International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 6, No. 3, (2022), pp. 392-398

Implementation of Naïve Bayes Method in Classification of Nutritional Status of Toddlers at Pasar Ujungbatu Sosa Public Health Center

Heri Santoso[1], *Armansyah*[2], *Fitri Handayani Siregar*[3],
[1,2,3]*Computer Science, Universitas Islam Negeri Sumatera Utara, Indonesia*
email: herisantoso@uinsu.ac.id[1], *armansyah@uinsu.ac.id*[2],
fitrihandaayanii075@gmail.com[3]

Abstract

Health is a very important field in human life, there have been many studies or studies conducted in the health sector, for example nutrition problems. Nutrients are needed by humans to live healthy in order to be able to move and carry out daily activities. For the fulfillment of nutrition in toddlers is usually influenced by social and economic factors of the family. Toddlers' bodies need balanced nutrition to be able to grow and develop properly. The results of the SSGI in 2021 the stunting rate nationally decreased by 1.6% per year from 27.7% in 2019 to 24.4% in 2021. The data used in this study was 1114 toddler data. From the results of training and data testing consisting of 5 attributes, namely gender, age, weight, height, and upper arm circumference and there are 4 classes for class division, namely over nutrition, good nutrition, less nutrition and poor nutrition. And it is known that the accuracy by using 10 data samples gets an accuracy value of 80%. Thus, the system built using the Naive Bayes method is considered successful in classifying the nutritional status of children under five.

Keywords: *Naive Bayes, Classification, Nutritional Status of Toddlers, System*

1. Introduction

Every human being needs a balance of nutrients to produce a healthy body. Likewise, the nutritional status of toddlers should be paid more attention because toddler age is a very important growth age to maintain the pattern of development. For the fulfillment of nutrition in toddlers is usually influenced by social and economic factors of the family. Toddlers' bodies need balanced nutrition to be able to grow and develop properly, have a good nutritional status, have a normal or healthy weight, the body is not susceptible to infectious diseases, increased work productivity, and is protected from chronic diseases and premature death. Meanwhile, poor nutritional status, undernutrition and nutrition are more related to diet. Malnutrition or poor nutrition status usually has abnormal weight and height and unbalanced nutritional intake, which usually causes the body to be more susceptible to disease and if it occurs in the long term will result in stunting nutrition[1].

The results of the SSGI 2021 national stunting rate decreased by 1.6% per year from 27.7% in 2019 to 24.4% in 2021. There are several ways to determine the nutritional status of a child, one of which is the anthropometric method, this method compares weight with height, which will produce an analysis in the form of weight per age, height per age, and weight per height. which then to determine the type of nutrition look at the results of the analysis of weight per height[2]. Based on the problems above, the researcher will conduct a research entitled "Implementation of the Naïve Bayes Method in Classification of the Nutritional Status of Toddlers at the Ujungbatu Sosa Health Center". This study will use the Naive Bayes method to predict future opportunities according to previous experiences[3].

The purpose of this research is to determine the classification of the nutritional status of children under five using the Naive Bayes method. With the implementation of Naive

ISSN : 2580-7250

Bayes, it can help to classify the nutritional status of toddlers to determine the growth development of toddlers.

2. Research Methodology

2.1. Data Mining

Data mining or also known as Knowledge Discovery in Database (KDD) is an activity related to data collection, the use of historical data to find knowledge, information, regularities, patterns or correlations in large data[4]. Data mining has four basic functions, namely prediction functions, description functions, classification functions, and association functions[5].

2.2. Classification

Classification is used to determine examples or functions that compare or contrast concepts or data classes, with the aim of being able to estimate the class of an object whose label is unknown[6].

2.3. Naive Bayes

Naive Bayes builds and evaluates models very quickly and scales linearly on the number of predictions and rows. Naïve Bayes is also a description that represents each object class based on a probabilistic conclusion or recapitulation and finds the most likely class that is appropriate for each object that will be affected by its class from test objects that are based on attributes or variables whose values are known[7]. Here's the Naive Bayes formula:

$$P(H|X) = \frac{P(H|X) * P(H)}{P(X)} \quad (1)$$

Information :

H = speculation
X = events related to speculation
P(H) = possibility of speculation
P(X) = probability of event
P(E | H) = possible number of X in H
P(H | E) = possible number of H in X

2.4. Nutritional status of toddlers

Nutritional status is a state of the body as a result of the relationship between energy intake and protein and other essential nutrients with the state of body health. Toddlers' bodies need balanced nutrition to grow and develop properly. If nutritional intake is not met according to their needs, the golden age will be a critical period that will disrupt the process of child growth and development[8].

2.5. Confusion matrix

Confusion matrix is a method that is usually used to perform accuracy calculations on the concept of data mining or Decision Support Systems[9]. which is where the algorithm for classification is based on the number (count) of test records that are correctly and incorrectly predicted by the model.

3. Results and Discussion

The discussion in this section is how to implement the naive bayes method in the classification of the nutritional status of children under five and how the results of the classification of nutritional status of children under five by applying the naive bayes method. The other discussion is a detailed description of the manual calculation of the naïve bayes method. The stages of classifying the nutritional status of children under five

using the nave Bayes method are: 1.Data collection, 2.Data transformation, 3. Naive Bayes process, and 4. Testing.

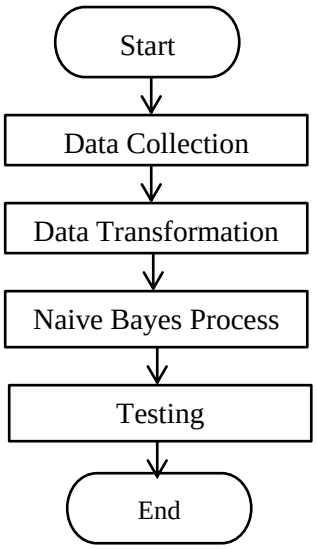

Figure 1. Stages of Classification the Naive Bayes method

3.1. Data collection

In the classification of the nutritional status of children under five, data collection and analysis of needs were carried out, the data obtained from the Pasar Ujungbatu Sosa Public Health Center. The total number of data on toddlers at the Pasar Ujungbatu Sosa Health Center is 1114 data or 1114 toddlers which are used as data sets. Then the data set is divided into training data as many as 1114 toddlers and testing data as data samples, namely as many as 10 toddlers taken from training data. The classicization of nutritional status of children with overnutrition was 18 toddlers, 877 good nutrition, 180 less nutrition and 39 poor nutrition.. The attributes used in this study were: gender, age, weight, height, and upper arm circumference. The classification of the nutritional status of children under five is divided into 4 classes of nutrition, namely: over nutrition, good nutrition, less nutrition and poor nutrition.

Table 1. Toddler Data at the Health Center

No.	Name	Gender	Age	Weight	Height	Arm Circumference	Nutritional Status
1	Siti Humairoh	W	31	13.3	81	15	Over Nutrition
2	M Rafael	M	36	15.7	92.1	17	Over Nutrition
3	Zay Rizky	M	26	12.8	84.5	15	Over Nutrition
4	Aulia R	W	44	12	91.9	15	Good Nutrition
5	M. Aidil	M	43	14.5	96.6	16	Good Nutrition
6	Ratu Aqilah Hsb	W	41	12.8	96.8	16	Good Nutrition
7	Taher Hasibuan	M	12	6.5	74.9	13	Poor Nutrition.
8	Adzra Tri Dianto	M	37	10.7	93.2	14	Poor Nutrition.
9	Delisa	W	24	9.9	98	13	Poor Nutrition.
10	Faiz Ananda	M	50	11.8	94.9	13	Less Nutrition
11	Yazid Az-Zikra	M	33	10	87.8	13	Less Nutrition
12	Diky Saputra	M	13	8.2	79	13	Less Nutrition
…..	……..	…	…	…	…	…	…….
1114	Aditia Pratama	M	15	6.4	64.4	12	?

3.2. Data Transformation

The following is explained in tabular form for a description of the above data transformation. The purpose of this data transformation is to simplify the process to find

out the results of data analysis. The data that has been collected from the analysis of data obtained from the puskesmas will be made a nave Bayes model in classifying the nutritional status of toddlers to make it easier to read and understand. And then the class probability and parameter probability in each class of nutritional status of children under five will be calculated or what is more often called the probability of occurrence. Below is made in the form of tables for data transformation.

Table 2. Toddler Data at the Health Center

Gender	Age (month)	Weight (kg)	Height (cm)	Arm (cm)
M = Man	Age1 = 0 - 6	WT1 = 0 – 2	HT1 = < 20	Arm1 = 11
W = Women	Age2 = 7 - 12	WT2 = 3 – 4	HT2 = 21 – 30	Arm2 = 12
	Age3 =13 – 18	WT3 = 5 – 6	HT3 = 31- 40	Arm3 = 13
	Age4 = 19 - 24	WT4 = 7 – 8	HT4 = 41 - 50	Arm4 = 14
	Age5 = 25 - 30	WT5 = 9 – 10	HT5 = 51 - 60	Arm5 = 15
	Age6 = 31 - 36	WT6 = 11 – 12	HT6 = 61 - 70	Arm6 = 16
	Age7 = 37 - 42	WT7 = 13 – 14	HT7 = 71 – 80	Arm7 = 17
	Age8 = 43 - 48	WT8 = 15 – 16	HT8 = 81 - 90	Arm8 = 18
	Age9 = 49 - 54	WT9 = 17 – 18	HT9 = 91 - 100	Arm9 = 19
	Age10 = 55 – 60	WT10 = 19 – 20	HT10 = >101	Arm10 = 20

3.3. Naive Bayes Process

Researchers conducted 10 experiments, namely with test data derived from the training data itself, and test data and training data contained in the data set. The following is an example of test data in the training data: there is a male toddler aged 15 months, weight 6.4 kg, height 64.4 cm and upper arm circumference 12 cm. What is the nutritional status of the toddler?

In the completion, a naive bayes calculation process is carried out, namely calculating the data that has been transformed to make it easier to calculate the classification of nutritional status of children under five, the transformed data is numeric data and the categories are transformed into all category data.

To find out the nutritional status of the toddlers above, the table in section 1. The Naïve Bayes model and the data must be transformed first. The transformation data can be seen in table 3 of the toddler data transformation. Below is the transformed toddler data:

Table 3. Transformation of Toddler Data

No	Name	Gender	Age	Weight	Height	Arm	Nutritional Status
1	Aditia Pratama	M	Age3	WT3	HT6	Arm2	?

Solution:

Table 4. Probability value of events in each class

	Over nutrition	Good nutrition	Less nutrition	Poor nutrition
Man	12/18	434/877	111/180	32/39
Age3	0/18	44/877	23/180	5/39
WT3	2/18	26/877	10/180	16/39
HT6	7/18	74/877	13/180	14/39
Arm2	1/18	44/877	12/180	15/39

a. Calculate Probability of Over Nutrition

= (NS = Overnutrition | Gender = M)*(NS = Overnutrition | Age = Age3)*(NS = Overnutrition |Weight = WT3) * (NS = Over Nutrition | Height = HT6) * (NS Over Nutrition | Arm = Arm2)*(Over Nutrient)

= (0.666) * (0) * (0.111) * (0.388) * (0.055)*(0.016)

= 0.

b. Calculate Probability of Good Nutrition
= (NS = Goodnutrition | Gender = M)*(NS = Goodnutrition | Age = Age3)*(NS = Goodnutrition |Weight = WT3) * (NS = Goodnutrition | Height = HT6) * (NS Good nutrition | Arm = Arm2)*(Over Nutrient)
= (0.494) * (0.050) * (0.029) * (0.084) * (0.050)*(0.787)
= 0.000002.

c. Calculate Probability of Lees Nutrition
= (NS = Leesnutrition | Gender = M)*(NS = Leesnutrition | Age = Age3)*(NS = Leesnutrition |Weight = WT3) * (NS = Leesnutrition | Height = HT6) * (NS Good nutrition | Arm = Arm2)*(Over Nutrient)
= (0.616) * (0.127) * (0.055) * (0.072) * (0.066)*(0.161)
= 0.000003.

d. Calculate Probability of Poor Nutrition
= (NS = Poornutrition | Gender = M)*(NS = Poornutrition | Age = Age3)*(NS = Poornutrition |Weight = WT3) * (NS = Poornutrition | Height = HT6) * (NS Poor nutrition | Arm = Arm2)*(Over Nutrient)
= (0.820) * (0.128) * (0.410) * (0.358) * (0.384)*(0.035)
= 0.000209

After all the results are known, it remains only to compare the results. The nutritional value of the higher value is the result of the classification of the nutritional status of the toddler. So the result of the nutritional status of the toddler is Malnutrition because the result of poor nutritional value is higher than all other nutritional values, namely 0.000209. The following is the display of the Nave Bayes calculation for the classification of the nutritional status of children under five using the system that was built. Prediction of the classification of nutritional status of children under five, in which the supporting attributes must be filled in to find out the classification class using the system. Below is how it looks on the system.

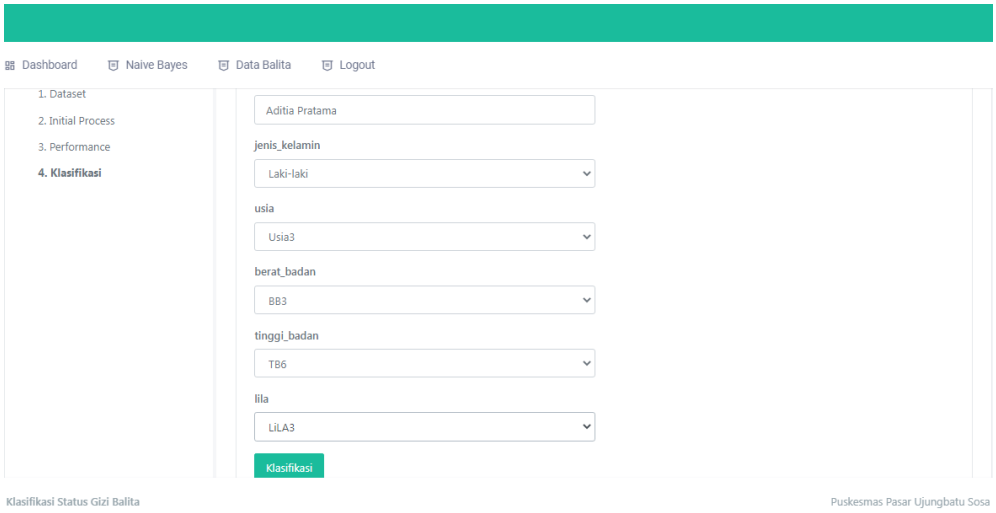

Figure 2. Classification of nutritional status of toddlers

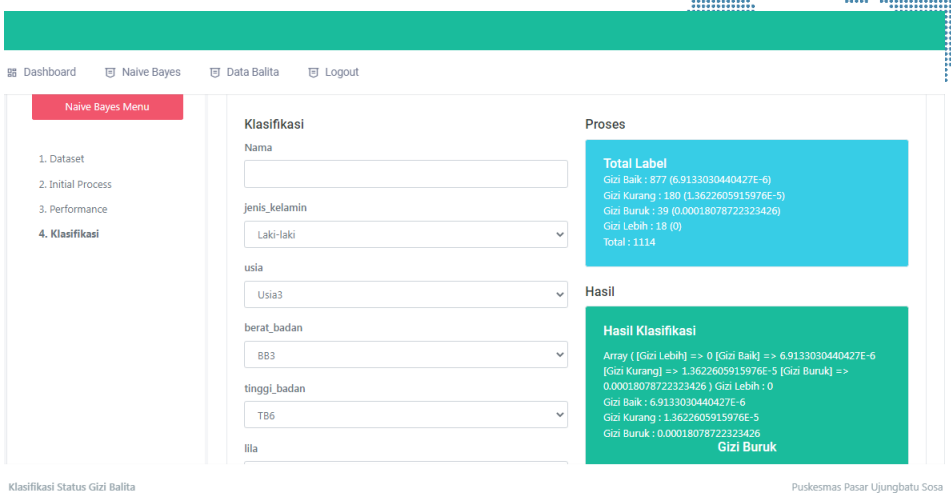

Figure 3. Completion of Classification Results

At this stage the naive bayes solution is carried out by taking the highest value of the probability label of the event probability.

3.4. Testing

Manual and system testing uses 10 toddler data and 5 parameters. Data testing is done by comparing the results of manual calculations and the results of system calculations using the Naïve Bayes method. The following are the results of testing the classification of nutritional status of children under five using the nave Bayes method, where the prediction results are compared with the original data on the nutritional status of children under five.

Table 5. Accuracy

No	Nutritional Status	Manual		System		Valid
1	Over nutrition	Good nutrition	0.00069	Good nutrition	0.00069	No
2	Good nutrition	Good nutrition	0.000356	Good nutrition	0.000355	Yes
3	Good nutrition	Good nutrition	0.001213	Good nutrition	0.001212	Yes
4	Less nutrition	Less nutrition	0.000254	Less nutrition	0.000254	Yes
5	Less nutrition	Less nutrition	8.09E-05	Less nutrition	8.21E-05	Yes
6	Good nutrition	Good nutrition	0.000104	Good nutrition	0.000104	Yes
7	Good nutrition	Good nutrition	0.001412	Good nutrition	0.001411	Yes
8	Less nutrition	Less nutrition	0.001401	Less nutrition	0.001401	Yes
9	Good nutrition	Over nutrition	0.000115	Over nutrition	0.000114	No
10	Poor nutrition	Poor nutrition	0.000209	Poor nutrition	0.000208	Yes

For system testing on test data, the results are the same as manual calculations for test data. Which is the correct data from the test data on system 8 and the wrong one is 2.
Calculate Accuracy:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\%$$

$$\text{Accuracy} = \frac{8}{10} \text{ x } 100\% = 80\%$$

Table 6. Configuration Matrix

		Prediction			
		Over nutrition	Good nutrition	Less nutrition	Poor nutrition
	Over nutrition	0	1	0	0
Actual	Good nutrition	1	4	0	0
	Less nutrition	0	0	3	0
	Poor nutrition	0	0	0	1

Above is a confution matrix table of 10 sample data tested using the nave Bayes model. And obtained an accuracy of 80%. With the formula for the correct number of data divided by the total number of data multiplied by 100%. The following will be summarized in the confution matrix test table. Based on the tests that have been carried out with the confusion matrix, it can be stated that the Nave Bayes method is a good method for classifying the nutritional status of toddlers.

Table 7. Confution Matrix Results

Precision	Recall	F1-Score	Accuracy
0.7	0.7	0.7	80%

From the test results, it is known that the model testing using the confusion matrix gets an accuracy value of 80%, although the accuracy results are not 100% but the accuracy results are classified as good accuracy. From the test results, it is known that the model test results from 10 correct sample data are 8 and the error is 2.

4. Conclusion

From the final project research that has been carried out, it can be concluded that: the classification of the nutritional status of toddlers at the Pasar Ujungbatu Sosa Health Center as many as 1114 toddlers can be classified properly and accurately with the implementation of the Naïve Bayeas method. In the 10 sample data tested, an accuracy of 80% was obtained by manual nave Bayes calculations or with the system built. So it was concluded that Naive Bayes could be used to classify the nutritional status of toddlers using 5 attributes. Applications for the classification of the nutritional status of children under five can be used by puskesmas officers to collect data on the nutritional status of children under five. The system can speed up the toddler nutrition classification process quickly and accurately.

References

[1] C. Litaay, "*Kebutuhan Gizi Seimbang"*, Zahir Publishing, Yogyakarta, (2021).
[2] Sulfianti, Sutrio, N. Vina, S. Eflariyani dan J. Dini , "*Penentuan Status Gizi"*, Yayasan Kita Menulis, Medan, (2021).
[3] Y.N.S. Sidiq, "*Metode Klasifikasi Menentukan Kenaikan Level UKM Bandung Timur Dengan Algoritma Naïve Bayes Pada Sistem JURAGAN Berbasis Komunitas"*, Cv.Kreatif Nusantara, Bandung, (2020).
[4] E. Buulolo, "*Data Mining Untuk Perguruan Tinggi"*, Penerbit Deepublish, Yogyakarta, (2020).
[5] W.Anjar, "*Data Mining : Algoritma dan Implemtasi"*, Yayasan Kita Menulis, Medan, (2020).
[6] Mulaab, "*Data Mining Konsep Dan Aplikasi"*, Media Nusa Creative, Malang, (2019).
[7] A. Muhammad, "*Data mining Algoritma dan Implementasi"*, Penerbit Andi, Yogyakarta, (2020).
[8] H. Tutik,"*Pendamping Gizi Pada Balita"*, DEEPUBLISH, Yogyakarta, (2019).
[9] A.Wulandari dan A. Sifa, "*Perbandingan Efektivitas Klasifikasi ALgoritma C.45 Dan Algoritma Naive Bayes Dengan Menggunakan Weka"*, LLPM, Jombang, (2019).