Classification of Toddler Nutrition Status with Anthropometry Calculation using Naïve Bayes Algorithm

1st Riris Aulya Putri
Electrical Departement
State University of Malang
Jalan Semarang No. 5, Malang
Indonesia
rhirisaulya@gmail.com

2nd Siti Sendari
Electrical Departement
State University of Malang
Jalan Semarang No. 5, Malang
Indonesia
siti.sendari.ft@um.ac.id

3th Triyanna Widiyaningtyas
Electrical Departement
State University of Malang
Jalan Semarang No. 5, Malang
Indonesia
triyannaw.ft@um.ac.id

***Abstract*— Growth and development process for toddler become the key point for growth and development in the next period. Nutrition needs must be given precisely, so children have a good nutrition status. Nutrition status monitoring for toddlers could be done with Anthropometry calculations, based on 3 index, weight for age (WFA), height for age (HFA), and weight for height (WFA). According to survey from Desa Tunjungtirto's Posyandu (Pos Pelayanan Terpadu/Integrated Service Post), they have not done the calculation of nutrition status, based on Anthropometry standards. To make it easier, we can use classification method. Classification is one of the data mining methods that find models or functions process that explain or differentiate the data class, its function is for predicting class from one unknown label object. Naïve Bayes is one of the popular classification algorithms and categorized as 10 best algorithms for data mining. The purpose of this research is to classified the toddlers's nutrition status based on 3 anthropometry index used Naïve Bayes algorithm. This classification will be tested with k-fold cross validation method to know the success of classification process. According to the results, can be concluded that the process of the toddler nutrition status classification, for each index, have 88% accuracy for WFA index, 64% accuracy for HFA index and 68% accuracy for WFH index.**

Keywords—Anthropometry, Classification, HFA (Height for Age), Naive Bayes, Toddler, WFA (Weight for Age), WFH (Weight for Height).

I. INTRODUCTION

Toddler is a term used to address children under 5 years old. That period of life is often called golden age which is an important period in child growth process since it will determine children's physical formation, psychic, and intelligence [1]. In this toddler period, children should be given balanced and precise nutritional needs so that they have decent nutritional status. Monitoring toddlers' nutritional status is a way to know the development of their growth. Anthropometry calculation is a way of examining the nutritional status that has ratified by World Health Organization (WHO) and Ministry of Health Republic of Indonesia [2]. Anthropometry is used to measure nutritional status deficiency in the form of decreased functional levels of body tissues, especially to determine protein imbalance, chronic energy deficiency, malnutrition, and showing the past nutritional history of a person [3].

Anthropometric index used to assess the nutritional status of children under five years are: 1) Weight for age (WFA) to find out if the toddler is malnutrition, lack of nutrition, having decent nutrition, or having excessive nutrition; 2) Body height for age (HFA) to determine whether or not the toddlers' height appropriate for their age, that it can be very short, short, normal and tall; 3) Body weight for height (WFH) to find out if the toddlers' body balanced between their weight and height, which it can be very thin, thin, normal and fat. Toddler's nutritional status threshold on each index is shown as in Ref. [2].

Although the calculation of nutritional status according to the standard of anthropometry has been settled and determined by the Ministry of Health to assess the nutritional status of toddlers, based on the survey conducted in Tunjungtirto village, the Maternal and Child Health Care in Tunjungtirto village still has not used the Anthropometry calculation. This happens since there are still many Maternal and Child Health Care cadres who do not understand Anthropometry, so that in Maternal and Child Health Care only use medical chart called KMS (Kartu Menuju Sehat/ Integrated Service Post) which assess the changing progress of body weight according to age. KMS provides only the information on changes in weight for age of the toddlers in each month, whether it experiences increase or decrease. This can be called as not specific, as toddlers are called healthy if they always gain weight and height which is appropriate for their age. According to these problems, current technological developments can be used for classification process to facilitate the determination of nutritional status of children under the Anthropometry standards.

Classification is a process to find models or functions that can explain and differentiate the class or concept from a data. The existing model is derived from an analysis of training data sets (data objects where classes of labels are known) [4]. Classification includes two processes on its job. The first process, classification will perform an analysis of data training to be able to form the classification rule which commonly referred to as the learning or training process. The second process, the classification will analyze the data testing that used to test the accuracy of classification rules that have been obtained from the previous process. To perform both classification process, there are four basic components of classification: class, attribute, training dataset, and testing dataset. The classification process can be done using algorithms such as Naïve Bayes. Naïve Bayes is one of the most popular classification algorithms and belongs to the top 10 algorithms in data mining [5]. In addition, Naïve Bayes

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 10,2025 at 10:20:20 UTC from IEEE Xplore. Restrictions apply.

often has both high accuracy and speed on various classification problems [6].

Several studies related to classification problem of toddlers' nutritional status using Naïve Bayes algorithm have similarities in the use of attributes such as gender, age, weight and height. However, the results of the classification output from these studies are different. Sari [7] which resulted in toddler's nutritional status become malnutrition, lack of nutrition, having decent nutrition, and having excessive nutrition with accuracy of 70%. Pramitarini et al [8] added the attributes of three comorbid (diarrhea, endemias, and pneumonia) that resulted in decent nutritional status, moderate nutrition, lack of nutrition, excessive nutrition, and malnutrition with an accuracy of 88%. Abdul [9] which resulted in the outcome of toddler's nutritional status based on 2 anthropometry index with accuracy result for WFA 47.99% and HFA 41.33%. Based on the descriptions of several previous studies, Naïve Bayes has proven to have a good classification process accuracy in Sari and Pramitarini research. Besides, previous studies have not assessed toddlers' nutritional status based on 3 anthropometric indexes: WFA, HFA and WFH. Therefore, this study will perform the calculation of nutritional status of toddler under three anthropometric index using Naive Bayes algorithm with the use of attributes of gender, age, weight, and height.

II. METHODOLOGY

Design of this research can be seen on Fig. 1.

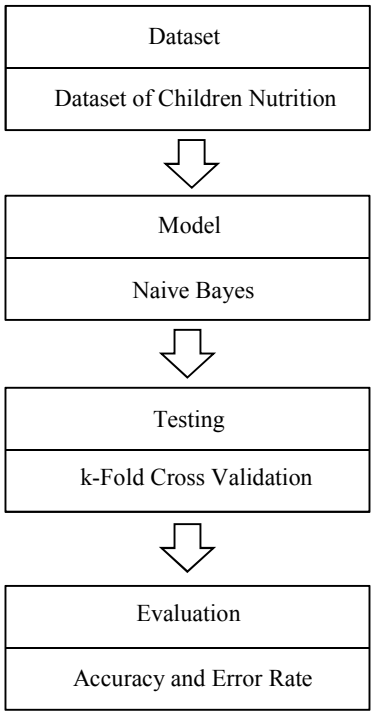

Fig.1. Design of this Research

A. Dataset

The result of data research got from three direct source "Posyandu" (Pos Pelayanan Terpadu/Integrated Health Post) by doing the research to the Posyandu in Desa Tunjungtirto. We got 100 Children contain of name, gender - Male (M), Female (F), age (in a month) weight count in kilogram, and height count in centimeter. Data found still in the process so it can't be used as a classification process, it is need to calculating the nutrient aspec of the baby based on Antropometri nutrient status. Based on it, we have three Antroppometri indexs and they are : WFA for the malnutrition status, lack of nutrition, having decent nutrition and having excessive nutrition. HFA in very short nutrient, short, normal, and high. WFH in a very thin status, thin, normal and fat.

B. Model

Naïve Bayes did the classification use probalilities and statistic method by using the idea of Bayes theory to predict a chance in a future based on the past eperience. Value of the probalility get from the frequence calculating and the combination of value based on the exist data collect [10]. Bayes theory is combined with *Naïve* that give three conclusions the condition between the attribute is free and high independent [4]. With the that result, it is applied Naïve Bayes theory as the same as in (1).

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

Based on Naïve Bayes theory if the i-attribute discrete , so $P(X_i|C)$ estimate as relative frequent from the sample which is having the X_i as the i-attribute in a C class. If the i-attribute continue or numeric , so $P(X_i|C)$ could be estimated using the function of Densitas Gauss as the same as in the (2).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad (2)$$

After we count based on Densitas Gauss, we can search the maximum probability constitute from the result of classification class of Naïve Bayes. The research classification process contain of two processes, they are training and testing process. Training process from the dataset will be read and then the data which is not numeric will be calculated the probability value, mean while the data which have numeric value will be calculated 'mean' and deviation standart. For the clearer classificaton prcess of Naïve Bayes it shows on Fig. 2.

C. Testing

Classification process need to do the testing process for knowing the result of the research. This testing classification use the cross validation method k-fold, because the acccuration of this method done by effective experiment and theoritical proof [11]. Testing technique become the recommendation if we would like to measure the accuration and error analysis from the classification result [12]. Total fold used in this research is 10 folds. From 100 datasets it will be divided in 10-folds to do the testing process. So, there will be 90 datasets training and 10 datasets testing in 10 folds.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 10,2025 at 10:20:20 UTC from IEEE Xplore. Restrictions apply.

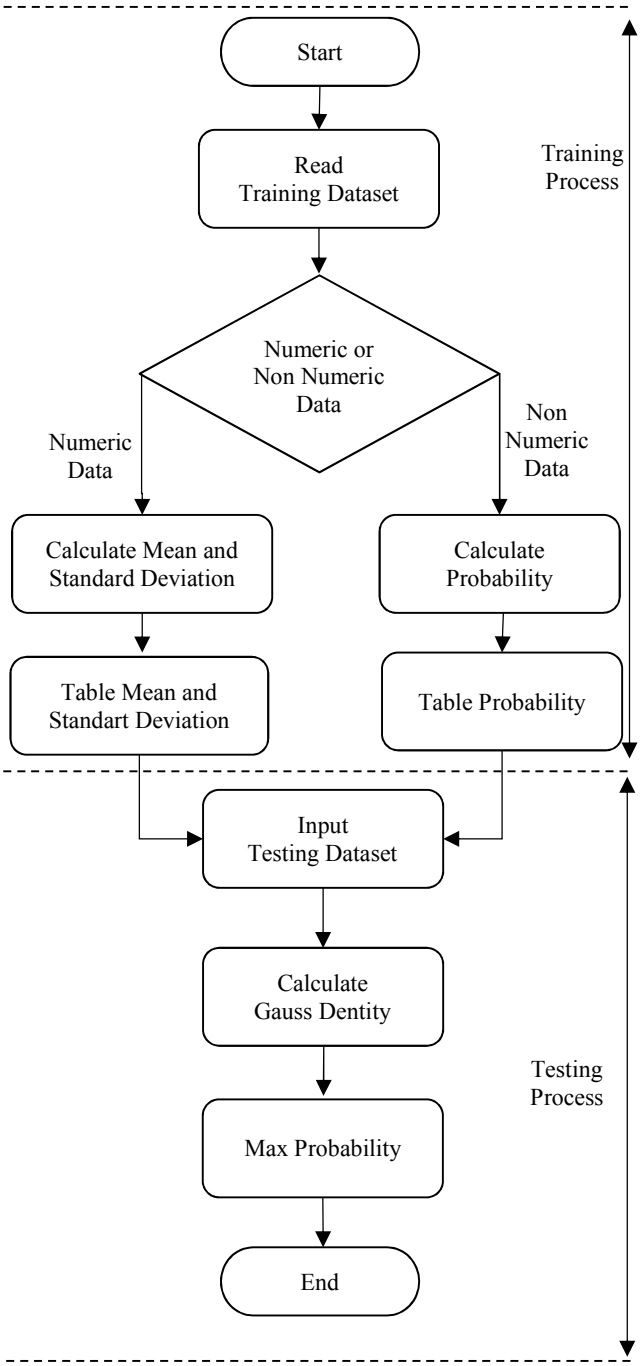

Fig. 2. Classification Process of Naive Bayes

D. Evaluation

The result of the test will be evaluated to get the accurate as in (4) and the error rate as in (5). After we know the accurate value and error rate in every fold, than the total result will be calculate by finding the 'mean' so that the final result of accurate and error rate from the alogarithm classification by Naive Bayes in the third chilren nutrient index status.

$$\text{Accuracy} = \frac{\sum \text{True perdiction}}{\sum \text{All Prediction}} \text{ x } 100\% \quad (4)$$

$$\text{Error rate} = \frac{\sum \text{False Prediction}}{\sum \text{All Prediction}} \text{ x } 100\% \quad (5)$$

III. Result and Discussion

A. Dataset

Composition research data could be seen in Table 1 until Table 9. In Table 1 shows total class and sex in every class for WFA index. In Table 2 shows the total class and sex in every class for HFA index. Table 3 tell us about total class and sex in every class for WFH index.

TABLE 1.
TOTAL CLASS AND SEX IN EVERY CLASS FOR WFA INDEX

Exp	Malnutrition	Lack	Decent	Excessive
Male	0	3	44	1
Female	1	4	44	3
Total	1	7	88	4

TABLE 2.
TOTAL CLASS AND SEX IN EVERY CLASS FOR HFA INDEX

Exp	Very Short	Short	Normal	Tall
Male	3	9	34	2
Female	8	12	30	2
Total	11	21	64	4

TABLE 3.
TOTAL CLASS AND SEX IN EVERY CLASS FOR WFH INDEX

Exp	Very Thin	Thin	Normal	Fat
Male	2	5	36	5
Female	1	5	38	8
Total	3	10	74	13

In Table 4 tell us about the data composition from the age attribute in WFA index. Table 5 shows about the data composistion from weight attribute in WFA index.

TABLE 4.
DATA COMPOSITION FROM THE AGE ATTRIBUTE IN WFA INDEX

Exp	Mal-nutrition	Lack	Decent	Excessive	Total
0-12 month	0	0	22	0	22
13-24 month	0	1	16	0	17
25-36 month	1	3	25	0	29
37-48 month	0	2	20	1	23
49-60 month	0	1	5	3	9
Total	1	7	88	4	100

TABLE 5.
DATA COMPOSITION FROM WEIGHT ATTRIBUTE IN WFA INDEX

Exp	Mal-nutrition	Lack	Decent	Excessive	Total
2-9 kg	1	2	24	0	27
9-18 kg	0	5	64	0	69
18-27 kg	0	0	0	4	4
Total	1	7	88	4	100

In Table 6 shows the data compsition from the age attribute in HFA index. Table 7 will tell us about the data composition from height attribute in HFA index.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 10,2025 at 10:20:20 UTC from IEEE Xplore. Restrictions apply.

TABLE 6.
DATA COMPOSITION FROM THE AGE ATTRIBUTE IN HFA INDEX

Exp	Very Short	Short	Normal	Tall	Total
0-12 month	2	5	15	0	22
13-24 month	4	3	10	2	19
25-36 month	2	7	18	0	27
37-48 month	3	3	15	2	23
49-60 month	0	3	6	0	9
Total	11	21	64	4	100

TABLE 7.
DATA COMPOSITION FROM HEIGHT ATTRIBUTE IN HFA INDEX

Exp	Very Short	Short	Normal	Tall	Total
40-68 cm	2	5	8	0	15
68-96 cm	9	16	40	2	67
96-124 cm	0	0	16	2	18
Total	11	21	64	4	100

In Table 8 refer to data composition from weight attribute in WFH index. In Table 9 indicate the data composition from height attribute in WFH index.

TABLE 8.
DATA COMPOSITION FROM WEIGHT ATTRIBUTE IN WFH INDEX

Exp	Very Thin	Thim	Normal	Fat	Total
2-9 kg	0	3	21	3	27
9-18 kg	3	7	52	5	67
18-27 kg	0	0	1	5	6
Total	3	10	74	13	100

TABLE 9.
DATA COMPOSITION FROM HEIGHT ATTRIBUTE IN WFH INDEX

Exp	Very Thin	Thim	Normal	Fat	Total
40-68 cm	0	0	10	5	15
68-96 cm	2	8	54	4	68
96-124 cm	1	2	10	4	17
Total	3	10	74	13	100

B. Testing

After we knew the nutrient status class from each children data then we will clasify them using algorithm classification by Naïve Bayes. Classification process will be done in every Antropometri index. This is needed to do since in every index have the different attribute and the nutrient status result. Testing classification process done by using k-fold cross validation method in amount of fold is 10-folds. From 100 datasets will be divied into 10 parts amount in 90 in every data training and 10 for data testing. For example, testing dataset in the first fold and the result for WFA index could be seen in Table 10 and Table 11, for HFA index could be seen in Table 12 and Table 13, and for the WFH index could be viewed in Table 14 and Table 15. The result of this test will be calculated accurate value with (4) and error rate value with (5).

TABLE 10.
FOLD 1 OF TESTING DATASET FOR WFA INDEX

No	Gender	Age	Weight	WFA
1	F	40	10	Lack
2	M	7	8	Decent
3	M	41	12	Decent
4	M	45	14	Decent
5	M	8	8	Decent
6	F	49	22	Excessive
7	F	15	11	Decent
8	F	29	12	Decent
9	F	41	10	Lack
10	F	40	18	Decent

TABLE 11.
RESULT TESTING DATASET FOLD 1 OF WFA INDEX

No	WFA	Naive Bayes	Correction
1	Lack	Decent	False
2	Decent	Decent	True
3	Decent	Decent	True
4	Decent	Decent	True
5	Decent	Decent	True
6	Excessive	Decent	False
7	Decent	Decent	True
8	Decent	Decent	True
9	Lack	Decent	False
10	Decent	Decent	True

TABLE 12.
FOLD 1 OF TESTING DATASET FOR HFA INDEX

No	Gender	Age	Height	HFA
1	F	40	75	Very Short
2	M	7	69	Normal
3	M	41	92	Normal
4	M	45	98	Normal
5	M	8	66	Short
6	F	49	104	Normal
7	F	15	87	Tall
8	F	29	87	Normal
9	F	41	85	Very Short
10	F	40	93	Normal

TABLE 13.
RESULT TESTING DATASET FOLD 1 OF HFA INDEX

No	HFA	Naive Bayes	Correction
1	Very Short	Normal	False
2	Normal	Normal	True
3	Normal	Normal	True
4	Normal	Normal	True
5	Short	Normal	False
6	Normal	Normal	True
7	Tall	Normal	False
8	Normal	Normal	True
9	Very Short	Normal	False
10	Normal	Normal	True

TABLE 14.
FOLD 1 OF TESTING DATASET FOR WFH INDEX

No	Gender	Weight	Height	WFH
1	F	10	75	Normal
2	M	8	69	Normal
3	M	12	92	Normal
4	M	14	98	Normal
5	M	8	66	Normal
6	F	22	104	Fat
7	F	11	87	Normal
8	F	12	87	Normal
9	F	10	85	Normal
10	F	18	93	Fat

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 10,2025 at 10:20:20 UTC from IEEE Xplore. Restrictions apply.

TABLE 15.
RESULT TESTING DATASET FOLD 1 OF WFH INDEX

No	WFH	Naive Bayes	Correction
1	Normal	Normal	True
2	Normal	Normal	True
3	Normal	Normal	True
4	Normal	Normal	True
5	Normal	Normal	True
6	Fat	Normal	False
7	Normal	Normal	True
8	Normal	Normal	True
9	Normal	Normal	True
10	Fat	Normal	False

C. Evaluation

Result from the test in Table 11 we determine the accuration for WFA test is 70% with error rate 30%. From the test in Table 13 accuration shows about HFA test is 60% with eror rate 40%. Based on Table 15 we knew that accuration show WFH in 80% with error rate 20%. The final result from this 10-fold test cross validation perform in Table 16 for WFA index, Table 17 for HFA index, and WFH index in Table 18.

TABLE 16.
FINAL RESULT FROM 10-FOLD TESTING FOR WFA INDEX

Fold	Accuracy	Error Rate
1	70%	30%
2	90%	10%
3	100%	0%
4	70%	30%
5	80%	20%
6	90%	10%
7	100%	0%
8	90%	10%
9	100%	0%
10	90%	10%
Mean	88%	12%

TABLE 17.
FINAL RESULT FROM 10-FOLD TESTING FOR HFA INDEX

Fold	Accuracy	Error Rate
1	60%	40%
2	80%	20%
3	50%	50%
4	50%	50%
5	50%	50%
6	90%	10%
7	60%	40%
8	70%	30%
9	50%	50%
10	80%	20%
Mean	64%	36%

TABLE 18.
FINAL RESULT FROM 10-FOLD TESTING FOR WFH INDEX

Fold	Accuracy	Error Rate
1	80%	20%
2	20%	80%
3	70%	30%
4	90%	10%
5	60%	40%
6	70%	30%
7	80%	20%
8	70%	30%
9	70%	30%
10	70%	30%
Mean	68%	32%

IV. CONCLUSION

Based on the result and discussion above we can conclude that, alogarithm of Naive Bayes could be used to classify the matter of determination children nutrient status based on Antropometry index. The best accuration test consist of WFA index with a value 88%, the next best accuration is WFH index with a value 68%, and the lowest accuration index is on HFA with a value 64%. The lowest error rate found that there is WFA index value 12%, then WFH in 32% , and HFA have the biggest value 36%. Meanwhile, in orde to get the best research, we suggest adding the training data with amount in every classess should be balances. Diffenetiate the nutrient status counting for the male and female baby. Or use other alogarithm classification like.

REFERENCES

[1] T. D. Pratiwi, Masrul, and E. Yerizel, "Hubungan Pola Asuh Ibu dengan Status Gizi Balita di Wilayah Kerja Puskesmas Belimbing Kota Padang," *J. Kesehat. Andalas*, vol. 5, no. 3, pp. 661–665, 2016.

[2] Kemenkes, *Keputusan Mentri Kesehatan Republik Indonesia Nomor: 1995/Menkes/ SK/XII/2010 tentang Standar Antropometri Penilaian Status Gizi Anak*. 2011.

[3] E. L. Amalia, H. S. Dachlan, B. Santoso, and G. Ag, "Integrasi Sistem Pakar dan Algoritma Genetika untuk Mengidentifikasi Status Gizi pada Balita," *J. EECCIS*, vol. 8, no. 1, pp. 1–6, 2014.

[4] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Second. San Francisco: Morgan Kaufmann, 2006.

[5] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.

[6] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *J. Teknol. Inf.*, vol. 18, no. 1, pp. 9–16, 2013.

[7] A. E. Sari, "Pengklasifikasian Kondisi Kesehatan Balita Menggunakan Naive Bayes Classification," *Simki-Techsain*, vol. 01, no. 07, pp. 2–11, 2017.

[8] W. Abdul, "Penentuan Status Gizi Dan Tinggi Balita Menggunakan Metode Naïve Bayes Classifier Berbasis Web," pp. 1–7.

[9] Y. Pramitarini, I. K. E. Purnama, and M. H. Purnomo, "Analisa Rekam Medis Untuk Menentukan Status Gizi Anak Balita Menggunakan Naive Bayes Classifier," *Pros. Semin. Nas. Manaj. Teknol. XVII*, pp. 1–8, 2013.

[10] T. R. Patil and S. S. Sherekar, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification," *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 256–262, 2013.

[11] S. K. Lidya, O. S. Sitompul, and S. Efendi, "Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," *Semin. Nas. Teknol. Inf. dan Komun.*, pp. 1–8, 2015.

[12] J. D. Rodríguez, A. Pérez, and J. A. Lozano, "Sensitivity Analysis of Kappa-Fold Cross Validation in Prediction Error Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 10,2025 at 10:20:20 UTC from IEEE Xplore. Restrictions apply.