Volume 4 No. 2 (July - December 2024), pp. 84-99 | ISSN 2962-2360 (online) | ISSN 2962-3669 (print)
Arcitech: Journal of Computer Science and Artificial Intelligence
http://journal.iaincurup.ac.id/index.php/arcitech/article/view/11635 | DOI : http://dx.doi.org/10.29240/arcitech.v4i2.11635

Predicting Early Childhood Readiness to Enter Elementary School Using the Naive Bayes Classification

Sukma Puspitorini[1], Novhirtamely Kahar[2], Ikah Kartika Sari[3]
[123] Informatics Engineering Study Program, Faculty of Computer Science, Nurdin Hamzah University, Indonesia
Email : pipietsukm4@gmail.com, novmely@ymail.com, ikahkartika640@gmail.com

Article Information

Article history

Received November 26, 2024
Revised December 8, 2024
Accepted December 9, 2024
Available Desember 30, 2024

Keywords

Naïve Bayes
Data Mining,
Fast Miner
Web Application
Predictive Model

Corresponding Author:

Sukma Puspitorini,
Nurdin Hamzah University,
Email : pipietsukm4@gmail.com

Abstract

This study aims to examine the readiness and maturity of early childhood in entering elementary school using the Naïve Bayes method. This analysis involves variables such as gender, age, aspects of physical-motor, cognitive, social-emotional development, and literacy skills which include reading, writing, arithmetic, and children's level of independence. The readiness category is classified into two classes, namely "ready" and "not ready". This prediction model is designed to provide a comprehensive understanding of the factors that affect the classification results, so that the evaluation process can be carried out in a transparent, objective, and data-driven manner. This research is expected to be a reference for other educational institutions in implementing a similar model to evaluate student readiness systematically. By adjusting variables and data according to local needs, this model has the potential to support more accurate and standardized decision-making, as well as improve the quality of early childhood preparation in entering formal education. The results show that the Naïve Bayes method is able to achieve an accuracy level of 93.33%, confirming its effectiveness in identifying early childhood readiness optimally.

Keywords : *Naïve Bayes, Data Mining, Fast Miner, Web Application, Predictive Model*

1. Introduction

Early Childhood Education plays an important role in the formation of the foundation of a child's cognitive, affective, and psychomotor development. Law Number 20 of 2003 concerning the National Education System mandates Early Childhood Education (PAUD) as an effort to develop children from birth to the age of six through the provision of educational stimuli that include physical-motor, language, art, religious, cognitive, social-emotional, and independence aspects, with the aim of optimizing potential and readiness to enter the next level of education. Thus, Early Childhood education is a level of education before basic education, which is a coaching effort carried out through the provision of educational stimuli to help growth enter further education, which is held on formal, non-formal, and informal pathways in

Indonesia (Rosita, Alawiyah, & Diananda, 2021) (Sartika & Wulandari, 2023) (Hidayat & Nurlatifah, 2023).

The Puspita Muatiara Play Group (KB), as the organizing institution of PAUD, applies a pedagogical approach based on experiential learning by integrating play activities while learning in the age group of 3-6 years. Institutional orientation is aimed at developing children's competencies which include aspects of attitude, knowledge, skills, and creativity, as mandated in the 2013 Curriculum, in order to ensure children's readiness to adapt to the elementary school environment. Although the conceptual framework of early childhood education prioritizes all-round development, phenomena in the field show a difference between theory and practice. Many parents tend to focus on mastering calistung (reading, writing, and arithmetic) as the only indicator of school readiness. This overemphasis on academic skills can overlook other important aspects, such as social-emotional and motor development, which also affect a child's readiness to enter primary school (Susilarini, 2021) (Munisa et al., 2024) (Kamphorst et al., 2021).

Previous research has generally focused on measuring children's school readiness using methods such as intelligence tests or manual assessments by teachers. Such as research conducted by Syahrizal (2021) on a Descriptive Study of Children's Readiness to Enter Elementary School (SD) Through NST Test and IQ Test. This research was conducted in kindergarten and aimed to roughly determine the readiness and maturity of kindergarten-age children to enter elementary school based on the results of the test (Nijmeegse Schoolbekwaamheids Test (NST). Furthermore, another study by Susilariani (2021) on Early Detection of Readiness to Enter Elementary School Through the NST and CPM Children Personal Matrix tests). Gaps from these previous studies include the subjectivity of assessment which is influenced by factors such as observer bias or differences in assessment standards between teachers. In addition, there is a lack of Predictive Approaches for Early Identification. These studies are more descriptive, namely describing the condition of children's readiness at a certain time. However, it lacks a predictive model that can be used to identify children who have the potential to experience difficulties in primary school from an early age (Syahrizal, 2021) (Susilarini, 2021).

This study aims to answer this gap by developing a prediction model based on the Naïve Bayes algorithm that not only accurately identifies school readiness, but is also able to overcome the limitations of traditional approaches through data-driven analysis. The Naive Bayes approach offers significant improvements with its ability to accommodate diverse data, give weight to relevant features, and generate probabilistic predictions. This allows for early identification of children at risk of difficulties as well as more informative decision-making. In addition, the automation of the classification process with Naive Bayes increases efficiency and reduces subjectivity bias.

Classification is a process used to find models or functions that describe and differentiate data into classes with the aim of finding patterns that have more value than relatively large to very large data (Wijaya & Dwiasti, 2020, p.3). In simple terms, classification is a grouping where data is grouped into groups that have been provided previously (Saputra, 2023). Naïve Bayes is one of the machine learning methods in the classification that is rooted in Bayes' theorem which uses probability and statistics, namely predicting future opportunities based on previous past experience. The main characteristic of this naïve Bayesian classifier is a very strong (naïve) assumption of independence from individual conditions or events . In this study, the Naive Bayes Algorithm is used because of its easy but also effective and reliable algorithm in processing high-dimensional data and its ability to produce accurate and easy-to-interpret classification models.

The prediction model is expected to provide comprehensive and objective information about school readiness, so that it can be used by early childhood stakeholders, including teachers, parents, and policymakers, in formulating intervention strategies and optimizing early childhood learning programs. It is hoped that this research is not only relevant for KB Puspita Mutiara, but also has the potential to be adopted by other educational institutions with adaptation to the local context, so that the results are wider and more general (Damanik et al., 2022) (Astuti et al., 2020) (Wickramasinghe & Kalutarage, 2021).

2. Previous Findings

There are several previous studies that have been conducted, related to the title of the research taken by the researcher, namely research related to early childhood readiness to enter elementary school using several methods as follows:

Research by Sungkowo and Ambarwati (2023) on the Classification of Readiness of Kindergarten Children Entering Elementary School Using the Naive Bayes Method. This study uses independent variables to measure children's readiness such as age, reading, writing, arithmetic, and independence. However, these factors may influence each other. For example, reading ability may be closely related to writing ability, or children's independence may be affected by cognitive and social-emotional maturity, so it is necessary to normalize data to handle these interrelated variables. Furthermore, research on the Classification of Children's Readiness to Enter Elementary School Using the Naïve Bayes Algorithm and the C4.5 Algorithm by Fanani and Sintia (2024). Naïve Bayes' algorithm assumes independence between attributes. In this study, attributes such as reading, arithmetic, and material comprehension may be interrelated, so further analysis is needed. Furthermore, research by Rahayu et.al (2022) on predicting school readiness using machine learning based on a combination of Adam and Nestrof Momentum. This study did not include variables that directly reflect children's cognitive abilities, such as early reading ability or numeracy ability and social-emotional ability. The focus of this research is more on demographic factors such as age, gender, parental education, preschool status than on children's cognitive indicators.

Based on the gap in the previous study, this study will discuss the prediction of early childhood readiness at KB Puspita Mutiara to find out the class of student readiness in continuing education to elementary school more comprehensively by considering other variables besides demographic variables such as gender, age, namely variables of physical and motor development, cognitive development, social-emotional development, literacy (reading, writing, and counting), and independence. The classification of readiness classes is divided into two, namely the 'ready' and 'not ready' classes.

3. Research Methodology

This study uses a case study design, which focuses on analyzing early childhood readiness at KB Puspita Mutiara Kumpeh Ulu to continue to the elementary school level. This case study was chosen because it aims to explore in depth the factors that affect children's readiness in a given context. In addition, a quantitative approach is used where student assessment data is converted into nominal form so that the data normalization process can be carried out. This approach was chosen to get a clear picture of the readiness and maturity of early childhood to enter elementary school at KB Puspita Mutiara Kumpeh Ulu.

Student data collection was carried out, among others, by conducting interviews with the Principal of KB Pulpit Mutiara Kumpeh Ulu to obtain information about the indicators of student readiness and maturity assessment, which include aspects of physical-motor development, language, art, religion, cognitive, and social-emotional development. In addition, administrative data collection was also carried out, such as age, gender, and data on student development reports from KB Pulpit Mutiara, with an age range of 5-7 years as many as 100 students from the 2018-2024 school year. Furthermore, the identification of research instruments that will be used as variables in this study is carried out, including: a) Demographic data of KB Pulpit Mutiara students such as name, gender, and age, b) Student Development Assessment Data which includes physical-motor, cognitive, social-emotional aspects, reading, writing, arithmetic, and independence. The raw data on the learning achievement of KB Pulpit Mutiara students obtained are as follows:

Table 1. Student Learning Achievement Data

NO	Age	Physic Motoric	Cog-nitive	Social Emotional	Read	Write	Count	Indepe ndency	Readiness
1	7	BSB	BSH	BSB	BSH	BSB	BSH	TIDAK	Siap
2	5	BSH	MB	MB	BSH	MB	MB	YA	Belum Siap
3	6	BSH	BSH	BSH	MB	BSH	MB	TIDAK	Siap
4	6	MB	BSH	MB	MB	MB	MB	YA	Belum Siap
5	6	BSH	MB	BSH	BSH	BSB	BSH	TIDAK	Siap
6	6	BSB	BSH	MB	MB	BSH	BSH	TIDAK	Siap
7	6	MB	BSH	MB	MB	MB	MB	YA	Belum Siap
8	7	BSB	MB	BSH	BSH	BSB	BSB	TIDAK	Siap
…	…	…..	…..	…..	…..	…..	…..	…..	…..
85	7	BSH	BSB	BSH	BSB	BSH	BSH	TIDAK	Siap

Furthermore, data pre-processing is carried out where the data that has been collected in the previous process is cleaned and converted into a format that is in accordance with the Naive Bayes algorithm. The stages of preprocessing this research data are as follows:

a. Data cleaning. In this process, data that is less relevant to the classification process will be cleaned. In this study, the demographic data of name and gender were not used as classification variables. Then the variables of student development in reading, writing, and counting are combined into literacy variables. Reading and writing literacy is included in language literacy while numeracy is included in mathematical numerical literacy skills where all three are important to measure children's readiness to receive learning in elementary school. In addition, the combination of variables aims to simplify the number of variables that need to be analyzed but still comprehensively represent students' abilities (Saint & Septriyanti, 2023).
b. Data normalization. The main goal of data normalization is to make the data more consistent because all variables have the same scale. This can facilitate the comparison of data from different variables so that it can improve the accuracy of the model. In this study, the pre-processing data carried out includes:
   1. Physical-motor, cognitive, social-emotional, reading, writing, and arithmetic variable data in the form of categorical data are first converted into numerical data, namely MB (Starting to Develop) = 1, BSH (Developing as Expected) = 2, and BSB (Developing Very Well) = 3.
   2. Literacy variable data was obtained from the average of writing, reading, and arithmetic scores.
   3. Student independence data was obtained from physical-motor, cognitive, and social-emotional values. According to Kurniawati and Hayati (2020), a child's independence can generally be associated with the ability to do everything on their own without having to depend on others (physical-motor), be able to take

the initiative and overcome daily problems (cognitive), and be willing to share and control emotions (social-emotional). The calculation of children's independence is carried out as follows equation (1): (Kurniawati et al., 2020)

$$Kemandirian\ anak\ (KA) = \frac{skor\ perolehan\ anak}{skor\ maksimal}\ x\ 100 \qquad (1)$$

The results of preprocessing data for Physical-Motor, Cognitive, Social-emotional, Literacy, and Independence variables are seen in the Table 2 below.

Table 2. Results of Variable Data Preprocessing

No	Age	Physical Motor	Cognitive	Social Emotional	Literacy	Independence
1	7	3	2	3	2,3	98,7
2	5	2	1	1	1,3	59,2
3	6	2	2	2	1,3	78,9
4	6	1	2	1	1,0	65,8
5	6	2	1	2	2,3	72,4
6	6	3	2	1	1,7	78,9
7	6	1	2	1	1,0	65,8
8	7	3	1	2	2,7	85,5
….	….	….	….	….	….	….
85	7	2	3	2	2,3	92,1

4. The data normalization process further to ensure that data is within a uniform value range. The next normalization of physical-motor, cognitive, social-emotional, literacy, and independence variables is the Min-Max Scalling Method method which will transform the data in the range of 0-1 as the formula in the equation (2) (Permana & Salisah, 2022)

$$X' = \frac{Xi - Xmin}{Xmax - Xmin} \qquad (2)$$

Where X' = data value after normalization and X = data value I. Results of the process. Normalization can be seen in Table 3 below.

Table 3. Results of Variable Data Normalization

No	Age	Physical Motor	Cognitive	Social Emotional	Literacy	Independence
1	1,00	1,00	0,50	1,00	0,80	1,00
2	0,00	0,50	0,00	0,00	0,20	0,14
3	0,50	0,50	0,50	0,50	0,20	0,57
4	0,50	0,00	0,50	0,00	0,00	0,29
5	0,50	0,50	0,00	0,50	0,80	0,43
6	0,50	1,00	0,50	0,00	0,40	0,57
7	0,50	0,00	0,50	0,00	0,00	0,29
8	1,00	1,00	0,00	0,50	1,00	0,71
….	….	….	….	….	….	….
85	1,00	0,50	1,00	0,50	0,80	0,86

c. Data sharing. In this research, 85 data were used as training data and the remaining 15 data were used as testing data. Data distribution is carried out randomly (random sampling) to ensure a representative distribution of variables. The training data was then classified with the Naive Bayes Algorithm to calculate the probability of the elementary school readiness class. The flowchart of the Naïve Bayes Algorithm in general is shown in Figure 1 below

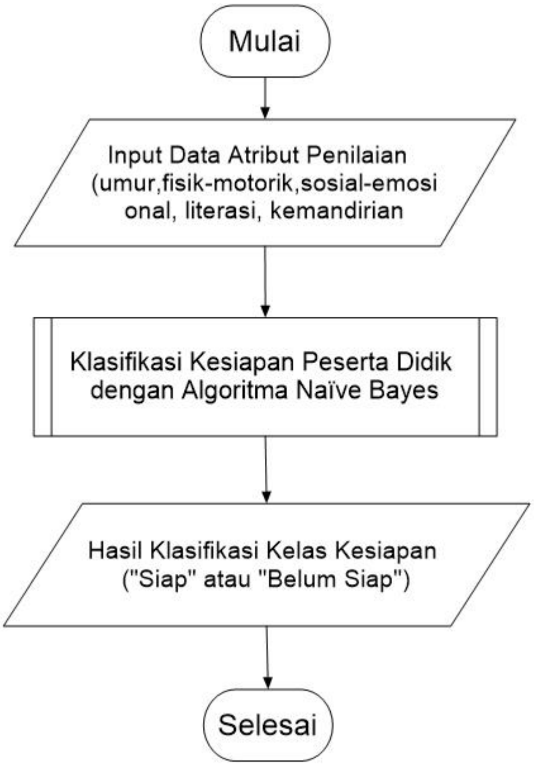


Figure 1. Naïve Bayes Algorithm Classification Flowchart

4. Results and Findings Analysis

4.1 Result

Furthermore, the training data is processed using Weka 3.8.6 tools. WEKA is a Java-based data mining tool that supports various tasks such as classification, clustering, regression, and visualization with a user-friendly GUI interface. Supporting file formats such as ARFF and CSV, WEKA offers flexibility, model evaluation capabilities such as cross-validation, as well as modularity to add new algorithms. Designed to make data analysis easier without the need for coding, WEKA is suitable for different skill levels of users. In this study, Weka was used to train the Naive Bayes classification model and evaluate its performance in predicting the readiness of KB Puspita Mutiara students to enter elementary school. In the preprocessing stage in Weka, the statistics of the 'Ready' and 'Not Ready' classes are shown in Figure 2 below. (Preet et al., 2020) (Pavic , 2023)

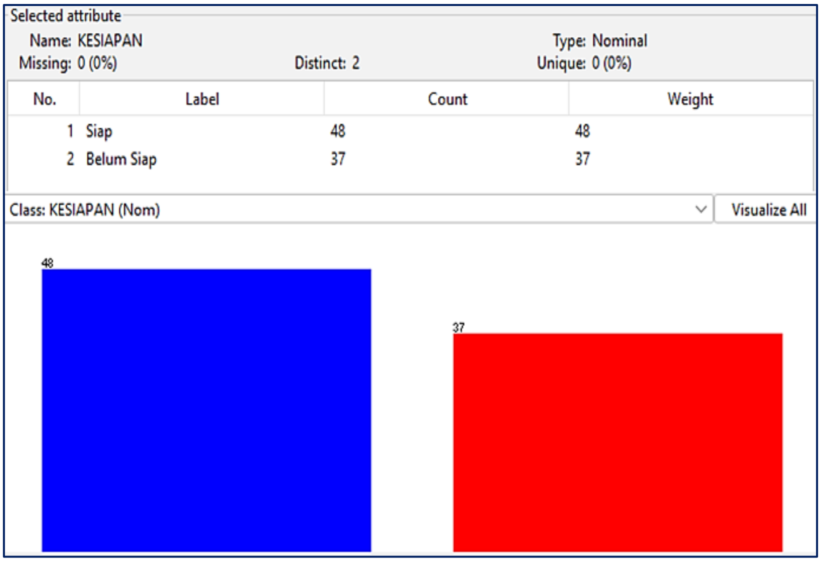


Figure 2. Amount of Data in Ready and Not Ready Classes

The results of the evaluation in the training process with the test option using the 'Use trainig set' are shown in Figure 3 below

```
Classifier output
=== Summary ===

Correctly Classified Instances          80               94.1176 %
Incorrectly Classified Instances         5                5.8824 %
Kappa statistic                          0.88
Mean absolute error                      0.0785
Root mean squared error                  0.2367
Relative absolute error                 15.9646 %
Root relative squared error             47.7393 %
Total Number of Instances               85

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.958    0.081    0.939      0.958    0.948      0.880    0.955     0.918     Siap
                 0.919    0.042    0.944      0.919    0.932      0.880    0.955     0.966     Belum Siap
Weighted Avg.    0.941    0.064    0.941      0.941    0.941      0.880    0.955     0.939

=== Confusion Matrix ===

  a  b   <-- classified as
 46  2 |  a = Siap
  3 34 |  b = Belum Siap
```

Figure 3. Classifier Output Results on Weka

Based on the output given in the figure above, the Naive Bayes model has shown quite good performance in classifying data. The explanation of the classification results which includes several measurements is as follows (Singh et al., 2023)

a. The number of correctly classified data is 80 out of 85 data while the number of incorrectly classified data is 5 out of 85 data.
b. Kappa statistics: A measure of agreement between the predicted outcome and the actual value, with a value close to 1 indicating excellent agreement.
c. Error metrics: Several error metrics such as Mean Absolute Error, Root Mean Squared Error, and others are used to measure how far the prediction is from the actual value.
d. Accuracy. Overall Accuracy shows that the model successfully classifies about 94% of the data correctly. This shows that the model is quite accurate in predicting instance classes.
e. In the Confusion Matrix, the results were obtained: True Positives (TP): 46 instances of the "Ready" class (or 54%) were correctly classified and Negatives (TN): 34 instances of the "Not Ready" class (or 40%) were correctly classified.
f. F-Measure: The harmonic average value between Precision and Recall. This provides an overall overview of the model's accuracy. A high F-Measure value in this readiness class model (i.e. 0.9) indicates that the model is both in Precision and Recall.
g. MCC (Matthews Correlation Coefficient): A measure of correlation between predicted and actual values, with a value close to 1 indicating an excellent correlation.
h. ROC Area: The area under the ROC curve, which is a measure of the model's overall performance. A value close to 1 in the readiness class model in this study shows good performance.

Other results obtained from the data training process with the Naïve Bayes classification algorithm using Weka include mean values, standard deviation, weight sum, and precision as shown in Figure 4 below.

Attribute	Class Siap (0.56)	Belum Siap (0.44)
Usia		
mean	0.7396	0.3378
std. dev.	0.2498	0.286
weight sum	48	37
precision	0.5	0.5
Fisik Motorik		
mean	0.5625	0.2973
std. dev.	0.1654	0.2954
weight sum	48	37
precision	0.5	0.5
Kognitif		
mean	0.5208	0.1622
std. dev.	0.3529	0.3087
weight sum	48	37
precision	0.5	0.5
Sosial Emosional		
mean	0.5521	0.2297
std. dev.	0.293	0.2985
weight sum	48	37
precision	0.5	0.5
Literasi		
mean	0.7375	0.1784
std. dev.	0.1922	0.1905
weight sum	48	37
precision	0.2	0.2
Kemandirian		
mean	0.6786	0.2934
std. dev.	0.1688	0.2462
weight sum	48	37
precision	0.1429	0.1429

Figure 4. Output classifier value results for each class

Mean can be used to determine the center value of a feature in each class whereas standard deviation is used to understand how spread the data is around the mean. To make predictions on new data based on the classification model that has been generated, a Threshold is used that can be determined between the average of the Ready and Not Ready classes. The threshold value added to each attribute can be used as a separator between the "Ready" and "Not Ready" classes. In general, if the value of an attribute on an instance exceeds that threshold, then the instance is likely to be classified into the "Ready" class. Conversely, if the value of the attribute is below the threshold, then the instance is likely to be classified into the "Not Ready" class (Firmansyah & Yulianto, 2023).

With these probability-based prediction rules, a website-based prediction interface can be built. The application built is expected to help and make it easier for users to process data and identify the readiness of new student data from KB Puspita Mutiara. The implementation of the website interface can be seen as follows:

1. Student data menu

The student data display in Figure 5 is a student data page used by operators to search for or manage student demographic data such as information such as NIS, Name, Gender, place/date of birth, school year and address.

Figure 5. Data Input Menu

2. Manage Data Menu

The menu in figure 6 is the data management page, this page is used by users to add data such as training data, testing data, and prediction data. On this page, the data is stored according to categories and processed to perform calculations.

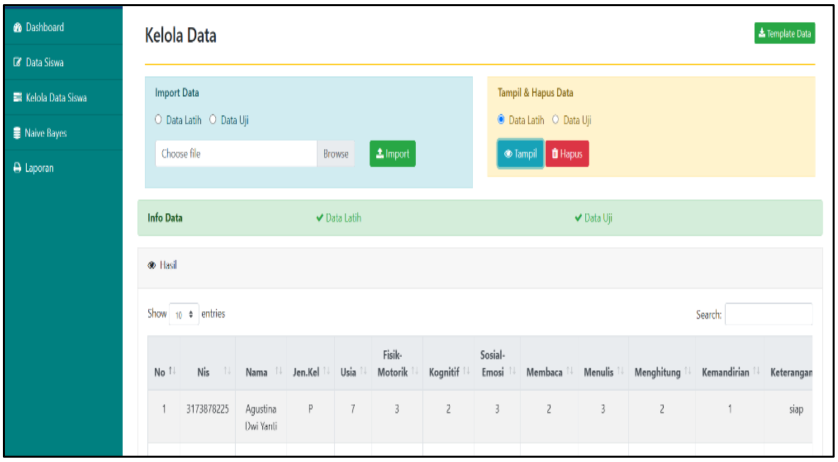
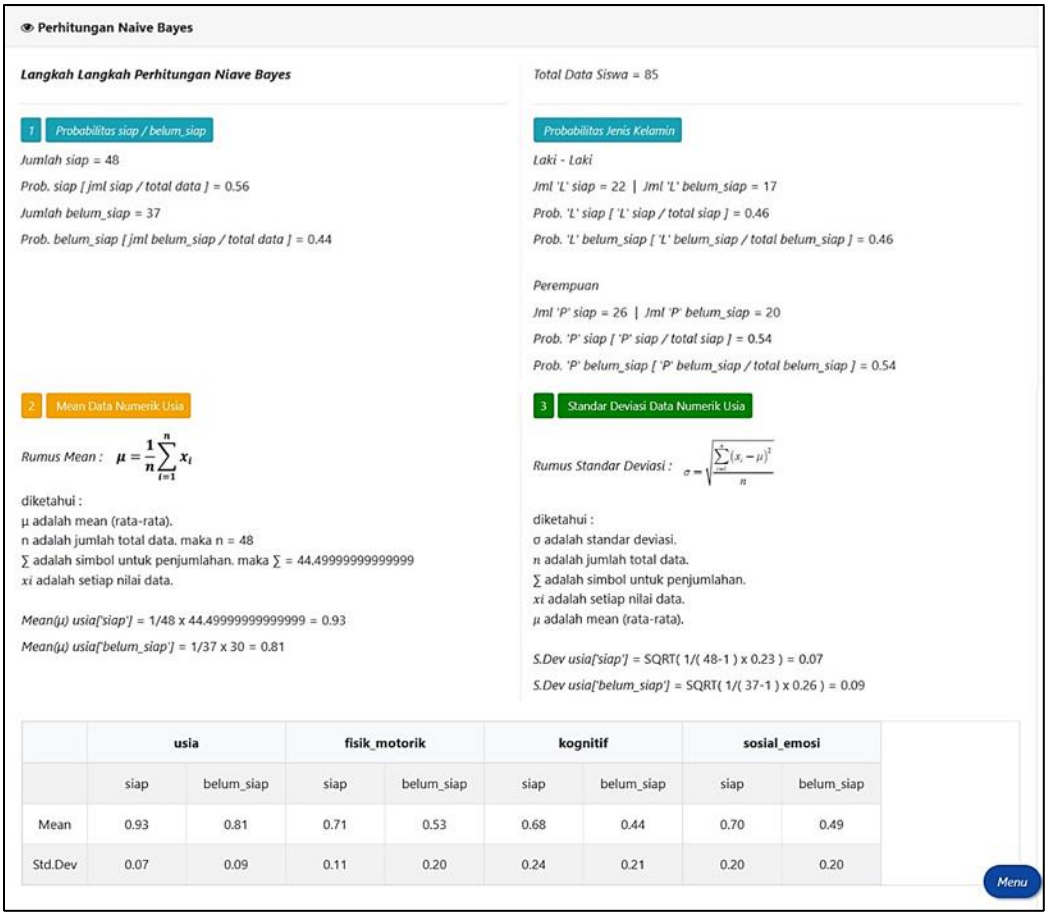


Figure 6. Manage Data Menu

3. Naïve Bayes Process Menu

Figures 7 to 9 show the process of calculating naïve bayes. On this page, training data testing is performed to determine the naïve bayes classification pattern and to calculate the prediction of the readiness class of the test data.



Perhitungan Naive Bayes

Langkah Langkah Perhitungan Niave Bayes

1 *Probabilitas siap / belum_siap*

Jumlah siap = 48

Prob. siap [jml siap / total data] = 0.56

Jumlah belum_siap = 37

Prob. belum_siap [jml belum_siap / total data] = 0.44

Total Data Siswa = 85

Probabilitas Jenis Kelamin

Laki - Laki

Jml 'L' siap = 22 | Jml 'L' belum_siap = 17

Prob. 'L' siap ['L' siap / total siap] = 0.46

Prob. 'L' belum_siap ['L' belum_siap / total belum_siap] = 0.46

Perempuan

Jml 'P' siap = 26 | Jml 'P' belum_siap = 20

Prob. 'P' siap ['P' siap / total siap] = 0.54

Prob. 'P' belum_siap ['P' belum_siap / total belum_siap] = 0.54

2 Mean Data Numerik Usia

Rumus Mean : $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

diketahui :
μ adalah mean (rata-rata).
n adalah jumlah total data. maka n = 48
∑ adalah simbol untuk penjumlahan. maka ∑ = 44.49999999999999
xi adalah setiap nilai data.

Mean(μ) usia['siap'] = 1/48 x 44.49999999999999 = 0.93

Mean(μ) usia['belum_siap'] = 1/37 x 30 = 0.81

3 Standar Deviasi Data Numerik Usia

Rumus Standar Deviasi : $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$

diketahui :
σ adalah standar deviasi.
n adalah jumlah total data.
∑ adalah simbol untuk penjumlahan.
xi adalah setiap nilai data.
μ adalah mean (rata-rata).

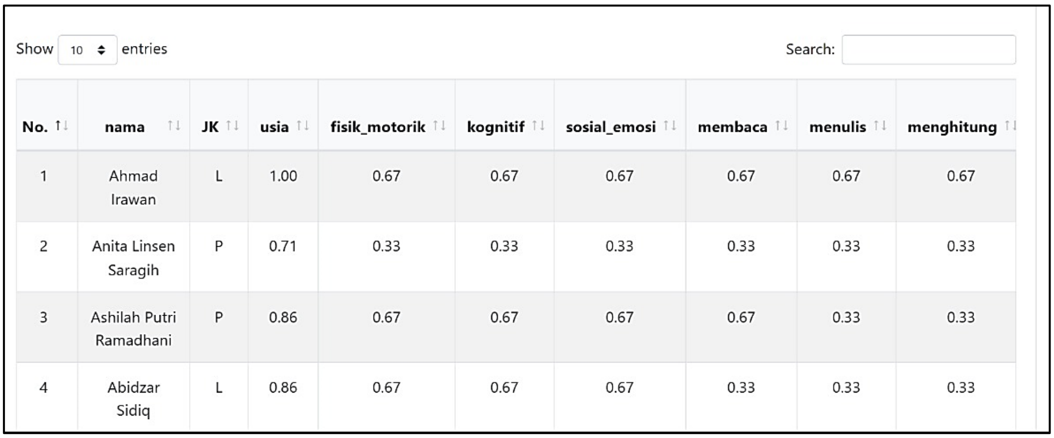
S.Dev usia['siap'] = SQRT(1/(48-1) x 0.23) = 0.07

S.Dev usia['belum_siap'] = SQRT(1/(37-1) x 0.26) = 0.09

	usia		fisik_motorik		kognitif		sosial_emosi	
	siap	belum_siap	siap	belum_siap	siap	belum_siap	siap	belum_siap
Mean	0.93	0.81	0.71	0.53	0.68	0.44	0.70	0.49
Std.Dev	0.07	0.09	0.11	0.20	0.24	0.21	0.20	0.20

Menu

Figure 7. Naïve Bayes Process Menu : Determine Class Probability



Show 10 entries Search:

No.	nama	JK	usia	fisik_motorik	kognitif	sosial_emosi	membaca	menulis	menghitung
1	Ahmad Irawan	L	1.00	0.67	0.67	0.67	0.67	0.67	0.67
2	Anita Linsen Saragih	P	0.71	0.33	0.33	0.33	0.33	0.33	0.33
3	Ashilah Putri Ramadhani	P	0.86	0.67	0.67	0.67	0.67	0.33	0.33
4	Abidzar Sidiq	L	0.86	0.67	0.67	0.67	0.33	0.33	0.33

Figure 8. Prediction Results Table View

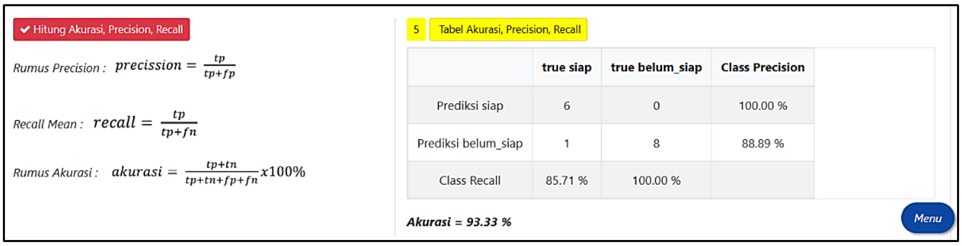


Figure 9. Accuracy Result Display

In Figure 9 above, the display of the accuracy of *Bayes' naïve* calculation shows an accuracy rate of 93.33%. This shows the same level of accuracy as *the naïve bayes calculation process* in the RapidMiner app.

4.2 Result Analysis

Based on the results of performance measurements carried out by the naïve bayes algorithm on the rapidminer application. The test is carried out to be the same as the manual calculation and the running system, the analysis of the results that can be taken is as follows:

1. The performance generated by Naïve Bayes' algorithm using the RapidMiner application results in 100% accuracy results. The results show that the results of student predictions for students who already have the readiness and maturity to enter elementary school are developing very well and in accordance with expectations.
2. The Simple Distribution value contained in the RapidMiner application indicates the probability value of the class it has. The values in the RapidMiner application and the manual excel calculation process have the same class probability values, i.e., the ready class is 0.56 and the unready class is 0.44.
3. The testing process on the website system with an accuracy result of 93.33% where this result is the same as the accuracy result at RapidMiner.
4. For testing on Puspita Mutiara KB students, the researcher used 15 student data in the 2023/2024 academic year which was used as test data. The results obtained from manual calculations, the RapidMiner application, and the website system have the same level of accuracy. Table 1 below shows a comparison of the results of the classification of student readiness at KB Puspita Mutiara using Weka and web prediction.

Table 4. Comparison of Readiness Classification Predictions

No	NIS	JK	Age (year)	Weka Results	Application Prediction	Fit/Not Fit
1	3167162066	L	7	Siap	Siap	Fit
2	3185817744	P	5	Belum Siap	Belum Siap	Fit
3	3175535758	P	6	Belum Siap	Siap	Not Fit
4	3176589376	L	6	Belum Siap	Belum Siap	Fit
5	3194753892	P	6	Siap	Siap	Fit
6	3177492041	L	6	Siap	Siap	Fit
7	3194653840	L	5	Belum Siap	Belum Siap	Fit
8	3165819539	P	6	Siap	Siap	Fit
9	3176612838	L	6	Belum Siap	Belum Siap	Fit
10	3170463306	L	7	Siap	Siap	Fit
11	3109354882	P	7	Siap	Siap	Fit
12	3184528330	P	6	Belum Siap	Siap	Not Fit
13	3162717094	P	7	Siap	Siap	Fit
14	3164836936	P	6	Belum Siap	Belum Siap	Fit
15	3178127527	L	6	Belum Siap	Belum Siap	Fit

Based on the results of the comparison table above, it can be concluded that the results of comparing real data with data in the website system have different results. The accuracy result is 93.33% for predictions using the web, and the accuracy result is 100% for the RapidMiner application. This difference occurs in data with numbers 3 and 12 that do not fit.

5. Conclusion

This study proves that the Naïve Bayes algorithm is very effective in assessing early childhood readiness to enter elementary school. The analysis conducted using data from 100 students of KB Puspita Mutiara showed an accuracy level of 93.33%, both in implementation through the WEKA application and web-based systems. The use of WEKA provides accurate evaluation results with metrics such as accuracy, kappa statistic, F-measure, and ROC area that is close to the ideal value. Meanwhile, the web-based system simplifies the data processing process directly, allowing readiness evaluations to be carried out quickly and efficiently. Theoretically, this research supports the development of the Naïve Bayes algorithm as a data-driven evaluation tool that can improve objectivity in education. Practically, web-based applications offer solutions that are easy for other educational institutions to implement.

In terms of policy, this model contributes to creating a standardized and data-based student readiness evaluation system, thereby supporting fairer and more measurable decision-making. For future studies, it is recommended to use a larger dataset with wider coverage of the region to improve the generalization of results. In addition, the addition of variables such as the influence of family environment, culture, and access to education can enrich the analysis model. The research can also integrate other algorithms, such as Decision Tree or Random Forest, to compare performance and find the most optimal approach. Finally, the development of predictive features in web-based applications, such as visualization of evaluation results and intervention recommendations, can provide greater practical benefits to users.

6. Acknowledgements

We would like to thank the researchers for their contributions to this paper: Ikah Kartika, and Mrs. Novhirtamely Kahar, ST. M.Kom as a co-author who provided support and advice during the research.

7. Author's Note

The author states that there is no conflict of interest regarding the publication of this article. The author emphasizes that the data and papers are free from plagiarism.

References

Astuti, R. W., Puspitorini, S., & Kurnia Akbar, F. (n.d.). *Implementation of the naïve bayes theorem in the analysis and prediction of the field of work of alumni of the stmik informatics engineering study program, nurdin hamzah jambi.*

Aziz, S. Al, & Septriyanti, Y. (2023). Correlation between Indonesian Literacy and Mathematical Numeracy Literacy of Students in Solving Mathematics Problems. *Lattice Journal : Journal of Mathematics Education and Applied*, *3*(1), 14. https://doi.org/10.30983/lattice.v3i1.6324

Damanik, A. R., Hartama, D., Damanik, I. S., Information, S., & Tunas Bangsa, S. (n.d.). *SYSTEMASI: Journal of Prediction Information System of Mortality Rate in Indonesia due to Covid-19 Using the Naïve Bayes Algorithm*. http://sistemasi.ftik.unisi.ac.id

Firmansyah, F., & Yulianto, A. (2023). Prediction of Learning Outcomes Using Naïve Bayes Classifier at the Elementary School Level. *Remik*, *7*(2), 1174–1182. https://doi.org/10.33395/remik.v7i2.12375

Hidayat, Y., & Nurlatifah, L. (2023). Comparative analysis of the level of early childhood development achievement (stppa) based on permendikbud no. 137 of 2014 with permendikbudristek no. 5 of 2022. *Journal of Intisabi*, *1*(1), 29–40. https://doi.org/10.61580/itsb.v1i1.4

Kamphorst, E., Cantell, M., Van Der Veer, G., Minnaert, A., & Houwen, S. (2021). Emerging School Readiness Profiles: Motor Skills Matter for Cognitive- and Non-cognitive First Grade School Outcomes. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.759480

Kurniawati, N., Hayati, T., Islam, P., Age, A., Uin, D., Gunung, S., Bandung, D., Cimencrang, J., Bage, G., Bandung, K., & Barat, J. (n.d.). *Increasing children's independence through practical life skill activities*.

Munisa, Nofianti, R., & Andini, N. R. (n.d.). *The readiness of children aged 5-6 years to enter elementary school.*

Pavic, J. (n.d.). *202307_PavicJakov_WEKA*.

Permana, I., & Salisah, F. N. S. (2022). Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, *2*(1), 67–72. https://doi.org/10.57152/ijirse.v2i1.311

Preet, K., Attwal, S., & Singh Dhiman, A. (2020). Exploring data mining tool-weka and using weka to build and evaluate predictive models. In *Advances and Applications in Mathematical Sciences* (Vol. 19, Issue 6). http://www.cs.waikato.ac.nz/ml/weka/downloading.html.

Sartika, L., & Wulandari, D. (2023). Early childhood islamic education institutions. *Al Athfaal Journal*, *1*(1), 2–6. https://doi.org/10.62214/jalfal.v1i1.92

Singh, S., Kaur, H., Kanozia, R., & Kaur, G. (2023). Empirical Analysis of Supervised and Unsupervised Machine Learning Algorithms with Aspect-Based Sentiment Analysis. *Applied Computer Systems*, *28*(1), 125–136. https://doi.org/10.2478/acss-2023-0012

Susilarini, T. (n.d.). *Early detection of readiness to enter elementary school through the Nijmeegse Schoolbekwaamheiid Test (NST) and Children Personal Matrix Test (CPM) (a quantitative descriptive study on kindergarten children who will enter elementary school).*

Syahrizal, S. (2021). A Descriptive Study of Children's Readiness to Enter Elementary School (SD) Through NST Test and IQ Test at Kindergarten Pembina Kisaran Asahan Regency. *Social Library Journal, 1*(3), 101–106. https://doi.org/10.51849/sl.v1i3.47

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, *25*(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6