

APPLICATION OF NAÏVE BAYES ALGORITHM IN DETERMINING STUDENT CONCENTRATION IN MATHEMATICS LEARNING PROCESS

Hadi Kusmanto¹, Nina Siti Munawaroh², Herlinda Nur'afwa Sofhya³
^{1,2,3} UIN Siber Syekh Nurjati Cirebon, Indonesia

Article Information

Article History:

Received December 20, 2024
Revised December 31, 2024
Published December 31,
2024

DOI:

<https://doi.org/10.58557/eduinsights.v2i2.100>

Keyword:

Student Concentration
Data Mining
Naïve Bayes
RapidMiner

ABSTRACT

This research on student concentration in the math learning process by applying the Naïve Bayes algorithm which aims to (1) determine the Naïve Bayes Algorithm in determining student concentration in the math learning process, (2) determine the chances of student concentration in the math learning process. This research uses the Naïve Bayes classification data mining method. This research was conducted at SMPN 4 Cirebon City Indonesia with a total sample of 101 students. The instrument used was a student questionnaire. The results showed: data mining has 4 stages consisting of datasets, data cleaning, data grouping, namely data grouping, and Naïve Bayes algorithm modeling. A model was obtained to predict the class by multiplying the probability of each criterion and then multiplying it by the probability of concentration or the probability of less concentration. There are 0.7 concentration classes and 0.3 less concentration classes. Based on the RapidMiner application, there is an Accuracy value of 94.00%. Class precision on concentration prediction has a value of 94.44%, while the prediction of less concentration has a value of 92.85%. Class recall on true concentration has a value of 97.14%, while on true lack of concentration has a value of 85.57%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Correspondence Author:

Hadi Kusmanto
UIN Siber Syekh Nurjati Cirebon, Indonesia
Email: hadikusmanto@syekhnurjati.ac.id

1. INTRODUCTION

Basic education is one of the formal education units, as stated in Permendikbud No 6 of 2019 Chapter 1 Article 1 Paragraph 2, "Basic education is a level in the formal education pathway that underlies the secondary education level, which is organized in educational units in the form of Primary Schools (SD) and Madrasah Ibtidaiyah (MI) or other equivalent forms and becomes an integral part of the continuation of education in educational units in the form of Junior High Schools (SMP) and Madrasah Tsanawiyah (MTs) or other equivalent forms".

Improving the quality of education depends on the implementation of teaching and learning activities in junior high schools, both in public and private educational institutions. The interaction between teachers and students in Teaching and Learning Activities (KBM) is of key importance. When the KBM takes place, the teacher acts as a moderator who delivers the material, while the students act as an audience who receives the material. The role of students in the learning process begins with the level of seriousness and concentration, because seriousness in understanding the material is the key to a smooth learning process. Good concentration is closely related to good understanding, while lack of concentration can have an impact on poor understanding. (Yusalis, 2020).

The learning process is an exploration of knowledge and experience that brings changes in a person's behavior and reaction capabilities, which are relatively permanent due to the interaction of individuals with their environment. There are various factors that have an influence on the learning process (Nurfadilah & Hakim, 2019). However, the fact shows that many students do not feel interested in math subjects. They think that math tends to be monotonous and scary because it is identified with numbers and formulas. Negative attitudes towards mathematics arise due to misperceptions associated with this subject. One of the factors is the assumption that mastering math requires a special talent that is not possessed by every individual.

Learning concentration is one of the things that is difficult for students to overcome, because in the learning process many things can affect student concentration. It takes a long time in order to help students concentrate in learning, student concentration in learning can increase when students get guidance and attention from teachers and teacher patience in dealing with students. (Riinawati, 2021). Concentration in the learning process can be reflected in several aspects, such as focus on the subject matter, the level of attention students pay during learning, their ability to answer questions, and the courage to ask questions.

Naïve Bayes Classifier (NBC) is a classification technique that belongs to the category of statistical classifiers, where the probability of class membership can be predicted (Annur, 2018). Naïve Bayes efficiently performs classification by focusing on estimating the correct probabilities, especially when facing challenged assumptions.

The problem that often occurs in SMP Negeri 4 Kota Cirebon is the lack of concentration of students so that when students get questions from the teacher they are often unsure when answering questions and the lack of courage of students when expressing their opinions and the shyness of students which causes them not to ask the teacher when there is something they don't understand about math lessons. This is similar to the problems studied by (Az Zahra & Solikhun, 2021) related to the lack of student concentration in the teaching and learning process and math teachers also often complain that the average learning process is often disrupted due to lack of student concentration. To overcome this problem, SMP Negeri 4 Cirebon City needs to get the right information about the level of student concentration. This information will be a guide for math teachers to better understand the situation of students during math learning, helping them to assess whether students really understand the material being taught or not. Therefore, accurate data on student concentration is needed in order to improve the effectiveness of mathematics teaching in this school.

2. METHOD

In this study, it uses a quantitative approach by applying the Naïve Bayes algorithm and using a questionnaire given to students of SMP Negeri 4 Kota Cirebon. The purpose of this research is none other than to determine student concentration data so that it can be used as a reference for teachers in knowing the condition of students. The following is a research design in determining student concentration in the math learning process using the Naïve Bayes algorithm:

- 1) Problem analysis
Analyze the problems involved in determining student concentration in the mathematics learning process and determine what indicators to use.
- 2) Studying the literature
This research is certainly based on references used to obtain information in the research.
- 3) Instrument testing

The data collected is in the form of a questionnaire given to students of class VII-I SMP Negeri 4 Kota Cirebon for instrument testing.

- 4) Testing data using SPSS
Data testing using SPSS software was carried out to assess the validity of the data used.
- 5) Data collection
The data collected is in the form of a questionnaire given to students in grades VII, VIII, and IX of SMP Negeri 4 Kota Cirebon.
- 6) Processing data with algorithms
Perform data processing using Naïve Bayes algorithm calculation
- 7) Testing the data
Data testing using the Rapid Miner application as a system to determine whether students have concentrated or not during the math learning process.
After collecting data, the next stage is data analysis techniques, as for the stages, namely:
 - 1) Preliminary analysis, at this stage will provide an overview of the research starting from the situation of student concentration problems found in SMP Negeri 4 Kota Cirebon.
 - 2) Data analysis, at this stage the data is tested using SPSS to determine whether the data is valid or not.
 - 3) Researchers started by taking data at SMP Negeri 4 Cirebon City.
 - 4) After the researcher gets the data, data processing is immediately carried out using *excel* software, after the data is well organized, *sort & filter* is carried out which aims to divide the attributes and values so that it can be done in data grouping.
 - 5) Application of the *Naïve Bayes* algorithm, this activity is an analysis of all the data that has been obtained. The data processed using the *Naïve Bayes* method is quantitative data. Quantitative data is data in the form of numbers. The following are the steps of the *Naïve Bayes* algorithm.
 - a. Calculate $P(C_i)$ for each class
 - b. Calculate $P(X|C_i)$ for each criterion and each class
 - c. Display classification results
 - d. Finish
 - 6) Data testing using the RapidMiner application

Naïve Bayes

Naïve Bayes Classifier is one of the problem solving algorithms included in the classification method in data mining. Naïve Bayes Classifier adopts statistics by using probability theory to solve a supervised learning case, meaning that in the set there is a label, class or target as a reference or teacher. (Nofriansyah & Nurcahyo, 2019).

Naïve Bayes Classifier works very well compared to other classifier models. This is proven in the journal Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone. "Naïve Bayes vs. decision trees vs. neural networks in the classification of training web pages, 2009", says that "Naïve Bayes Classifier has a better accuracy rate than other classifier models." (Jollyta et al., 2020).

Bayes prediction refers to the application of Bayes theorem using this general formula:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

$P(H|X)$ is the probability that hypothesis H is true for evidence X, or in other words, $P(H|X)$ is the posterior probability of H under condition X, $P(X|H)$ is the probability of evidence X is true for hypothesis H or the posterior probability of X under condition H, $P(H)$ is the prior probability of hypothesis H and $P(X)$ is the prior evidence for confidence in proving hypothesis X. In data mining, X is a tuple or data item and H represents the hypothesis or assumption that tuple X belongs to class C. $P(H|X)$ can be calculated as the probability that hypothesis H is true in the classification work of *tuple* X, or $P(H|X)$, is the probability that *tuple* X belongs to Class C while $P(H)$ is the initial probability of H for each *tuple* independent of attribute values and $P(X)$ is the initial value of the origin probability of *tuple* X.

3. RESULT AND DISCUSSION

This section explains how the stages and process of Naive Bayes in determining student concentration on the math learning process. Here is the Naive Bayes algorithm training data:

Table 1. Naive Bayes Algorithm Training Data

R	X1	X2	X3	X4	..	X10	Results
R1	R	S	S	TS	..	SS	K
R2	S	S	R	S	..	R	K
R3	S	SS	S	R	..	TS	KK
R4	S	R	SS	SS	..	R	K
R5	S	TS	R	TS	..	TS	KK
..
R100	S	S	S	R	..	R	K

After the data has been determined, the next step is to calculate the amount of concentration and lack of concentration based on Table IV.9. From the 100 training data used, it is known that the concentration class is 70 data, and the less concentration class is 30 data. Calculation of the probability of the possible concentration class in determining student concentration on the math learning process, namely;

$$P(\text{Konsentrasi}) = \frac{70}{100} = 0,7$$

While the calculation of the probability of lack of concentration is:

$$P(\text{Kurang Konsentrasi}) = \frac{30}{100} = 0,3$$

After the probability of each class is known, then the author calculates each probability of each criterion used. The criteria used by the author are the existence of acceptance or attention to the subject matter, interest in the subjects studied, the presence of appropriate limb movements in accordance with the teacher's instructions, the readiness of the knowledge obtained immediately appears when readiness is needed, responds to the material taught, is able to apply the knowledge obtained, is able to express ideas / opinions, is able to analyze the knowledge obtained, is not bored with the learning process that is passed, physical health.

In determining the probability of each criterion, the author calculates the parts contained in each criterion, in this study the parts contained in each criterion use a 5 linkert scale, namely SS, S, R, TS, and STS. So that in determining the probability of each criterion is done by calculating the amount of concentration and lack of concentration on the linkert scale used. So that the calculation of the probability of each criterion can be seen in the following tables.

To calculate the probability of the possibility of acceptance or attention to the subject matter, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 2. Probability of Criteria1

Criteria1 (X1)	Probability	
	Concentration	Lack of Concentration
SS	0,4857	0,1333
S	0,4143	0,5667
R	0,1	0,1
TS	0	0,1
STS	0	0,1
Total	1	1

The probability on the criteria for acceptance or attention to the subject matter, namely in the concentration category with the SS scale has a probability of 0.4857, S has a probability of 0.4143, R has a probability of 0.1 and TS and STS have a probability of 0. So that the total probability of concentration is 1. While in the category of lack of concentration with the SS scale has a

probability of 0.1333, S has a probability of 0.5667, R, TS and STS have a probability of 0.1. So that the number of probabilities of less concentration is 1.

To calculate the probability of the criterion of being interested in the subjects studied, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 3. Probability of Criteria2

Criterion2 (X2)	Probability	
	Concentration	Lack of Concentration
SS	0,4143	0,1
S	0,4857	0,3667
R	0,0714	0,2333
TS	0,0286	0,1667
STS	0	0,1333
Total	1	1

The probability on the criteria for being interested in the subjects studied, namely in the concentration category with the SS scale has a probability of 0.4143, S has a probability of 0.4857, R has a probability of 0.0714, TS has a probability of 0.0286 and STS has a probability of 0. So that the total probability of concentration is 1. While in the category of lack of concentration with the SS scale has a probability of 0.1, S has a probability of 0.3667, R has a probability of 0.2333, TS has a probability of 0.1667 and STS has a probability of 0.1333. So that the total probability of lack of concentration is 1.

To calculate the probability of the criterion of having the right limb movement according to the teacher's instructions, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 4. Probability of Criteria3

Criterion3 (X3)	Probability	
	Concentration	Lack of Concentration
SS	0,1	0,0667
S	0,2857	0,1333
R	0,2714	0,3
TS	0,1	0,2
STS	0,2429	0,3
Total	1	1

The probability on the criterion of the existence of appropriate limb movements in accordance with the teacher's instructions, namely in the concentration category with the SS scale has a probability of 0.1, S has a probability of 0.2857, R has a probability of 0.2714, TS has a probability of 0.1 and STS has a probability of 0.2429. So that the total probability of concentration is 1. Whereas in the less concentration category with the SS scale has a probability of 0.0667, S has a probability of 0.1333, R has a probability of 0.3, TS has a probability of 0.2 and STS has a probability of 0.3. So that the total probability of less concentration is 1.

To calculate the probability of the possibility of the readiness criteria, the knowledge obtained immediately appears when readiness is needed, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows.

Table 5. Criterion Probabilities 4

Criterion4 (X4)	Probability	
	Concentration	Lack of Concentration
SS	0,0714	0,0333
S	0,1857	0,1
R	0,4143	0,4333
TS	0,2857	0,2667
STS	0,0429	0,1667
Total	1	1

The probability on the readiness criteria for knowledge obtained immediately appears when readiness is needed, namely in the concentration category with a scale SS has a probability of 0.0714, S has a probability of 0.1857, R has a probability of 0.4143, TS has a probability of 0.2857 and STS has a probability of 0.0429. So that the total probability of concentration is 1. While in the category of lack of concentration with the SS scale has a probability of 0.0333, S has a probability of 0.1, R has a probability of 0.4333, TS has a probability of 0.2667 and STS has a probability of 0.1667. So that the total probability of lack of concentration is 1.

To calculate the probability of the possibility of the criteria responding to the material taught, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 6. Probability of Criteria5

Criterion5 (X5)	Probability	
	Concentration	Lack of Concentration
SS	0,0286	0,0667
S	0,2429	0,1333
R	0,3571	0,3333
TS	0,2857	0,3
STS	0,0857	0,1667
Total	1	1

The probability on the criteria for responding to the material taught, namely in the concentration category with the SS scale has a probability of 0.0286, S has a probability of 0.2429, R has a probability of 0.3571 TS has a probability of 0.2857 and STS has a probability of 0.0857. So that the total probability of concentration is 1. While in the category of lack of concentration with the SS scale has a probability of 0.0667, S has a probability of 0.1333, R has a probability of 0.3333, TS has a probability of 0.3 and STS has a probability of 0.1667. So that the total probability of lack of concentration is 1.

To calculate the probability of the possibility of the criterion of being able to apply the knowledge gained, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 7. Criterion Probabilities 6

Criterion6 (X6)	Probability	
	Concentration	Lack of Concentration
SS	0,1857	0,0333
S	0,3714	0,2333
R	0,4143	0,3
TS	0,0143	0,2333
STS	0,0143	0,2
Total	1	1

The probability of being able to apply the knowledge gained, namely in the concentration category with the SS scale has a probability of 0.1857, S has a probability of 0.3714, R has a probability of 0.4143, TS has a probability of 0.0143 and STS has a probability of 0.0143. So that the total probability of concentration is 1. Whereas in the category of less concentration with the SS scale has a probability of 0.0333, S has a probability of 0.2333, R has a probability of 0.3, TS has a probability of 0.2333 and STS has a probability of 0.2. So that the total probability of lack of concentration is 1.

To calculate the probability of the possibility of the criterion of being able to express ideas/opinions, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the following table:

Table 8. Criterion Probabilities 7

Criterion7 (X7)	Probability	
	Concentration	Lack of Concentration
SS	0,1571	0
S	0,4571	0,2
R	0,2714	0,5
TS	0,1	0,2667
STS	0,0143	0,0333
Total	1	1

Probability on the criteria for being able to express ideas / opinions, namely in the concentration category with a scale SS has a probability of 0.1571, S has a probability of 0.4571, R has a probability of 0.2714 TS, has a probability of 0.1 and STS has a probability of 0.0143. So that the total probability of concentration is 1. While in the category of lack of concentration with the SS scale has a probability of 0, S has a probability of 0.2, R has a probability of 50%, TS has a probability of 0.2667 and STS has a probability of 0.0333. So that the number of probabilities of less concentration is 1.

To calculate the probability of the possibility of the criterion of being able to analyze the knowledge gained, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 9. Criterion Probabilities8

Criterion8 (X8)	Probability	
	Concentration	Lack of Concentration
SS	0,2714	0
S	0,5	0,2333
R	0,2	0,5667
TS	0,0286	0,0667
STS	0	0,1333
Total	1	1

The probability of being able to analyze the knowledge gained, namely in the concentration category with the SS scale has a probability of 0.2714, S has a probability of 0.5, R has a probability of 0.2 TS, has a probability of 0.0286 and STS has a probability of 0. So that the total probability of concentration is 1. Whereas in the category of lack of concentration with the SS scale has a probability of 0, S has a probability of 0.2333, R has a probability of 0.5667, TS has a probability of 0.0667 and STS has a probability of 0.1333. So that the total probability of lack of concentration is 1.

To calculate the probability of the possibility of not being bored with the learning process, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the table as follows:

Table 10. Criterion Probabilities9

Criterion9 (X9)	Probability	
	Concentration	Lack of Concentration
SS	0,3857	0,1667
S	0,2714	0,2333
R	0,2143	0,1667
TS	0,0714	0,2
STS	0,0571	0,2333
Total	1	1

The probability on the criterion of not getting bored with the learning process that is passed, namely in the concentration category with the SS scale has a probability of 0.3857, S has a probability of 0.2714, R has a probability of 0.2143 TS, has a probability of 0.0714 and STS has a probability of 0.0571. So that the total probability of concentration is 1. Whereas in the less concentration category with the SS scale has a probability of 0.1667, S has a probability of 0.2333, R has a probability of 0.1667, TS has a probability of 0.2 and STS has a probability of 0.2333. So that the total probability of lack of concentration is 1.

To calculate the probability of the physical health criteria, the excel software uses the formula =COUNTIFS/COUNTIF, so it can be seen in the following table:

Table 11. Criterion Probabilities 10

Criterion10 (X10)	Probability	
	Concentration	Lack of Concentration
SS	0,4857	0,1333
S	0,4143	0,5667
R	0,1	0,1
TS	0	0,1
STS	0	0,1
Total	1	1

Probability on physical health criteria, namely in the concentration category with a scale SS has a probability of 0.1143, S has a probability of 0.2857, R has a probability of 0.2143 TS, has a probability of 0.2429 and STS has a probability of 0.1429. So that the total probability of concentration is 1. Whereas in the less concentration category with the SS scale has a probability of 0.1667, S has a probability of 0.1, R has a probability of 0.1, TS has a probability of 0.4667 and STS has a probability of 0.1667. So that the total probability of lack of concentration is 1.

After each criterion probability is known, the next step is to make a memorization when there are respondents R101 to R105 by answering the statements from the questionnaire that has been given, then determining the classification value. The following is the testing data that will be used.

Table 12. Naive Bayes Algorithm Testing Data

R	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Results
R101	SS	R	SS	TS	R	SS	S	R	S	SS	?
R102	R	SS	R	STS	TS	TS	R	S	S	S	?
R103	S	TS	R	S	STS	R	S	R	STS	TS	?
R104	S	S	TS	TS	R	STS	TS	S	R	S	?
R105	R	R	STS	SS	S	TS	STS	SS	TS	STS	?

The formula used in determining the prediction class can be seen in the following equation.

$$P(X_2|C)P(X_3|C) \dots P(C|X_1, \dots, X_n) = P(X_1|C) = \prod_{i=1}^n P(X_i|C)$$

So to calculate the concentration value on respondent data R101 to R105 is as follows:

$$\begin{aligned}
 &P(R101|Concentration) \\
 &=P(Criterion1=SS|Concentration) \times P(Criterion2=R|Concentration) \times P(Criterion3=SS|Concentration) \\
 &\times P(Criterion4=TS|Concentration) \times P(Criterion5=R|Concentration) \times P(Criterion6=SS|Concentration) \\
 &\times P(Criterion7=S|Concentration) \times P(Criterion8=R|Concentration) \times P(Criterion9=S|Concentration) \times P \\
 &(Criterion10=SS|Concentration) \times P(Concentration). \\
 &=0,4857 \times 0,0714 \times 0,1 \times 0,2857 \times 0,3571 \times 0,1857 \times 0,4571 \times 0,2 \times 0,2714 \times 0,1143 \times 0,7 \\
 &=0,000000130428841
 \end{aligned}$$

$$\begin{aligned}
 &P(R101|Lack of Concentration) \\
 &=P(Criterion1=SS|Lack of Concentration) \times P(Criterion2=R|Lack of Concentration) \times P(Criterion3=SS|Lack of Concentration) \\
 &\times P(Criterion4=TS|Lack of Concentration) \times P(Criterion5=R|Lack of Concentration) \times P(Criterion6=SS|Lack of Concentration) \\
 &\times P(Criterion7=S|Lack of Concentration) \times P(Criterion8=R|Lack of Concentration) \times P(Criterion9=S|Lack of Concentration) \times P(Criterion10=SS|Lack of Concentration) \\
 &\times P(Lack of Concentration). \\
 &=0,1333 \times 0,2333 \times 0,0667 \times 0,2667 \times 0,3333 \times 0,0333 \times 0,2 \times 0,5667 \times 0,2333 \times 0,1667 \times 0,3 \\
 &=0,000000008119477
 \end{aligned}$$

Or in excel *software* using the formula =VLOOKUP(criteria1)*VLOOKUP(criteria2)*...*VLOOKUP(criteria10)*probability value (concentration/lack of concentration). After getting the results, then the author compares the results of the probability of concentration and lack of concentration by looking for the largest probability then that is the answer to the prediction class, or in excel *software* by =IF(...; "Concentration"; "Lack of Concentration"). So that it can be known that the student is included in the prediction class category of concentration or lack of concentration. Here are the comparison results.

Table 13. Prediction Class Classification Results

Respondents	Probability of Concentration	Probability of Lack of Concentration	Prediction Class
R101	0,000000130428841	0,000000008119477	K
R102	0,000000014515309	0,000000028576799	KK
R103	0,000000018819208	0,000000524715537	KK
R104	0,00000006290999	0,000000229920159	KK
R105	0,000000000011921	0	K

So from the comparison data, it can be seen that the *testing* data from respondent data R101 and R105 has a concentration prediction class classification. Meanwhile, the data from respondents R102, R103, and R104 have a prediction class classification of less concentration.

RapidMiner

This section describes the stages of using the RapidMiner application in classifying the *Naïve Bayes* method. The end of using this application will provide how much accuracy the *Naïve Bayes* method uses. The RapidMiner used is RapidMiner Studio Educational 10.3. The following is an image of the initial appearance of the RapidMiner application.

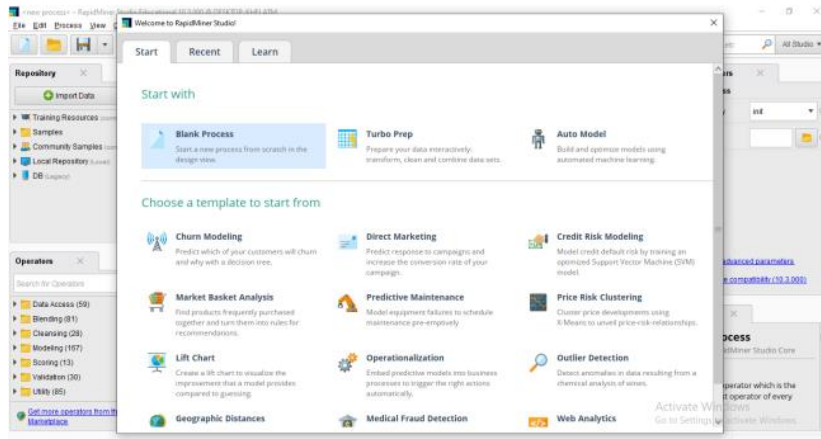


Image 1. RapidMiner Initial View

In Figure 1. is the initial display in the RapidMiner application. Since the data to be tested is new data, then select "Blank Process". The "Blank Process" menu contains the initial display if you are going to create a new process with new data that has never existed in this *tool*. Then a window will appear as shown below:

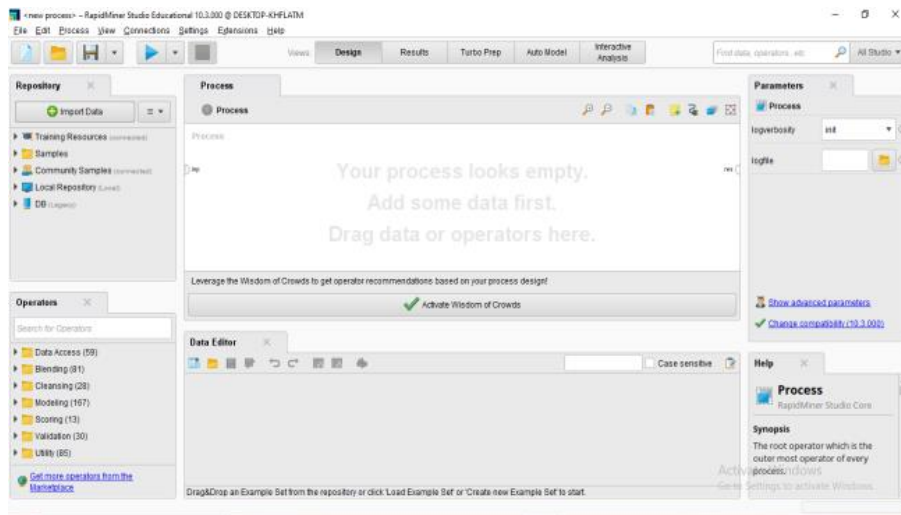


Image 2. RapidMiner Blank Process

In Figure 2. is a display of "Blank Process" in RapidMiner which is ready to enter new data that will be used in the form of excel data. How to enter the data is by selecting "Import Data" then the following display will appear.

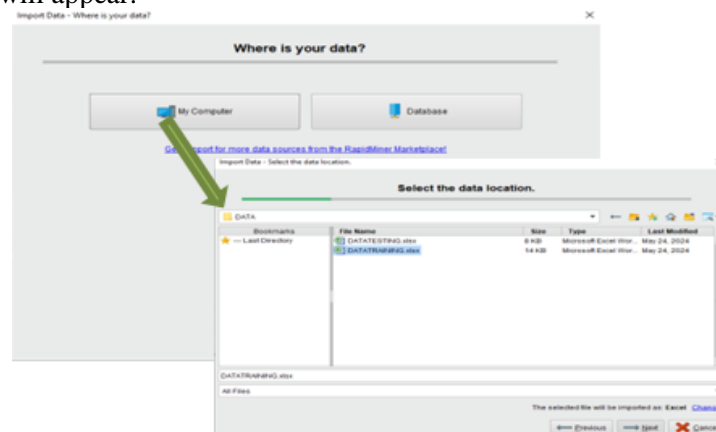


Image 3. Import Excel Data

In Figure 3. select "My Computer" to enter the excel data that has been saved on the computer, then select the location of the excel data that has been prepared. Then select "Next" then the following display will appear.

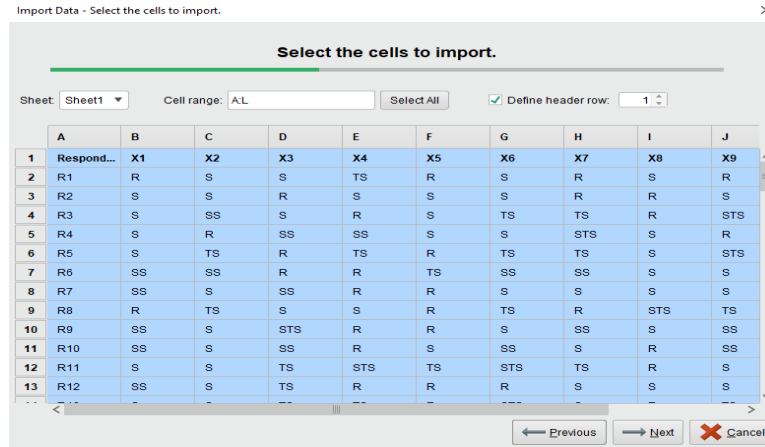


Image 4. Import Excel Data Stage1

In Figure 4. above shows the contents of the data to be processed. Select the sheet on the data and the data block to be processed, then select next and it will appear as shown below.

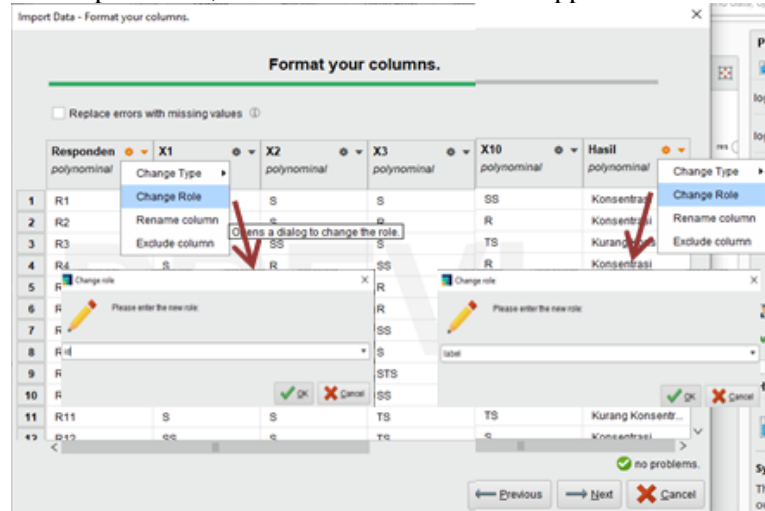


Image 5. Import Excel Data Stage2

In Figure 5. is stage 2 data import which will be carried out by selecting the data type where the respondent's *change role* is changed to the type "id" and the result is given the type "label". Then choose OK. Then the following image will appear.

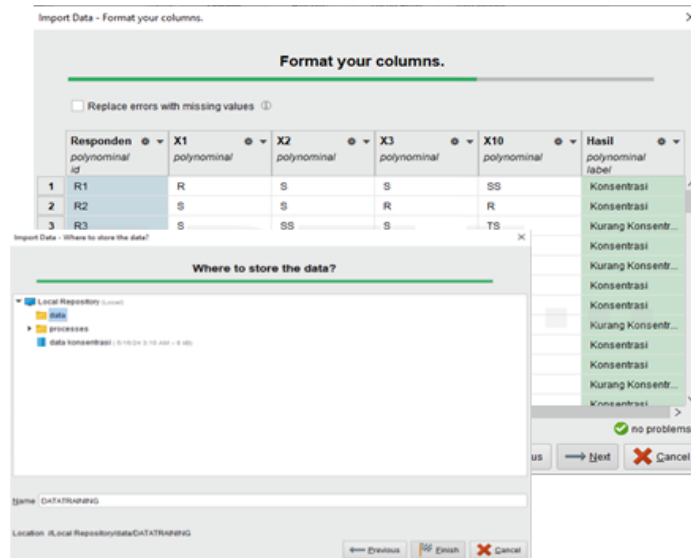


Image 6. Import Excel Data Stage3

In Figure 6. is the data import stage 3 in the results section change "polynomial" to "binominal" because there are only 2 labels available. Then select "Next", then select the data storage area to be processed then specify the name of the data to be processed. Then choose finish. The imported data will look like the following image.



Image 7. Imported Data

In Figure 7. is data that has been imported and the data is ready for processing. System processing at this stage will explain the stages of the process of using *Naïve Bayes* in Rapidminer. The following are the stages that can be seen in the picture below:



Image 8. Stage 1 of Naive Bayes Data Processing

The first stage can be seen in Figure 8. by *dragging* or clicking and holding the training data that has been stored in the repositories to the main *process* worksheet. Then find "Naïve Bayes" in the operator section that is already available *drag* it to the main *process* worksheet. Then find "Apply Model" in the operator section and *drag* it to the main *process* worksheet. Then find "Performance Classification" on the operator then *drag* it to the main *process* worksheet. Can be seen below:



Image 9. Stage 2 Naive Bayes Data Processing

After the required *tools* are available in the worksheet as shown in Figure 9. next to form a decision tree, the data needs to be connected with the operators that are already in the worksheet. Can be seen in the picture below.



Image 10. Stage 3 Naive Bayes Data Processing

The data that has been connected to the operators on the worksheet, then the next step is to see on the main sheet the process of the operators does not have an error mark. If there are still errors, it is possible that the input data is invalid or cannot be processed because there is still data that does not match. And if the data and operators have no error signs, then the data process can be continued to get the desired results. Then the data process ends at stage 3.

To get the results in the form of a decision tree, the next stage can be done by clicking the blue arrow located at the top of the *toolbar*. At this stage will display the final results and the last step in using this RapidMiner *tool*. The final result that will be displayed is in the form of "SimpleDistribution" which determines the number of values of concentration and less concentration data. Can be seen below:



Image 11. Simple Distribution values

Based on Figure 11. explains that the concentration class has a classification/probability value of 0.700 while the less concentration class gets a classification/probability value of 0.300. Then the accuracy of the *Naïve Bayes Classifier* Algorithm Model test results is shown below:

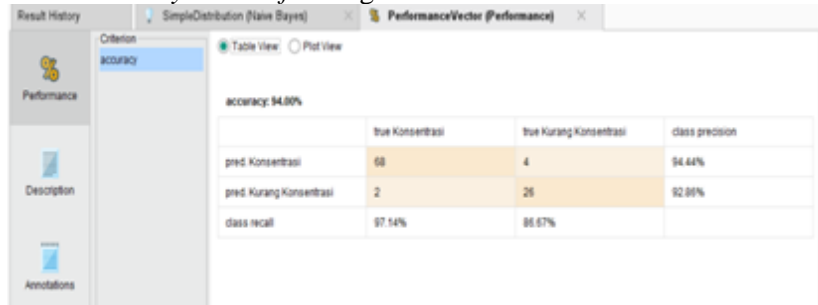


Image 12. Accuracy Value

In Figure 12. the *Accuracy* value is 94.00%. *Class precision* on concentration prediction has a value of 94.44%, while the prediction of less concentration has a value of 92.85%. *Class recall* on *true* concentration has a value of 97.14%, while on *true* less concentration has a value of 85.57%. With the number of predictions of concentration and true concentration is 68 respondents (*true positive*). The number of predictions of concentration and true lack of concentration is 4 respondents (*false positive*). The number of predictions of lack of concentration and true concentration is 2 respondents (*false negative*). The number of predictions of lack of concentration and the reality of true lack of concentration is 26 respondents (*true negative*).

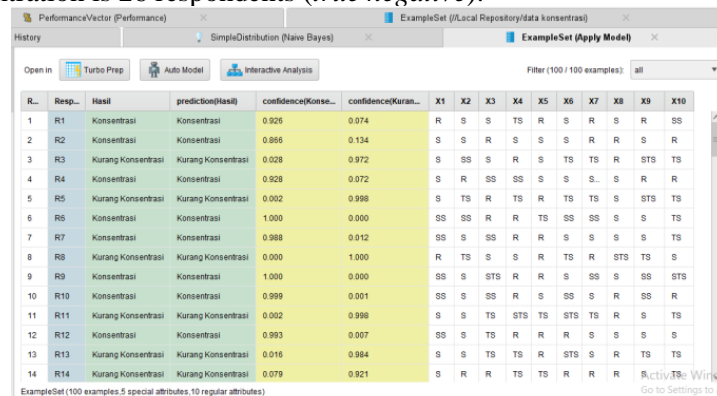


Image 13. Prediction Results of RapidMiner Application

Based on Figure 13. prediction of concentration and true lack of concentration are 4 respondents (*false positive*) found in respondents R29, R33, R38, and R96. While the prediction of lack of concentration and the reality of true concentration is 2 respondents (*false negative*) found in respondents R31 and R50.

4. CONCLUSION

The Naïve Bayes algorithm in determining student concentration is by using the ranges found by categorizing into concentration classes and less concentration classes. The range with a percentage of 61% to 100% includes the concentration class. Then the percentage below 61% includes the less concentration class. The data obtained came from a questionnaire distributed at SMPN 4 Cirebon City with 100 respondents. After that, suppose 5 testing data to simulate the Naïve Bayes model in predicting concentration classes. Of the 5 data, there are 2 concentration classes and 3 less concentration classes.

The probability of student concentration obtained is that there is a concentration class with a probability of 0.7 and a less concentration class with a probability of 0.3. With details of 70 concentration class respondents and 30 less concentration class respondents. The application of the Naïve Bayes algorithm in determining student concentration is by using the RapidMiner Studio Educational 10.3 application which can display 2 classes of classification results with an accuracy rate of 94.00% with details of 68 respondents predicting concentration and true concentration (true positive), 4 respondents predicting concentration and true lack of concentration (false positive), 2 respondents predicting less concentration and true concentration (false negative), and 26 respondents predicting less concentration and true lack of concentration (true positive). Future researchers are expected to compare the Naïve Bayes algorithm with other algorithms in order to compare the accuracy results..

REFERENCES

- Annur, H. (2018). Classification of Poor Communities Using the Naive Bayes Method. *ILKOM Scientific Journal*, 10(2), 160-165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Az Zahra, F. S., & Solikhun. (2021). Application of Naive Bayes Algorithm in Determining Student Concentration on the Teaching and Learning Process at Taman Asuhan Junior High School. *RESISTOR JOURNAL*, 4(2), 142-155. <https://doi.org/https://doi.org/10.31598>
- Jollyta, D., Ramadhan, W., & Zarlis, M. (2020). *Data Mining Concepts and Their Applications* (First). CV Budi Utama.
- Nofriansyah, D., & Nurcahyo, G. W. (2019). *Data Mining Algorithms and Testing*. CV Budi Utama.
- Nurfadilah, S., & Hakim, D. L. (2019). Student Learning Independence in the Mathematics Learning Process. *Proceedings of Sesiomadika*, 1214-1223.
- Riinawati. (2021). The Relationship between Student Learning Concentration and Student Learning Achievement during the Covid-19 Pandemic in Elementary Schools. *EDUCATIVE: JOURNAL OF EDUCATIONAL SCIENCE*, 3(4), 2305-2312. <https://doi.org/https://doi.org/10.31004/edukatif.v3i4.886>
- Yusalis, E. (2020). *Teacher's Efforts in Maintaining the Learning Concentration of Santri of Madrasah Diniyah Nurus Syifa Dusun Gupak Warak Krebet Village, Jambon Subdistrict, Ponorogo Regency*. Thesis IAIN Ponorogo.