Classification System Of Toddler Nutrition Status using Naïve Bayes Classifier Based on Z-Score Value and Anthropometry Index

Tiara Eka Putri¹, Ridho Taufiq Subagio², Kusnadi³, Petrus Sobiki⁴
¹Universitas Catur Insan Cendekia, Cirebon, Jawa Barat
²Universitas Catur Insan Cendekia, Cirebon, Jawa Barat
³Universitas Catur Insan Cendekia, Cirebon, Jawa Barat
⁴Universitas Catur Insan Cendekia, Cirebon, Jawa Barat

E-mail: tiara.ekaputri@cic.ac.id

Abstract. The system of nutritional status assessment for a toddler is crucial to monitor the growth of a toddler. This present study was carried out to build a classification system for determining the assessment of toddler nutritional status using naive bayes classifier based on value the z-score and index of Anthropometry. The data was used to perform classification include gender, age, height and weight. The data was calculated using the z-score to get nutritional status based on anthropometric indices-weight-for-age, height-for-age, weight-for-height for classified use Naive Bayes Classifier. This study used 225 data of toddlers. Testing system used 55 data as training and 175 data as testing with 100% accuracy. The results of this study was a system that could be used to perform classification of nutritional status based on a combination of three anthropometric indices using the Naive Bayes Classifier. Naive Bayes Classifier performed classification interpretation of nutritional toddler consisting of malnutrition, normal and over-nutrition. This study showed that the classification of nutritional status was on 175 data generating the highest percentage was malnutrition of 44.58%.

1. Introduction
Toddler growth could be monitored by doing an assessment of nutritional status on a regular basis by performing Anthropometry measurements and the calculation of z-score [1]. The parameters that are used in the measurement of Anthropometry for toddlers are the age, weight, and height. In addition, the effect of Gender in the measurement of Anthropometry is considerable [2]. According to these parameters, it generates the index of the Anthropometry; height for age, weight for height, weight for age [3][4] it will be interpreted by using the classification of the NCHS [5]. Anthropometric index and z-score can be used to assess the nutritional status of children under five in developing countries [6]. Currently, the assessment of the nutritional status is done vulnerably to human error and takes a long time. This problem can be overcome by creating a system that performs classification of nutritional toddler status automatically. To interpret the nutritional status based on the third nutritional indicators can be done with the classification technique [7].

In the technique of classification, there are several methods that can be used, which one of them is a naïve bayes classifier method [8]. This classifier method is one of the methods used in the classification technique as well as the machine learning based on data training. Methods of naïve bayes classifier is good system in accuracy and efficiency to classify comparing with other classification methods, does not require a lot of training data to perform classification [9]. Other studies investigated classification of nutritional status, using the method of fuzzy inference system generating accuracy of 81.15% [10]. The combination of decision trees and neural networks can produce an accuracy rate of 77.17% and an error rate of 22.8% [11]. In the world of health classification using naïve bayes became new innovations.
Some experiments were conducted to assess the performance of naïve bayes to do five times a test validation and generate value accuracy of 98.54% [12]. Naïve Bayes Classifier classification did not require large data to do the learning data training [13]. By using the Naïve Bayes Classifier, it proposes a classification system that performs classification interpretation of nutritional status based on a combination of three toddlers index Anthropometry

The basic concept of the naïve bayes classifier is used based on the bayes theorem, which is a theorem in the statistics to calculate the possibility of class from each group will go into the class with maximum value. Bayes theorem assumes all attributes independently on each variable class value [8]. Naïve bayes can classify and guess the current value based on the previous values [14]. Bayes theorem is a simplification of the advantages of the classic way that is fully integral to gain marginal [13]. Naïve bayes classifier can classify to distinguish the irrelevant attribute [15].

This present study is aimed to implement the Naïve Bayes Classifier method and build a system to classify the interpretation of the nutritional status of children under five based on the z-score and anthropometric index. The z-score is used to determine the nutritional status of children under five based on the anthropometric index. While the Naïve Bayes Classifier method determines the classification of nutritional status based on a combination of three anthropometric indices. Some research on nutritional status of toddlers has been carried out, only discussing data mining methods or z-score methods. In this present study, it combines data mining techniques with the z-score of the anthropometric index.

2. Research Methodology
2.1. Naïve Bayes Classifier
Naïve Bayes Classifier method is the methods in data mining to classify data using probability calculations [9]. The classification is divided into two phases consisting of learning and classifying. Phases of learning reads the data known to the class, while the classification phase is formed the data that will be tested. The naïve bayes is classified and predicted for the future values based on the previous value. The following is the general equation of bayes's theorem [8]:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{1}$$

Naïve bayes classifier steps are as follows [6]:
1. If $P(X)$ is independent then $P(X|C_i)P(C_i)$ are calculated for achieving maximum values. If the prior class probabilities previously unknown then it is assumed that the same class is $P(C_1)=P(C_2)=...=P(C_m)$ to calculate $P(X|C_i)$ then $P(X|C_i)P(C_i)$. Previous class probability can be predicted by $P(C_i|D) = |C_i,D| / |D|$, when $|C_i,D|$ were the amount of data training for $C_i$, which is in $D$. Class prior Probabilities can be calculated using the equation of:

$$P(C_i) = \frac{|C_i,D|}{|D|} \tag{2}$$

2. Because the data set has many attributes, it will be very difficult in to calculate $P(X|C_i)$. To minimize the calculation of $P(X|C_i)$, naïve bayes is a term creation of independent classes. If the values of these attributes are independent, then the equation become:

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$$

$$= P(x_1|C_i) \times P(x_2|C_i) \ldots \times P(x_n|C_i) \tag{3}$$
According to the above equation then it can easily estimate the probability $P(x_1|\mathcal{C}_i)$, $P(x_2|\mathcal{C}_i)$..., $P(x_n|\mathcal{C}_i)$ of training data.

2.2. Nutrition Status

Nutritional Anthropometry is associated with a variety of body dimensions and measurements of body composition from different levels of age and level of nutrition [2]. In the evaluation of toddler nutritional status, anthropometry presented in the form of indexes that are associated with other variables. Variables of Anthropometry are used to classify the nutritional status such as age, weight, height and gender. These variables produce an index of anthropometry, namely height for age, weight for height, weight for age[4]. The combination of these three indices produce an interpretation of anthropometry “malnutrition”, “normal” and “overnutrition” [5].

Z-score growth reference is used by WHO in 2005 as the standard Anthropometry for under five years. That standards are differentiated according to gender men and women. To determine the classification of nutritional status the calculation of z-score table reference based on z-score is done first. The z-score is an index of Anthropometry which are used internationally to determine the nutritional status and growth, expressed as force of the standard deviation (SD) of the inhabitants, it is used to calculate nutritional status using the standard reference Anthropometry by three anthropometry index[5]. Z-score can be computed by this equation:

$$z = \frac{x - \mu}{\sigma}$$  \hspace{1cm} (4)

Where $x$ is the measure value $\mu$ is the mean in the reference of the inhabitants and $\sigma$ is the standard deviation of the references inhabitants.

After getting of the z-score, the next step is to determine the category of nutritional status based on the toddler's threshold. Categories and nutritional status threshold index based on Anthropometry toddler height for age, weight for height, weight for age. The three indexes are combined to Anthropometry determining the interpretation of the nutritional status of toddlers. Therefore, nutritional status of interpretation consists of “low”, "normal" and "high “.

3. Result and Discussions

The data was used as data testing to be resolved using calculation the naive bayes classifier based on the value of z-score and anthropometry index. For example, a toddler data is described below:

| The Identity                  |
|------------------------------|
| Name            : Salapudin (no real name) | Height : 75 cm |
| Gender          : Male                 | Weight for age : Underfed           |
| Date of birth   : December-11-2015    | Height for age : Normal             |
| Date of weigh   : February-8-2017     | Weight for height : Very thin       |
| Age             : 14 month             | Interpretation : ??                 |
| Weight          : 7.2 kg              |                                  |

3.1 The Computation of Z-Score

To get the nutritional status based on an index of Anthropometry, the first step to do was to calculate the value of the z-score data training consisting of gender, weight, length and age. The attribute generated the index height for age, weight for height, weight for age. Z-score calculation of each index anthropometry is shown in the following in equation (4):

$$z \ (weight \ for \ age) = \frac{7.2 - 10.1}{10.1 - 9.0} = -2.63$$
The results of the calculation of the data obtained from toddlers who had x (weight) 7.2 in the lower μ value (median) in the table of weight for age. Then it was divided into the results of the subtraction between the values of μ (median) and the lower threshold value or -1 S.D. (σ), this was due to the child’s weight was lower than the median value. Appropriate threshold value then Anthropometry index z-score −2.63 on the index weight for age was included in the category of malnutrition.

3.2 The Computation of Naïve Bayes Classification

After having obtained the status of nutrition based on an index of anthropometry, the next step was to discover the classification status of the interpretation as the third combination based on toddler nutrition index anthropometry. The first trials used data 55 toddlers as data training. Calculation step for naive bayes classifier used the classification as follows:

a) Calculate the probability from several classes or categories of nutritional status of interpretations using the equation (2). Class interpretation of nutritional status (C) consisted of malnutrition, normal and over-nutrition. The total number of training data (D) were 55 data toddlers, with the probability of a malnutrition was 19, normal was 25, and overnutrition was 11. The calculation of the probabilities for each category on the nutritional status of interpretation was indicated by the equation (2):

\[ P(C_i) = \frac{19}{55} = 0.35 \]

b) Calculate the probability of each category of nutritional status of interpretations using equation (2) for each category (C) total training with discrete data (D) were 55 data. The number of data (D) from each category’s interpretation of nutrition were 19 data for malnutrition, 25 for normal data and 11 for the over-nutrition.

c) The calculation of the probabilities for each category discrete data on the nutritional status of interpretation anthropometry is shown in the following in equation (2). The results of the calculation of the probability of discrete data for each category is shown in Table 2.

| Attribute discrete | Malnutrition | Normal | Overnutrition |
|--------------------|--------------|--------|---------------|
| Boys               | 11/19        | 11/25  | 6/11          |
| Girls              | 8/19         | 14/25  | 5/11          |
| W-f-A (malnutrition) | 6/19    | 7/25   | 0/11          |
| W-f-A (underfed)   | 9/19         | 7/25   | 0/11          |
| W-f-A (normal)     | 4/19         | 9/25   | 4/11          |
| W-f-A (overnutrition) | 0/19    | 2/25   | 7/11          |
| H-f-A (very thin)  | 0/19         | 13/25  | 7/11          |
| H-f-A (thin)       | 0/19         | 1/25   | 1/11          |
| H-f-A (normal)     | 10/19        | 8/25   | 3/11          |
| H-f-A (obese)      | 9/19         | 3/25   | 0/11          |
| W-f-H (very short) | 16/19        | 0/25   | 0/11          |
| W-f-H (short)      | 3/19         | 0/25   | 0/11          |
| W-f-H (normal)     | 0/19         | 25/25  | 0/11          |
| W-f-H(tall)        | 0/19         | 0/25   | 11/11         |

d) Calculate mean (\( \bar{x} \)) and standard deviation (\( S^2 \)) from each continuous data consisting of age, weight and height for each category’s interpretation of nutrition. Next, it computed the gaussian distributions each continuous data consisting of age, weight and height for each category of nutritional status of interpretations. Calculation of the gaussian distribution of age with the value input \( x = 14 \), \( x = 7.2 \) and
\[ x = 75 \] for each of the categories of status of nutritional interpretation is shown in the following in equation (4):

\[
g = \frac{1}{\sqrt{2\pi x_{14.39}}} \exp\left(-\frac{(14-28.16)^2}{2(14.39)^2}\right) = 0.06
\]

The results of the calculation of probability gaussian for each category is shown in Table 3.

**Table 3.** The results of the calculation of Gaussian distributions each continuous data

| Attribute | Malnutrition | Normal | Overnutrition |
|-----------|--------------|--------|---------------|
| Age       | 0.06         | 0.07   | 0.09          |
| Weight    | 0.22         | 0.15   | 0.03          |
| Height    | 0.04         | 0.19   | 0.10          |

e) After calculation of probability of discrete data (gender, height for age, weight for height, weight for age, and the distribution of continuous data gauss (age (months), weight (kg) and height (cm)), then the calculation of the likehood of each category's interpretation of nutritional status can be calculated using equation (3).

\[
P(\text{malnutrition}) = 0.58 \times 0.47 \times 0.53 \times 0.84 \times 0.065 \times 0.22 \times 0.04 = 2.25 \times 10^{-5}
\]

\[
P(\text{normal}) = 0.44 \times 0.28 \times 0.32 \times 0 \times 0.07 \times 0.11 \times 0.45 = 0
\]

\[
P(\text{overnutrition}) = 0.55 \times 0 \times 0.27 \times 0 \times 0.09 \times 0.03 \times 0.10 \times 0.2 = 0
\]

f) The results of the calculation was investigated the normalization so the likehood value derived value totaled 1.

g) The last stage is the process of comparison of the value of the results of the calculation to get the greatest value as a result of the classification. Based on a comparison of the results of the interpretation of the category then the likehood that had the highest value was the likehood of malnutrition with the value 1.

3.3 Accuration Test

Accuracy testing was performed to find out the level of accuracy that is generated by performing some test data on a system using the calculation result of the z-score and naïve bayes classifier. Testing was done using data testing were 175 data. The test results from the 175 data for the results of the classification of NBC showed zero wrong data. The classification results from 175 data of under-five children resulted in 78 under-five or 44.58% into the category of malnutrition, 64 under-five or 36.58% included in the category of normal nutrition and 33 under-five or 18.86% included in the category of over nutrition.

4. Conclusion

The combination of the Naïve Bayes Classifier method and the z-score value was able to classify the interpretation of toddler nutritional status based on the anthropometric index. The system was able to undertake the training data to determine learning classification interpretation of nutritional status using naïve bayes classifier method toddler classifier. The calculation results showed that the case study of toddler study havd malnutrition status by having the highest percentage of 44.58%. The accuracy of Testing using 175 data toddler as test data had accuracy of 100%.
References

[1] M. I. Waly, “Nutritional assessment of preschool children using Z-Score analysis,” *Canad J Clin Nutr*, vol. 2, no. 2, pp. 50–59, 2014.

[2] D. B. Jelliffe and W. H. Organization, *The assessment of the nutritional status of the community (with special reference to field surveys in developing regions of the world)*. World Health Organization, 1966.

[3] T. Maken and L. R. Varte, “Nutritional status of children as indicated by z-scores of the Hmars: a tribe of NE India,” *Anthrocom Online J. Anthropol.*, vol. 8, no. 1, pp. 213–227, 2012.

[4] M. De Onis and M. Blössner, “The World Health Organization global database on child growth and malnutrition: methodology and applications,” *Int. J. Epidemiol.*, vol. 32, no. 4, pp. 518–526, 2003.

[5] M. De Onis, M. Blossner, and W. H. Organization, “WHO global database on child growth and malnutrition,” Geneva: World Health Organization, 1997.

[6] M. A. Hoque, M. A. Sayeed, M. R. Ahsan, M. A. Al Mamun, and F. Salim, “Nutritional Status among under-5 Children of a selected slum in Dhaka city,” *North. Int. Med. Coll. J.*, vol. 7, no. 2, pp. 143–145, 2016.

[7] X. Dezhi and G. U. Ganegoda, “Rule based classification to detect malnutrition in children,” *Int. J. Comput. Sci. Enggineering*, 2011.

[8] J. Han, M. Kamber, and J. Pei, “Data mining: concepts and techniques, Waltham, MA,” *Morgan Kaufman Publ.*, vol. 10, pp. 971–978, 2012.

[9] W. Hadi, Q. A. Al-Radaideh, and S. Alhawari, “Integrating associative rule-based classification with Naïve Bayes for text classification,” *Appl. Soft Comput.*, vol. 69, pp. 344–356, 2018.

[10] S. Suharjito, J. Jimmy, and A. S. Girsang, “Mobile decision support system to determine Toddler’s nutrition using fuzzy sugeno,” *Int. J. Electr. Comput. Eng.*, vol. 7, no. 6, pp. 3683–3691, 2017.

[11] D. Thangamani and P. Sudha, “Identification of malnutrition with use of supervised datamining techniques—decision trees and artificial neural networks,” *Int J Eng Comput Sci*, vol. 3, no. September 9, pp. 8236–8241, 2014.

[12] Z. Markos, “Predicting Under nutrition status of under-five children using data mining techniques: The Case of 2011 Ethiopian Demographic and Health Survey.” Addis Ababa University, 2013.

[13] M. Karabatak, “A new classifier for breast cancer detection based on Naïve Bayesian,” *Measurement*, vol. 72, pp. 32–36, 2015.

[14] T. R. Patil and S. Sherekar, “S.(2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification,” *Int. J. Comput. Sci. Appl. ISSN 0974*, vol. 1011.

[15] U. K. Pandey and S. Pal, “Data Mining: A prediction of performer or underperformer using classification,” *arXiv Prepr. arXiv:1104.4163*, 2011.