

# Implementation of Naïve Bayes and K-NN Algorithms in Diagnosing Stunting in Children

Wulan Widhari<sup>1</sup>, Agung Triayudi<sup>2\*</sup>, Ratih Titi Komala Sari<sup>3</sup>

Faculty of Communication and Information Technology, Universitas Nasional, Jakarta, Indonesia

Author Email: wulanwdr21@gmail.com<sup>1</sup>, agungtriyudi@civitas.unas.ac.id<sup>2\*</sup>, ratih.titi@civitas.unas.ac.id<sup>3</sup>

**Abstract.** Indonesia faces a huge potential risk of stunting, as revealed in the Indonesian Nutrition Status Analysis according to 2022 data, the stunting rate reached 24.22% in 514 districts / cities throughout Indonesia. To prevent stunting in children, early detection can be done. This research was conducted to compare the performance of two algorithms Naive Bayes and K-NN to predict stunting cases in children, to get a better picture of how classification algorithms predict stunting cases with a better level of accuracy and responsiveness, comparison experiments of several algorithms are needed using specific datasets to develop an optimal classification model. Based on the results of performance testing on the K-Nearest Neighbor and Naive Bayes methods in testing the performance of accuracy, precision, recall, and f1-score, the results of performance testing on the naïve bayes method obtained performance values on 30% testing data are accuracy of 71%, precision 71%, recall 76%, and f1-score 73%. The performance results of the K-NN method using the euclidean distance measurement obtained the best performance value, namely accuracy of 97%, precision of 98%, recall of 96%, f1-score of 97% at a value of  $k = 3$ . Based on the performance results of the comparison of the Naive Bayes and K-NN methods, it shows that the best classification method on the stunting dataset is the K-NN method because it gets better performance than the Naive Bayes method.

**Keywords:** Classification, Comparison, Naive Bayes, K-Nearest Neighbors, Stunting

## 1 Introduction

Stunting refers to impaired growth in children aged two years or less that occurs during the first thousand days in the womb and will affect how long children live. Stunting can lead to being underweight, risk of obesity, poor reproductive health, and reduced productive capacity. Income, education, maternal knowledge, and family size are among the factors that can lead to stunting [1]. Indonesia faces a huge potential risk of stunting, as revealed in the Indonesian Nutrition Status Analysis. According to 2022 data, the stunting rate reached 24.22% in 514 districts/cities across Indonesia. Although the rate will decrease significantly, stunting is still considered a serious problem in Indonesia as the prevalence rate is always above 20% [2].

The percentage of children under five in a group who are stunted in their physical growth is called the prevalence of stunting. This indicator is used to evaluate the nutritional problems of children under the age of five in a region or country. Higher values indicate that the problem is more serious and requires greater effort to solve. The World Health Organization set a target of a stunting prevalence rate below 20% as part of the fulfillment of the Sustainable Development Goals (SDGs). In Indonesia, where there are many cases of inadequate nutrition in children under five, parents need to carefully understand their child's nutritional condition. Assessing a child's physical condition alone is not perfect for assessing their overall health. Therefore, the role of parents, especially mothers, is crucial in providing adequate food intake for their children, as healthy consumption patterns have a significant influence on children's growth and development [3].

To prevent stunting in children, early detection can be done. Prevention of stunting can be done by involving training for parents of toddlers, posyandu health teams, and related medical personnel. Training is expected to provide more understanding to parents about how to prevent stunting as well as enrich the insights of health cadres in identifying children under five who may experience stunting [4]. In the current era, where technology plays a central role in various activities, building a system that can identify stunting is a relevant solution to overcome this problem. A system for diagnosis is a computer system that functions like a decision-making expert in a particular field [5].

To solve the problem of evaluating stunting in children, more creative evaluation methods, such as machine learning, are needed to predict the likelihood of a child being diagnosed with stunting and to enable preventive

measures to be taken proactively and appropriately. For classification, ML uses many algorithms that can be used. To gain a better understanding of the specifics required to develop an optimal classification model. This research provides how classification algorithms predict stunting cases with better accuracy and responsiveness, comparison experiments of various algorithms using datasets that further understanding of how various algorithms work in stunting prediction, and emphasizes the importance of using ML in the diagnosis of stunting in children [6].

To improve these activities, which can safely store data, simplify data management, and have the ability to identify stunting symptoms in young children by utilizing the Bayes theorem approach, The Bayes theorem method is a method of drawing a conclusion and making a decision with a probability value [7]. The study discusses the application of NBC on the nutritional status classification of stunting toddlers using K-Fold Cross Validation testing, resulting in the nutritional status classification of stunting toddlers can be recognized and classified properly at each iteration. The classification performance results of the NBC algorithm validated with 10-Fold Cross Validation show that the 8th iteration has the highest accuracy of 95.14%, and the 3rd iteration has the lowest accuracy of 81.73%. Overall, the average accuracy for each iteration is 88.53%. The results of the classification can be used as a model to make predictions with the same characteristics or data variables. If there is data with the same variables, GUI-R can classify the data more quickly and effectively [8].

The application of the K-NN method using a website as the basis for creating a classification system for growth and development disorders in infants can provide assistance to posyandu officers in monitoring the development of toddlers. In the context of the toddlers' stunting status classification, the system managed to achieve an accuracy of 83% with an error rate of 0.167 [9]. This research will produce a classification system for stunting status in toddlers that can help determine their stunting status. Then calculations are carried out to evaluate the performance of the Naive Bayes classification data model created using the confusion matrix. The results of the confusion matrix test indicate that the accuracy, precision, and recall scores are 58%, 68%, and 58%, respectively. 30% of the training data and 70% of the testing data were used in the trials [10].

The implementation of the Naive Bayes algorithm in the Kalitengah Village Posyandu in the classification of nutritional status of toddlers indicates that the Naive Bayes method can successfully identify the nutritional status of toddlers with an accuracy of 87.33% of the fourteen test data [11]. In research that discusses the use of the K-NN algorithm in determining stunting conditions in children, the results of testing using 114 data indicate that the K-NN algorithm is able to categorize children's stunting status by considering anthropometric attributes such as age (U), body weight (BW), height (TB), and head circumference (LK). Tests were conducted to obtain the highest accuracy results for k values of 3, 5, 7, and 9. The results showed that the highest accuracy and lowest error rate were found at  $k = 3$ , reaching 83% with an error rate of 0.142 [12].

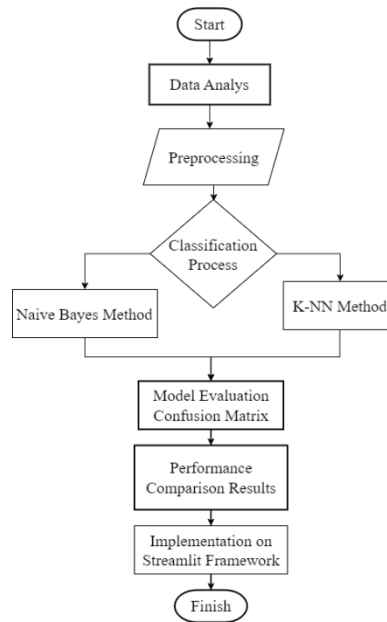
The evaluation results indicated that Random Forest produced the highest accuracy of 87.75%, then K-NN 84.8%, and Naive Bayes 83.2%. Although Random Forest produced good accuracy, K-NN stood out as they were able to get most of the actual stunting cases. However, Random Forest has a good balance between accuracy and recall, thus proving that due to the combination of good accuracy with the ability to find stunting cases better than other models in this evaluation, this model can be a good choice for diagnosing stunting [6].

By adding together all of the frequencies and value combinations from a certain collection of datasets, the Naive Bayes Classifier calculates a set of probabilities. The algorithm is based on Bayes' theorem and assumes that each attribute is considered independent or not interdependent, as given by the value of the class variable. K-Nearest Neighbor (K-NN) is one of the methods in the instance-based learning category and belongs to the lazy learning technique. The K-NN process involves finding the K objects in the training data that have the closest distance or highest similarity to objects in the new data or testing data [13].

By comparing the performance of two machine learning algorithms, namely Naive Bayes and K-Nearest Neighbor algorithms, for predicting child stunting cases, this study aims to determine the best classification model. The results will hopefully provide a better understanding of how classification algorithms work, which will enable the development of better stunting diagnosis models.

## 2 Metodologi

In this study, the data used is sourced from the Kaggle site, namely the Stunting Dataset. After the data is obtained, then the preprocessing stage is carried out, and then the data will be classified with the Naive Bayes and K-NN models. After being classified, it is continued with model evaluation using the Confusion Matrix testing method. Then the comparison of the two algorithms is carried out by comparing accuracy, precision, recall, and f1-score to find out which algorithm or method can perform the best classification. The following figure illustrates the process design flow.



**Figure 1.** Process Design Flow

## 2.1 Data Retrieval

This research uses data sourced from the Kaggle site, namely the Stunting Dataset. In the dataset there are 6500 data which contains 8 attributes, namely Sex, Age, Birth Weight, Birth Length, Body Weight, Body Length, Exclusive Breastfeeding, and Stunting. Of these attributes, the independent variable is 7 attributes and the dependent variable is 1 attribute. This data will be used to classify using the Naive Bayes and K-NN algorithms.

## 2.2 Preprocessing

Data processing is done at this point so that it can be utilised in the classification procedure. Data normalisation, data transformation, and data cleaning are the phases. The dataset that is used is split into two parts: 30% is used for testing and 70% is used for training. . Of the total 6500 data is divided into 4550 training data and 1950 testing data.

## 2.3 Naive Bayes

Naïve Bayes is used to estimate the probability of each class assuming that the classes are independent of each other. In this method, all attributes are considered independent and contribute to decision-making with equal weight. As a decision-making tool, Naive Bayes is also used to update the level of confidence in information [14]. The use of the Naive Bayes classification method uses probability calculations. The basic concept of this method is Bayes' theorem, which is applied in statistics to determine the chances of the Naive Bayes classifier to calculate the class probability for each existing group and the attributes that determine the optimal class [4]. The general form of the naive bayes method equation formula is shown in equation 1.

$$P(K|G) = \frac{P(G|K) \cdot P(K)}{P(G)} \quad (1)$$

## 2.4 K-Nearest Neighbor

K-NN has several advantages, including being able to handle large datasets and being easy to use. The K-NN technique uses the following steps, namely inputting training data, training data labels, k, and testing data, calculating the distance between each testing data to each training data, determining k from the training data that has the closest distance to the testing data, then testing, then checking the k data labels, then determining the label with the highest frequency, then entering the testing data into the class with the highest frequency [15].

Equation 2 applies a formula to determine the distance between two points, specifically training data points and testing data points.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2}$$

### 2.5 Evaluasi

Confusion Matrix is useful for understanding the differences between classes in classification cases. To measure the performance of Naive Bayes and K-Nearest Neighbor algorithms, Confusion Matrix is used in the evaluation stage. The Confusion Matrix table is a table containing predictions and actual values. The Confusion Matrix table shows the number of correct and incorrect data classified. True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) are the four values generated in the confusion matrix table [10]. The following figure shows an illustration of the confusion matrix table.

		Actual	
		Positive	Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 2. Confusion Matrix Table Illustration

Accuracy, Standard deviation, Precision, F1-Score, Recall and specificity are the evaluation criteria considered. In this study, the evaluation carried out is by calculating the value of accuracy, precision, recall, and F1-Score. [8]. The general form of the equation can be shown in the following equations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \tag{6}$$

### 3 Result and Discussion

The research conducted uses a stunting dataset derived from Kaggle. At this stage, the data is divided into training data and testing data and is determined to be 70% training data and 30% testing data. The application of manual calculations for the Naive Bayes and K-NN algorithms uses 11 data samples, namely 10 training data and 1 testing data. Tables 1 and 2 show the 11 data samples.

Table 1. Sample Dataset: Training Data

Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	ASI Eksklusif	Stunting
F	56	2,9	50	11	90	Yes	No
F	20	3,3	49	11,1	80,5	No	No
M	4	2,8	48	6,5	63	No	No
F	14	2	49	7	71	Yes	No
M	32	3,2	49	11	88,7	Yes	No
M	30	2,3	50	12	90	Yes	No
M	2	2,9	49	8,5	74,2	Yes	No
M	33	2,5	49	10	91,5	No	Yes

M	33	3	50	15	96	Yes	No
F	15	4	51	5,9	58,3	Yes	No
M	16	2,4	48	6,6	76	Yes	Yes
F	47	2,5	49	10	91,5	No	Yes
F	31	3,3	49	9,1	90	Yes	No
F	25	2,6	49	9,8	90	No	No
F	4	2,7	47	5,8	69,6	Yes	No
F	12	3,6	50	8,2	70,5	Yes	No
F	3	2,8	48	6	54	No	Yes
F	55	3	49	8,5	81	No	No
F	12	2,9	49	5,8	69,5	Yes	No

**Tabel 2.** Sample Dataset: Testing Data

Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	ASI Eksklusif	Stunting
M	7	3,2	49	9	68	Yes	?

### 3.1 Naive Bayes Implementation

The dataset used consists of 10 data as training data shown in Table 1 and Table 2 as testing data. After determining the next stage, the process of implementing the Naive Bayes method using equation 1. The data samples used for manual calculations are shown in Tables 1 and 2, The next process is data transformation or converting data into a form that is more in accordance with a format that can facilitate the stunting prediction process. The transformed category data can be seen in Tables 3 and 4 after the data is transformed.

**Table 3.** Category Data from Data Transformation

Category	Information
Age < 16	Age under 16 months
Age 16 – 30	Age between 16 to 30 months
Age > 30	Age above 30 months
Birth Weight < 3	Birth Weight under 3kg
Birth Weight >= 3	Birth Weight of 3 kg
Birth Length < 48	Birth Length below 48 cm
Birth Length >= 48	Birth Length above 47 cm
Body Weight < 10	Body Weight below 10 kg
Body Weight >= 10	Body Weight of 10 kg
Body Length < 90	Body Length below 90 cm
Body Length >= 90	Body Length of 90 cm

**Table 4.** After Data Transformation

Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	ASI Eksklusif	Stunting
F	> 30	< 3	>= 48	>= 10	>= 90	Yes	No
F	16 - 30	>= 3	>= 48	>= 10	< 90	No	No
M	< 16	< 3	>= 48	< 10	< 90	No	No
F	< 16	< 3	>= 48	< 10	< 90	Yes	No
M	> 30	>= 3	>= 48	>= 10	< 90	Yes	No
M	16 - 30	< 3	>= 48	>= 10	>= 90	Yes	No
M	< 16	< 3	>= 48	< 10	< 90	Yes	No
M	> 30	< 3	>= 48	>= 10	>= 90	No	Yes
M	> 30	>= 3	>= 48	>= 10	>= 90	Yes	No
F	< 16	>= 3	>= 48	< 10	< 90	Yes	No
M	16 - 30	< 3	>= 48	< 10	< 90	Yes	Yes

F	> 30	< 3	>= 48	>= 10	>= 90	No	Yes
F	> 30	>= 3	>= 48	< 10	>= 90	Yes	No
F	16 - 30	< 3	>= 48	< 10	>= 90	No	No
F	< 16	< 3	< 48	< 10	< 90	Yes	No
F	< 16	>= 3	>= 48	< 10	< 90	Yes	No
F	< 16	< 3	>= 48	< 10	< 90	No	Yes
F	> 30	>= 3	>= 48	< 10	< 90	No	No
F	< 16	< 3	>= 48	< 10	< 90	Yes	No
F	16 - 30	>= 3	>= 48	< 10	< 90	Yes	Yes
M	< 16	>= 3	>= 48	< 10	< 90	Yes	?

Based on the table above, where 20 data are used for training and determining the classification of testing data with data details in the form of:

Sex: Male  
 Age: 7 (<16)  
 Birth Weight: 3,2 (>=3)  
 Birth Length: 49 (>= 48)  
 Body Weight: 9 (<10)  
 Body Length: 68 (<90)  
 Exclusive breastfeeding (ASI) : yes

- It is known that the label class has two stunting classifications, Yes and No. Calculate the number of each label class divided by the total of all the data in the training data.  
 $P(Y = \text{Yes}) = 5 / 20 = 0,25$   
 $P(Y = \text{No}) = 15 / 20 = 0,75$
- Determine the probability of the same number of cases belonging to the same class.  
 $P(\text{Sex} = \text{Male} | Y = \text{Yes}) = 2/5 = 0,4$   
 $P(\text{Sex} = \text{Male} | Y = \text{No}) = 5/15 = 0,3$   
 $P(\text{Age} = <16 | Y = \text{Yes}) = 1/5 = 0,2$   
 $P(\text{Age} = <16 | Y = \text{No}) = 7/15 = 0,5$   
 $P(\text{Birth Weight} = >=3 | Y = \text{Yes}) = 4/5 = 0,8$   
 $P(\text{Birth Weight} = >=3 | Y = \text{No}) = 8/15 = 0,5$   
 $P(\text{Birth Length} = >=48 | Y = \text{Yes}) = 1/5 = 0,2$   
 $P(\text{Birth Length} = >=48 | Y = \text{No}) = 14/15 = 0,9$   
 $P(\text{Body Weight} = <10 | Y = \text{Yes}) = 3/5 = 0,6$   
 $P(\text{Body Weight} = <10 | Y = \text{No}) = 10/15 = 0,7$   
 $P(\text{Body Length} = <90 | Y = \text{Yes}) = 3/5 = 0,6$   
 $P(\text{Body Length} = <90 | Y = \text{No}) = 10/15 = 0,7$   
 $P(\text{Asi} = \text{Yes} | Y = \text{Yes}) = 2/5 = 0,4$   
 $P(\text{Asi} = \text{Yes} | Y = \text{No}) = 11/15 = 0,3$
- Calculating the total value of all variable results in each classification.  
 $P(\text{Sex} = \text{Male} | Y = \text{Yes}) * P(\text{Age} = <16 | Y = \text{Yes}) * P(\text{Birth Weight} = >=3 | Y = \text{Yes}) * P(\text{Birth Length} = >=48 | Y = \text{Yes}) * P(\text{Body Weight} = <10 | Y = \text{Yes}) * P(\text{Body Length} = <90 | Y = \text{Yes}) * P(\text{Asi} = \text{Yes} | Y = \text{Yes})$   
 $= 0,4 * 0,2 * 0,8 * 0,2 * 0,6 * 0,6 * 0,4$   
 $= 0,002$   
 $P(\text{Sex} = \text{Male} | Y = \text{No}) * P(\text{Age} = <16 | Y = \text{No}) * P(\text{Birth Weight} = >=3 | Y = \text{No}) * P(\text{Birth Length} = >=48 | Y = \text{No}) * P(\text{Body Weight} = <10 | Y = \text{No}) * P(\text{Body Length} = <90 | Y = \text{No}) * P(\text{Asi} = \text{Yes} | Y = \text{No})$   
 $= 0,3 * 0,5 * 0,5 * 0,9 * 0,7 * 0,7 * 0,3$   
 $= 0,010$
- Compare all class results in the classification.  
 From the results above, it can be seen that based on testing data that is not yet known, the highest probability value is obtained which results in a value of 0.010 so it can be concluded that the criteria in the testing data are categorised as No Stunting (No).

### 3.2 K-Nearest Neighbor Implementation

Data samples used for manual calculations on the implementation of the K-NN method are shown in Table 1 for training data samples and Table 2 for testing data samples. The next process is data transformation into a form that is more in accordance with the format that can facilitate the stunting prediction process. Data transformation can be seen in Table 5.

**Table 5.** Data Transformation

No.	Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	ASI Eksklusif	Stunting
1	1	56	2,9	50	11	90	1	No
2	1	20	3,3	49	11,1	80,5	2	No
3	2	4	2,8	48	6,5	63	2	No
4	1	14	2	49	7	71	1	No
5	2	32	3,2	49	11	88,7	1	No
6	2	30	2,3	50	12	90	1	No
7	2	2	2,9	49	8,5	74,2	1	No
8	2	33	2,5	49	10	91,5	2	Yes
9	2	33	3	50	15	96	1	No
10	1	15	4	51	5,9	58,3	1	No
11	2	16	2,4	48	6,6	76	1	Yes
12	1	47	2,5	49	10	91,5	2	Yes
13	1	31	3,3	49	9,1	90	1	No
14	1	25	2,6	49	9,8	90	2	No
15	1	4	2,7	47	5,8	69,6	1	No
16	1	12	3,6	50	8,2	70,5	1	No
17	1	3	2,8	48	6	54	2	Yes
18	1	55	3	49	8,5	81	2	No
19	1	12	2,9	49	5,8	69,5	1	No
20	1	19	3	52	7,7	57	1	Yes
21	2	7	3,2	49	9	68	1	?

In the manual calculation, a sample of 20 training data is used to calculate the Euclidean distance. The initial step in working on the manual calculation of the K-NN method is to determine the parameter value K, which in this study uses the parameter value  $k = 3$ . The next step is to calculate every testing data to all training data and then sort the results of manual calculations to calculate the distance using the Euclidean formula with the calculation of the formula in equation 2. After that, it is sorted according to the number of K that has been determined, and the results of the sequence are seen as the most dominant class. Considering the outcomes of this computation, which can be seen in Table 6, it can be determined that the most dominant class is No stunting.

**Table 6.** Calculation Results of K-NN Method Testing Data

No.	Distance Calculation Results (Euclidean Distance)	Nearest Order	Class (Stunting)
1	53,77	20	No
2	18,21	11	No
3	6,51	4	No
4	8,03	6	No
5	32,52	14	No
6	32,00	13	No
7	7,99	5	No
8	35,08	16	Yes
9	38,69	17	No

10	13,17	8	No
11	12,35	7	Yes
12	46,43	18	Yes
13	32,57	15	No
14	28,48	12	No
15	5,20	1	No
16	5,84	2	No
17	14,97	9	Yes
18	49,75	19	No
19	6,21	3	No
20	16,64	10	Yes

### 3.3 Performance Testing Evaluation

In the tests conducted in this study using the Naive Bayes and K-Nearest Neighbor algorithms, the results are displayed with a confusion matrix and performance testing consists of calculating the accuracy, precision, recall, and f1-score values in equations 3, 4, 5, and 6. The test results with a total of 1950 testing data implemented in the Naive Bayes method are addressed in the form of a confusion matrix table using the python program on Google Colab in Table 7 and the K-NN method in Table 8 as follows.

**Table 7.** Confusion Matrix of Naive Bayes Method Testing Results

	Actual Diagnosed Stunting	Actual Diagnosed No Stunting
Prediction of Classification Results Diagnosed with Stunting	TP = 766	FP = 310
Predicted Classification Result Diagnosed No Stunting	FN = 247	TN = 627

After obtaining the TP, FP, FN, and TN values, the accuracy, precision, recall, f1-score values can be calculated as follows:

$$Accuracy = \frac{766 + 627}{766 + 627 + 247 + 310} = 0.714358974358 = 71\%$$

$$Precision = \frac{766}{766 + 247} = 0.7118959107806692 = 71\%$$

$$Recall = \frac{766}{766 + 310} = 0.7561697926949654 = 76\%$$

$$F1 - Score = 2 \times \frac{0.7118 \times 0.7561}{0.7118 + 0.7561} = 0.7333652465294399 = 73\%$$



**Table 7.** Confusion Matrix of K-NN Method Testing Results

	Actual Diagnosed Stunting	Actual Diagnosed No Stunting
Prediction of Classification Results Diagnosed with Stunting	TP = 971	FP = 17
Predicted Classification Result Diagnosed No Stunting	FN = 42	TN = 920

After obtaining the TP, FP, FN, and TN values, the accuracy, precision, recall, f1-score values can be calculated as follows:

$$Accuracy = \frac{971 + 920}{971 + 920 + 42 + 17} = 0.9697435897435898 = 97\%$$

$$Precision = \frac{971}{971 + 17} = 0.9827935222672065 = 98\%$$

$$Recall = \frac{971}{971 + 42} = 0.9585389930898321 = 96\%$$

$$F1 - Score = 2 \times \frac{0.9827 \times 0.9585}{0.9827 + 0.9585} = 0.9705147426286856 = 97\%$$

The results of performance testing on the naïve bayes method obtained an accuracy value of 71%, precision 71%, recall 76%, and f1-score 73%. The performance results of the K-NN method using euclidean distance measurement with a value of k = 3 obtained the best performance value, namely accuracy of 97%, precision of 98%, recall 96%, f1-score 97%. Based on the results of the comparison of the Naïve Bayes and K-NN algorithms, the accuracy, precision, recall, and F1-Score of the two algorithms concluded that the K-NN algorithm is the best algorithm in this classification. The results of the K-NN algorithm classification are implemented in the interface to predict stunting so that it can be used by the wider community.



**Figure 3.** Stunting Prediction Interface

## 4 Conclusion

This research compares Naive Bayes and K-NN algorithms on stunting datasets. Based on the evaluation results of performance testing of the K-Nearest Neighbor and Naive Bayes methods in calculating accuracy, precision, recall, and f1-score. The results of performance testing on the naïve bayes method obtained performance values on 30% testing data are accuracy of 71%, precision 71%, recall 76%, and f1-score 73%. The performance results of the K-NN method using euclidean distance measurement obtained the best performance value on 30% testing data, namely accuracy of 97%, precision 98%, recall 96%, f1-score 97% at a value of  $k = 3$ . Based on the performance results of the comparison of the Naive Bayes and K-NN methods, it shows that the best classification method on the stunting dataset is the K-NN method because it gets better performance than the Naive Bayes method. Based on this research, further research is needed to expand the scope of algorithms and techniques used in stunting diagnosis. In further research, it can be done by adding variables to the dataset for testing in order to get a better accuracy value and also develop a website display to be more attractive and complete.

## References

- [1] Z. Zurhayati and N. Hidayah, "FAKTOR YANG BERTHUBUNGAN DENGAN KEJADIAN STUNTING PADA BALITA," *JOMIS (Journal of Midwifery Science)*, vol. 6, no. 1, pp. 1–10, Jan. 2022, doi: 10.36341/jomis.v6i1.1730.
- [2] Y. Rahma and D. Suhartini, "APLIKASI DIAGNOSA PENYAKIT STUNTING PADA BALITA," 2023. doi: 10.333365/jti.v17i2.2581.
- [3] T. Prasetya, I. Ali, C. L. Rohmat, and O. Nurdiawan, "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, vol. 5, no. 1, p. 93, Dec. 2020, doi: 10.51211/itbi.v5i1.1431.
- [4] L. Firgia, A. C. Nurcahyo, N. P, and M. Mira, "IMPLEMENTASI METODE NAÏVE BAYES SISTEM PAKAR MENDETEKSI STUNTING PADA BALITA BERBASIS WEBSITE," *Sebatik*, vol. 26, no. 2, pp. 543–548, Dec. 2022, doi: 10.46984/sebatik.v26i2.2117.
- [5] H. Indra Purwanto and S. Wibisono, "AHP-CBR Untuk Deteksi Dini Stunting Pada Balita Menggunakan Algoritma Similaritas KNN," *UNIVERSITAS BINA INSANI INFORMATICS FOR EDUCATORS AND PROFESSIONALS*, vol. 7, no. 1, pp. 64–73, 2022.
- [6] I. P. Putri, Terttiaavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 257–265, Jan. 2024, doi: 10.57152/malcom.v4i1.1078.
- [7] D. Gunawan and V. N. Andika, "Implementasi Teorema Bayes Pada Sistem Informasi Posyandu Dalam Mendeteksi Stunting Pada Balita," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 4, p. 692, Jun. 2023, doi: 10.30865/json.v4i4.6146.
- [8] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "APLIKASI NAÏVE BAYES CLASSIFIER (NBC) PADA KLASIFIKASI STATUS GIZI BALITA STUNTING DENGAN PENGUJIAN K-FOLD CROSS VALIDATION," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, May 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [9] N. A. Sholikhin and S. Atmojo, "Aplikasi Web Untuk Klasifikasi Stunting Pada Balita Dengan Menggunakan Metode K-Nearest Neighbours (Studi Kasus Posyandu Jawa Kidul)," *Journal of System Engineering and Technological Innovation (JISTI)*, vol. 1, no. 02, pp. 44–47, Oct. 2022, doi: 10.38156/jisti.v1i02.23.
- [10] A. Rozaq and A. J. Purnomo, "CLASSIFICATION OF STUNTING STATUS IN TODDLERS USING NAIVE BAYES METHOD IN THE CITY OF MADIUN BASED ON WEBSITE," *Jurnal Techno Nusa Mandiri*, vol. 19, no. 2, pp. 69–76, Sep. 2022, doi: 10.33480/techno.v19i2.3337.
- [11] Harliana and D. Anggraini, "Penerapan Algoritma Naïve Bayes Pada Klasifikasi Status Gizi Balita di Posyandu Desa Kalitengah," 2023.
- [12] O. Saeful Bachri and R. M. Herdian Bhakti, "Penentuan Status Stunting pada Anak dengan Menggunakan Algoritma KNN," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 02, pp. 130–137, Nov. 2021, doi: 10.46772/intech.v3i02.533.
- [13] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," *Indonesian Journal of Data and Science*, vol. 1, no. 3, Dec. 2020, doi: 10.33096/ijodas.v1i3.20.
- [14] Y. Saputra, N. Khoirunisa, and S. Arinal Haqq, "Classification of Health and Nutritional Status of Toddlers Using the Naïve Bayes Classification," *CoreID Journal*, vol. 1, no. 2, pp. 49–57, Jul. 2023, doi: 10.60005/coreid.v1i2.8.

- [15] A. Byna, *MONOGRAF ANALISIS KOMPARATIF MACHINE LEARNING UNTUK KLASIFIKASI KEJADIAN STUNTING*. CV. Pena Persada, 2020. doi: 10.31237/osf.io/dtcgz.