Nutrition Anthtropometric Status Model by Data Mining: Case Study in Palembang South Sumatera

Ermatita[1], Suci Destriatania[2], Yulnelly[3]

[1]*Information System, Computer Sciences Faculty, Sriwijaya University, Indonesia*
[2]*Faculty of Public Health, Sriwijaya University, Indonesia.*
[3]*Information System, Computer Sciences Faculty, UPN Veteran Jakarta, Indonesia*
ermatitaz@yahoo.com

ABSTRACT
Children under five nutrition status has been known as an important issue in every country. Monitoring of children nutritional status should be done to overcome nutrition problem in children under five years. Anthropometric is one way to determine nutrition status by measuring weight, height/ lenght and age. Nutrition status children under five years is classified by anthropometric indices based on latest World Health Organization (WHO) criteria: Weight for age z score (WAZ), height for age z score (HAZ), weight for height z score. Data mining is a technique in computer science to give information by prediction or classification. This study proposed decision tree model which combined with anthropometric method to monitoring nutritional status of children under five. Nutrition status of children which categorized by anthropometric indices is then classified with decision tree method to monitoring nutritional status by dividing into normal or chronic. This combination could gave quick information so that can be formulated solution to overcome.

Keywords: *Nutrition, anhropometric, data mining, decision tree.*

I. INTRODUCTION

Children growth and development is influenced by nutrition intake particularly in critical period. Evidence have indicated that inappropriate feeding practices can have profound consequences for the growth, development, and survival of infants and children, particularly in developing countries.

Nutritional status of children under five is determined based on age, weight and height or length which categorized by anthropometric indices weight for age z score (WAZ), height for age z score (HAZ), weight for height z score. Based on national current data, prevalence of "gizi buruk" in Indonesia is 19,6%. Prevalence of stunting in 2013 37,2% which is higher than 2010 (35,6%). From 37,2% is known prevalence of stunted is 18,2% and very stunted is 18% [1].

Nutritional status monitoring could be done by using technology of information. Data mining can be use in that process. Data mining model of nutritional status in using decision tree model. Zahra (2015) said that Classification techniques are implemented as a digital library search engine because it can help researchers to have the best response according to their demand. This model is developed by decision tree method to give an information about children nutrition status where this model is based on criteria of children nutritional status.

This study will give a model of children nutritional status so that children nutritional problem could be overcome earlier.

II. LITERATURE REVIEW

A. Data Mining Technique

Technique in data mining is called as method or algoritma that can be used to explore information from large data [2]. Many technique that has been called in data mining literature, for example: classification, neural network, genetic algorithm, etc is widely knon in information technology to manage an information. Data mining is categorized in many groups based on its function [3]:

1. Description
   Finding way to describe "pola" and possibility in data.
2. Estimation
   Model is developed by using compete record that show value of target variable an predictive value.
3. Prediction
   Prediction of value of result will be done in next step.Method that can be used in classification and estimation could be used also in prediction.

4. Classification
5. Clustering

   Clustering is set of record, observation and make of class of object. Cluster is different with classification that not have target variable.

6. Association

   The task of the association in data mining is to find attributes that appear in one time. In business commonly referred to as the market basket analysis. It is to find the goods purchased together and the goods that were never purchased simultaneously [3].

B. Modeling

Methods in data mining include the classification [4]. Classification establish a model that can be used to make predictions [5]. The resulting model is commonly called a classifier [6]. Model classifications widely used e.g Discriminant Analysis, Decision Tree, Neural Networks, Bayesian Networks, Support Vector Machine [2] [6] [8][10].

C. Nutritional status

Nutritional status is divided into five category: (1) children under five (2) addolescent (3) adult (4) chronic energy malnutrition (5) pregnant woman. Current data about nutritional staus is described in latest basic health surver report in 2013 based on province [1].

Almatsier (2004) said that nutritional status is a state as an aoutcome from nutrition intake and utilitu on nutritional substances. Nutrition status can be an expression of balance state in a form of variable. Jahari (2004) said that nutritional status is a condition as a result of balance state of intake and nutrition need for variety of biologic process

D. Nutritional Status Indicator

Based of Gibson (1990) in Indonesion Public Health, index weight for age z score gives a description of body mass. Body mass is sensitive with current changes particulary because of illness, appetite of decreasing of food intake (Riskesdas, 2013.)

E. Monitoring of Nutritional Status

Anthropometric is one way to determine nutrition status by measuring weight, height/ lenght and age. Nutrition status children under five years is classified by anthropometric indices based on latest World Health Organization (WHO) criteria: Weight for age z score (WAZ), height for age z score (HAZ), weight for height z score. World Health Organization-National Center for Health Statistics. (Indonesian public health)[7]

III. RESARCH METHODOLOGY

This research methodology is:

1. This study began identification of collecting data risk factor of the children nutritional status.
2. System need data nutritional status of children in the city of Palembang by anthropometry.
3. Further modeling of nutrient status classification is based on data mining method Decision Tree This Modelling star with:
   a. The stage of Data Preparation Phase at this stage of data collection and determination of the case data to be analysis
   b. Modelling phase: in this phase the collecting data are modelling by decision tree model. This phase aim tables are ready for classification process.
   c. Evaluation Phase This stage is part of the CRISP - DM process that includes checking whether a pattern or information
   d. The next stage is determining process of nutritional status of children based on anthropometric method Decision tree.

IV. RESULT AND DISCUSSION

This research was conducted by collecting data, data collected from several posyandu with permission from the city government of Palembang, obtained data toddlers conditions. Here is an example of data that has been retrieved from the neighborhood health center in the Palembang city.

Table 1. The condition of children under five years at posyandu

No	Name	Name of Parent		Date of birth	Age	Weight	Height
		Father	Mother				
1	Pratama	amawansya	Perawati	3/22/2012	3,1 th	15 kg	96,3 cm
2	Meisyah	hmad uslim	Fitriana	5/1/2012	3 th	1,6 kg	96,0 cm
3	Afgista Rahmad	elmianto	Tini	8/17/2011	3,9 th	13 kg	100,5 cm
4	Kaila Amanda	jang	Sam Sia	1/22/2011	4,3 th	3,8 kg	100,4 cm
5	Nayla	Agus man	Ria Reski	5/9/2012	3 th	0,2 kg	94 cm
6	Indah Permatasari	Bibit	Suswati	10/10/2011	3,7 th	4,4 kg	99,7 cm
7	Shyfa Salsabila	Panel	Suryati	8/20/2011	3,8 th	14 kg	101 cm
8	Nabil	Berkah	Ririn	9/27/2011	4,7 th	5,2 kg	106,5 cm
9	Lulu Adia	Alfarizi	Warningsih	5/3/2012	3 th	14 kg	97 cm
10	Anggi	Sapar	Een	11/11/2011	3,6 th	2,8 kg	99,5 cm
11	Saputra	Anton	Ponita	3/27/2012	3,1 th	2,8 kg	95,5 cm
12	Nabila	Basri	Ayu Wandira	3/22/2011	4,1 th	12 kg	103 cm
13	Dini	Romadi	Netti	8/5/2010	4,9 th	13 kg	109 cm
14	Pajar	Yadi	Iin	1/9/2012	3,3 th	1,5 kg	98,5 cm
15	Leni Aedelia	Agus Sasi	Ira	8/5/2010	4,9 th	4,2 kg	108 cm
16	Endi Andik	Asrol	Puput	10/25/2010	4,6 th	6,5 kg	107,3 cm
17	Nazila	Zulkarnain	Heni	12/26/2010	4,4 th	5,8 kg	104 cm
18	Bagas	Siswanto	Eliana	1/2/2011	4,4 th	5,5 kg	106,5 cm
19	Rahma	Azidin	Parida	8/18/2011	4,8 th	2,2 kg	105 cm
20	Citra	Yan	Tima	6/17/2010	4,9 th	15 kg	108 cm

The data in the table on the next in the analysis using data mining methods in the following:

A. Data Mining Analysis

a) Data Preparation Phase

Data from the activities of this child's weight, which will be the target data in this study was . While the attributes used as a determinant in the formation of the decision tree are: Location, weight for age z score (WAZ), height for age z score (HAZ), weight for height z score (WHZ).

In this case, the data selection are used as much as 50 data resulting from the process of weighing data ETL Posyandu toddler White Orchid and toddlers weighing data IHC clump which is then stored into the database kasus_status_gizi.

Furthermore, the results of the data analysis kasus_status_gizi classification on the nutritional status of children is GOOD, LESS NUTRITION or CHRONIC.

This study determined the classification to fill the target attribute value based on the likelihood that there will be a determinant of attributes. The number of possibilities that arise obtained from the attribute value attribute value Location x WAZ x value attribute WAZ = 2x3x3 = 18 possibilities for the classification of nutritional status that can be seen in the appendix.

Table 2 is the value of each attribute used to classify the nutritional status of children is GOOD, LESS NUTRITION or CHRONIC.Result of clasify show at Table 1. Data Preparation Process at Location Posyandu Anggrek Putih and Posyandu Serumpun

On location attributes, classified into 2 is based on data obtained activities of a child's weight.At WAZ, has a bad attribute value is less, and more normal. The classification is based on the Z-Score set by the Minister of Health. If:

• Less -3 SD to <-2 SD
• Good -2 SD and 2 SD
• More> 2 SD

At HAZ, have attribute values very thin, thin, normal, and overweight. The value can be enforced if:

• Thin SD -3 to <-2 SD
• Normal -2 SD and 2 SD
• Grease> 2 SD

Based on the conclusions of the assessment of nutritional status classification of nutritional status of children categorized as follows:

If the WAZ less and HAZ short while WHZ normal; conclusions state of child nutrition at this time either, but the child is experiencing a chronic problem, because the child's weight proportionate to height.

2. WAZ good; HAZ less; WHZ fat; conclusions of children suffered chronic malnutrition problems, and at this moment suffering from obesity (overweight) for more than proportional weight to height

3. WAZ less, HAZ short and WHZ thin; children are malnourished and chronic. This means that at the current state of child nutrition was not good and his past history was also not good.

4. WAZ good , HAZ normal and WHZ normal ; conclusions state child nutrition both in the present and past

5. WAZ less ; HAZ normal ; WHZ thin; conclusions of malnourished children who weight (lean) , the state of child nutrition in general is good , but weigh less proportional to Her height for the child's body tall

Table 2. Classification Of Nutrition Status

	WAZ	HAZ	WHZ
Bgood	Less	short	Normal
Kchronic	Good	less	fat
K malnutrition	Less	short	thin
Bgood	Good	Normal	Normal
K malnutrition	Less	Normal	thin

B. Modelling Phase

Preparation of process data that has been done , the resulting tables are ready for the process of classification , ie the data in the table can be found in appendix kasus_status_gizi .

Table kasus_status_gizi is a collection of data to be processes, the data used about 50 data . Then the training data consisting of each attribute above is formed into a decision tree using C4.5 algorithm . Which became the decision as to whether the good nutritional status , malnutrition or chronic .

To select an attribute as root , based on the highest gain value from existing attributes . To select the initial node then calculate the entropy of each attribute value of the data kasus_status_gizi .

$$I_{(S1,S2,S3)} = -\frac{42}{50}\log_2\frac{42}{50} - \frac{8}{50}\log_2\frac{8}{50} - \frac{0}{50}\log_2\frac{0}{50} = 0{,}63431$$

1. Atributte Location:

a. Posyandu Anggrek Putih

$$q_1 = -\frac{30}{30}\log_2\frac{30}{30} - \frac{0}{30}\log_2\frac{0}{30} - \frac{0}{30}\log_2\frac{0}{30} = 0$$

b. Posyandu Serumpun

$$q_2 = -\frac{12}{20}\log_2\frac{12}{20} - \frac{8}{20}\log_2\frac{8}{20} - \frac{0}{20}\log_2\frac{0}{20} = 0$$

$$E_{(lokasi)} = \frac{30}{50} * 0 + \frac{20}{50} = 0$$

$$Gain_{(lokasi)} = I_{(S1,S2,S3)} - E_{(lokasi)} = 0{,}63431 - 0 = 0{,}63431$$

Entropy value to attribute the resulting location is 0 , while the value of the gain or value effective for classifying attribute Location of 0.63431 .

A. For Attribute WAZ

a. Less

$$q_1 = -\frac{0}{3}\log_2\frac{0}{3} - \frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3} = 0$$

b. Good

$$q_2 = -\frac{41}{46}\log_2\frac{41}{46} - \frac{5}{46}\log_2\frac{5}{46} - \frac{0}{46}\log_2\frac{0}{46} = 0$$

c. Over

$$q_2 = -\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1} - \frac{0}{1}\log_2\frac{0}{1} = 0$$

$$E_{(BB/U)} = \frac{3}{50} * 0 + \frac{46}{50} * 0 + \frac{1}{50} * 0 = 0 \quad (9)$$

$$Gain_{(BB/U)} = I_{(S1,S2,S3)} - E_{(BB/U)} = 0{,}63431 - 0 = 0{,}63431$$

Result of entropy attribute value is 0 , while gain value for attribute classification **BB/U** is 0,63431.

B. Atribut WHZ

a. Thin

$$q_1 = -\frac{0}{8}\log_2\frac{0}{8} - \frac{8}{8}\log_2\frac{8}{8} - \frac{0}{8}\log_2\frac{0}{8} = 0$$

b. Normal

$$q_2 = -\frac{42}{42}\log_2\frac{42}{42} - \frac{0}{42}\log_2\frac{0}{42} - \frac{0}{42}\log_2\frac{0}{42} = 0$$

c. Fat

$$q_2 = -\frac{0}{0}\log_2\frac{0}{0} - \frac{0}{0}\log_2\frac{0}{0} - \frac{0}{0}\log_2\frac{0}{0} = 0$$

$$E_{(lBB/TB)} = \frac{8}{50} * 0 + \frac{42}{50} * 0 + \frac{0}{50} * 0 = 0$$

$$Gain_{(BB/TB)} = I_{(S1,S2,S3)} - E_{(BB/U)} = 0{,}63431 - 0 = 0{,}6343$$

Entropy value for the attribute BB / TB is generated is 0 , while the value of the gain or value effective to classify the attributes BB / TB at 0.63431 .

Table 3. Node 1

Node 1	Number of case (S)	good (S1)	less (S2)	Kronis (S3)	Entropy	Gain
Total	**50**	**42**	**8**	**0**	**0,63431**	
Location						0,63431
-Posyandu Anggrek Putih	30	30	0	0	0	
-Posyandu Serumpun	20	12	8	0	0	
BWAZ						0,63431
-less	3	0	3	0	0	
-good	46	41	5	0	0	
-more good	1	1	0	0	0	
BWHZ						0,63431
-Thin	8	0	8	0	0	
-Normal	42	42	0	0	0	
-Fat	0	0	0	0	0	

Results it can be seen that Tabel.3 Grain attribute value on all attributes are the same . In attribute Location and BB / TB each value has classified the case , but the majority of the data is BB / TB . Of the three attribute value , attribute value Petite already classifies cases into acute and attribute values Normal already classifies cases into Good But to attribute value Fat absence cases.

From these results can be described as a decision tree as shown in Figure 1 .

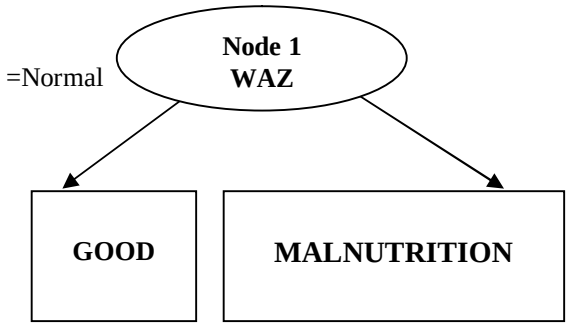

Figure 1. Decision Tree Node 1 Calculation Results

In Figure 1 , if the toddler has the index results Height Weight to Normal then the toddler has a Good nutritional status , so that the value of the attribute Normal stocked Good value . If the toddler has the index results Weight to Height Petite then the toddler has less nutritional status, so that the attribute value Petite fill filled Malnourished values.

C. Evaluation

After getting the decision tree of the modeling, phases of CRISP - DM process is the evaluation stage. This stage is part of the CRISP - DM process that includes checking whether a pattern or information found contradicted by the facts or hypotheses that existed before. In a decision tree technique, pattern or information generated from the data mining process is in the form of rules derived from a decision tree that has been used.

The following is a decision tree that has been formed in the data mining process as shown in Figure 2 below.

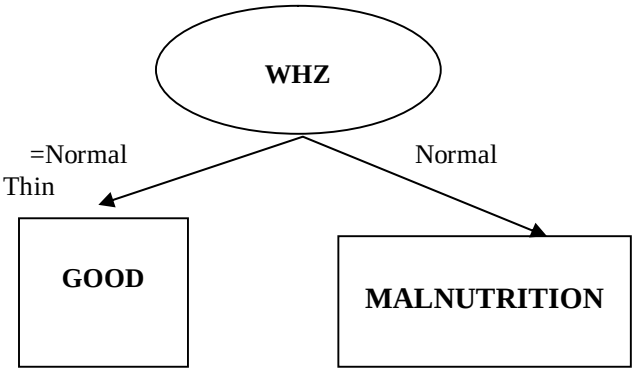

Figure 2. Decision Tree Node 1 Calculation Results

Results Figure 2. we will get rules
which can Classified nutritional status as shown in Table 4.

Table 4. Of Results Decission Tree

Rules	
1	**IF** WHZ = Normal **THEN** nutritional status of children under five = Good
2	**IF**WHZ= Kurus **THEN** nutritional status of children under five = less

After getting the rules of the decision tree , will be supplied test set . Supplied test set is to test how well the classification resulting from the data training . The results of the test set is supplied Correctly classified (the percentage of correct predictions) and incorrectly classified (percentage of incorrect predictions)

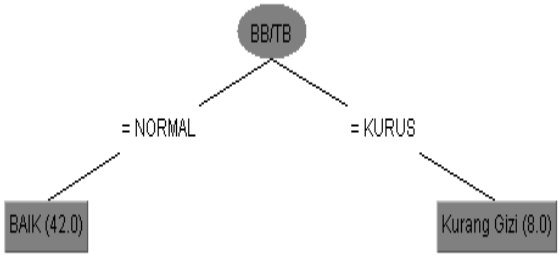

Figure 3. Test results using Weka Supplied

Results from the target of nutritional status of children of the above data is classified as malnourished 8 people and were classified as normal or good nutrition is 42 people

V. CONCLUSION

Modeling with data mining method Decision Tree that has been done to help the classification of nutritional status more quickly and with a wide range since it can use a lot of data. Implementation of the method Decision Tree helps monitoring toddler nutrition early, so that malnutrition in children can be addressed as early as possible .Data mining model result by Weka Application show malnourished 8 people and were classified as normal or good nutrition is 42 people.

REFERENCES

[1] Dep.Kes RI (2013), "*the Health Research*", Research and Development of the Ministry of Health

[2] Tan, P.N , Steinbach , M .. and Kumar , V.2006 , " *Intruduction To Data Mining* " , Pearson Education , Inc., Boston.

[3] Larose and Daniel , T. , 2005 , "*Discovering Knowledge in Data : An Introduction to Data Mining*" , John Willey & Sons.Inc.

[4] Ermatita and Ward , R , 2008, "*Survey classification in Data Mining Techniques*" , Proceedings / International Seminar The study results , University of Muhammadiyah . Purwokerto.

[5] Ermatita , Winarko , E and Ward , R , 2009, "*A Survey of Classificatiion Techniques and Application in Bionformatics*",

Proceedings of the International Conference on Creative Communication and Innovative Technology , 2009 (ICCIT - 2009) , pp.167-174 , Jakarta, Indonesias

[6] Santosa , B , 2007, *"Mechanical data utilization for business purposes : Theory and Application* " , Graha Science Yogyakarta

[7] Han , J and Kamber , M , 2001, "*Data Mining Consepts and Techniques*" , Morgan Kaufmann Publishers , San Francisco

[8] The monitoring of nutritional status, http://www.indonesian publichealth.com/2013/03/pemantauan-status-gizi.html

[9] Ibrahim . E, Balagatabi,H.N. 2015, "*Comparison of Decision Tree Methods in Classification of Researcher's Cognitive styles in Academic Environment***"**, Jurnal Teknologi Zahra, Roliana & Hossein / Jurnal Teknologi (Sciences & Engineering) 74:1 (2015),223–231 https://www.researchgate.net/publication/280632360_Comparison_of_Decision_Tree_Methods_in_Classification_of_Researcher%27s_Cognitive_styles_in_Academic_Environment.

[10] Saeed, S., Shaikh, A., Memon, M. A., & Naqvi, S. M. R. (2018). "*Impact of Data Mining Techniques to Analyze Health Care Data***"**. Journal of Medical Imaging and Health Informatics, *8*(4), 682-690.