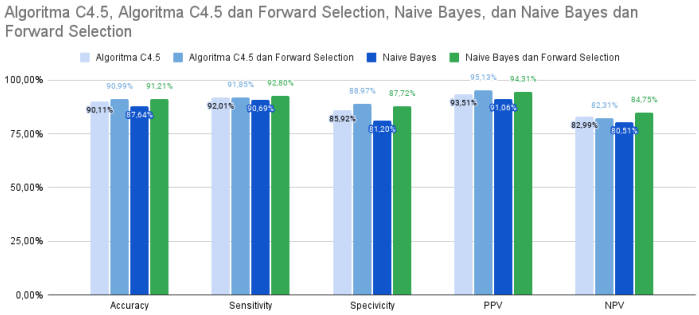

Figure 20 Comparison of the two models using feature selection and not using feature selection

From the results obtained, it can be seen that the Naïve Bayes algorithm and the C4.5 algorithm itself have a lower level of accuracy compared to the Naïve Bayes algorithm and the C4.5 algorithm with feature selection using Forward Selection. Because it is based on a calculation method: Naïve Bayes is a calcification algorithm based on Bayes' torema, which is a very simple and fast method for classifying data. So Naïve Bayes assumes that all features are independent of each other, and this is often a naive assumption. This algorithm is very suitable for modeling text classification, such as sentiment analysis or document classification (Hasibuan & Mahdiana, 2023).

The C4.5 algorithm is a decision tree algorithm used to classify data which is also the same as Naïve Bayes (Yenila et al., 2022). However, this algorithm focuses on selecting the best attributes at each level of the tree, so that it will maximize the information obtained from these attributes. The C4.5 algorithm is suitable for decision making, especially in situations where the interpretation of decision rules is critical. Like the modeling carried out in 4.5 to find the root of a decision tree, the C4.5 algorithm must calculate entropy, gain info, split info so that the gain ratio is found for each attribute. The Naïve Bayes and C4.5 algorithms with selection features are often used to determine student achievement or the level of diabetes sufferers. However, in this research, the Naïve Bayes and C4.5 algorithms were used with selection features to measure the value of Islamic boarding school administrators, especially the Nurul Islamic Boarding School, in determining exemplary administrators considering that in the current era the development of Islamic boarding schools is very limited. very fast.

CONCLUSION

Comparison between the Naïve Bayes and C4.5 algorithms with selection features. Forward Selection with the Naïve Bayes and C4.5 algorithms without selection features, can be applied to predict the best management nominations. Both models produce quite good accuracy. In this research, the Forward Selection algorithm is used to optimize existing attributes so that the attributes obtained are relevant and influential attributes. From the number

* Corresponding author

of attributes before optimization, there were 17 attributes which were then selected to become 8 attributes. After optimizing feature selection, the number of attributes becomes 6 for naïve Bayes and 5 for C4.5. There are 4 relevant attributes in both algorithm models, namely *BTQ, Fiqh, Tauhid and Akhlaq*. These four attributes are always used whether feature selection is used or not. So Forward Selection which selects the most relevant features can provide a more accurate and efficient model for Naïve Bayes, and for the C4.5 Algorithm it can help produce decision trees that are simpler and easier to understand. So what is certain is that both algorithms benefit from feature selection methods to improve performance and interpretability.

ACKNOWLEDGMENT

The author would like to thank us for providing the data we needed during the research, especially the Nurul Jadid Islamic Boarding School, as well as the support of facilities from Dian Nuswantoro University, without which this research would not have been possible.

REFERENCES

Angga. (2020). Reveal a School Branding Model Based on Pesantren. *JIEM (Journal of Islamic Education Management)*, *4*(2), 143–155. https://doi.org/10.24235/jiem.v4i2.7270

Castro, V. de S., & Oliveira, S. R. B. (2023). education sciences Diversity in Software Design and Construction Teaching : *Education Sciences*, *13*(303), 1–22. https://doi.org/https://doi.org/10.3390/educsci13030303

Dutra, A. C. C., & Dawson, J. A. (2023). Computational Design of Antiperovskite Solid Electrolytes. *Journal of Physical Chemistry C*, *127*(37), 18256–18270. https://doi.org/10.1021/acs.jpcc.3c04953

Folorunso, S. O., Awotunde, J. B., Adigun, A. A., Prasad, L. V. N., & Lalitha, V. L. (2023). A hybrid model for post-treatment mortality rate classification of patients with breast cancer. *Healthcare Analytics*, *4*(June), 100254. https://doi.org/10.1016/j.health.2023.100254

Hamdani, R., Sriani, & Darta, A. (2024). ANALISIS TINGKAT KEPUASAN PASIEN DI KLINIK PRATAMA. *Journal of Science and Social Research*, *4307*(1), 273–280.

Hasibuan, T. H., & Mahdiana, D. (2023). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta. *Skanika*, *6*(1), 61–74. https://doi.org/10.36080/skanika.v6i1.2976

Imaduddin, H., Hermansyah, B. A., & Alfajri, M. M. (2023). Klasifikasi Kematian Akibat Gagal Jantung Menggunakan Algoritma Logistic Regression Berbasis Forward Selection. *J I M P - Jurnal Informatika Merdeka Pasuruan*, *7*(3), 96. https://doi.org/10.51213/jimp.v7i3.565

Muhammad, R., & Zakaria, H. (2024). Penerapan Algoritma K-Means dalam Penentuan Siswa Bermasalah Berdasarkan Running Record (Studi Kasus : SMK Averus Jakarta). *JRIIN: Jurnal Riset Informatika Dan Inovasi*, *1*(7), 720–729.

Nurhalisha, T., Pranoto, W. J., Saputra, H., & Latipah, A. J. (2023). The application of particle swarm optimization (PSO) to improve the accuracy of the naive bayes algorithm in predicting floods in the city of Samarinda. *Journal of Intelligent Decision Support System*, *6*(3), 138–146.

Rahmayani, F. (2023). Penerapan Algoritma C4.5 Dengan Feature Forward Selection Untuk Analisis Capaian Indikator Kinerja Utama Berdasarkan Tracer Study (Studi Kasus: Fasilkom Unsika). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(4), 2732–2738.

Sahlan, M. (2023). Profiling of Religious Personality through Character Management Representations. *Journal Of Research in Educatiun Management*, *2*(1), 37–48.

Saleh, A., Dharshinni, N., Perangin-Angin, D., Azmi, F., & Sarif, M. I. (2023). Implementation of Recommendation Systems in Determining Learning Strategies Using the Naïve Bayes Classifier Algorithm. *Sinkron*, *8*(1), 256–267. https://doi.org/10.33395/sinkron.v8i1.11954

Saputra, M., Sinaga, I. P. L., & Sinaga, W. (2023). Analisis penerapan algoritma c4.5 untuk mengidentifikasi kelancaran kredit nasabah pada cu. dosnitahi pinang sori. *Jurnal TEKINKOM*, *6*(2), 737–743. https://doi.org/10.37600/tekinkom.v6i2.943

Suhendar, A. H., Rohmawati, A. A., & Prasetyowati, S. S. (2024). Performance of CART Time-Based Feature Expansion in Dengue Classification Index Rate. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *9*(1), 1–9.

Yenila, F., Wahyuni, S., Rianti, E., Marfalino, H., & Gusmita, D. (2022). Sistem Pakar Deteksi Hemangioma pada Batita menggunakan Metode Hybrid. *Jurnal Informasi Dan Teknologi*, *4*(4), 265–270. https://doi.org/10.37034/jidt.v4i4.250

Yulhendri, Y., Malabay, M., & Kartini, K. (2023). Correlated Naïve Bayes Algorithm to Determine Healing Rate of Hepatitis Patients. *International Journal of Science, Technology & Management*, *4*(2), 401–410. https://doi.org/10.46729/ijstm.v4i2.776

Zakaria, W., Yuniati, U., & Puspitasari, E. E. (2023). Strategi Membangun Brand Image Dalam Meningkatkan Daya Saing Lembaga Pendidikan. *Indonesian Journal of Digital Public Relations (IJDPR)*, *1*(2), 64. https://doi.org/10.25124/ijdpr.v1i2.5545

* Corresponding author