Toddler Nutritional Status Classification Using C4.5 and Particle Swarm Optimization

Alwis Nazir[1]*, Amany Akhyar[2], Yusra[3], Elvia Budianita[4]

[1,2,3,4]Department of Informatics Engineering, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Abstract.

Purpose: This research was conducted to create a classification model in the form of the most optimal decision tree. Optimal in this case is the combination of parameters used that will produce the highest accuracy compared to other parameter combinations. From this best model, it will be used to predict the nutritional status class for the new data.

Methods/Study design/approach: The dataset used is from Nutritional Status Monitoring in 2017 in Riau Province, Indonesia. From the dataset, the Knowledge Discovery in Database (KDD) stages were carried out to build several classification models in the form of decision trees. The decision tree that has the highest accuracy will then be selected to predict the class for the new data. Predictions for new data (unclassified data) will be made in a web-based system.

Result/Findings: Particle Swarm Optimization is used to find optimal parameters. Before PSO is used, there are 213 parameters in the dataset that can be used to do classification. However, using many such parameters is time-consuming. After PSO is used, the optimal parameters found are the combination of 4 parameters, which can produce the most optimal decision tree. The 4 chosen parameters are gender, age (in months), height, and the way to measure the height (either stand up or lie down). The most optimal decision tree has an accuracy of 94.49%. From the most optimal decision tree, a web-based system was built to predict the class for new data (unclassified data).

Novelty/Originality/Value: Particle Swarm Optimization (PSO) is a method that can help to select the most optimal parameters, or in other words produce the highest classification accuracy. The combination of parameters selected has also been confirmed by the nutritionist. The prediction system has been declared feasible to be used by nutritionists through the User Acceptance Test (UAT).

INTRODUCTION

Nutritional status is a result of daily nutritional intake that describes an individual [1]. The nutritional status could assist and monitor the optimizing process of individual growth and development [2]. The toddler age (also known as the golden age) is an important period to optimize the growth and development of toddlers [3]. Toddler nutritional status is evaluated based on three indicators, namely weight-for-age (W/A), weight-for-height (W/H), and height-for-age (H/A) [4]. Based on H/A, the nutritional status of toddlers can be categorized into very short, short, normal, and high [5]. The combination of very short and short categories is known as stunting. The results of Nutritional Status Monitoring in 2017 show that the problem of H/A nutritional status (stunting) is the highest nutritional problem compared to W/A (underweight) and W/H (wasting) as shown in Figure 1 and Figure 2 [4].

Stunting is a condition of toddlers who are too short for their age due to inadequate nutrition and repeated bouts of infection during the first 1000 days of their life [6]. There are several causes of this condition, including poor parenting and a lack of consuming nutritious food due to the mother's lack of knowledge. Stunting has an impact on intelligence and physical development, more susceptibility to diseases, and at-risk of decreasing productivity in the future. Another broad impact of stunting is that it inhibits economic growth and increases poverty [7].

---

* Corresponding author.

Email addresses: alwis.nazir@uin-suska.ac.id (Nazir), amanyakhayar@uin-suska.ac.id (Akhyar),
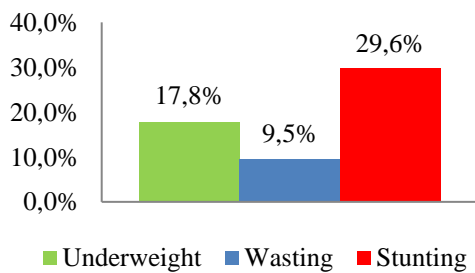yusra@uin-suska.ac.id (Yusra), elvia.budianita@uin-suska.ac.id (Budianita)

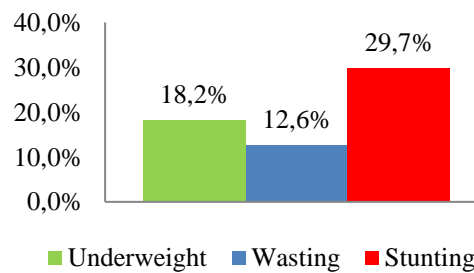Figure 1. Toddler nutritional problem in Indonesia

Figure 2. Toddler nutritional problem in Riau

One of the health development goals to reduce the number of stunting (also called the Stunting Intervention Program) is providing public nutrition education through the Public Nutrition Improvement Program. However, this program has not been done so much that the Stunting Intervention Program has not been effective [7]. One of the actions recommended by [8] to deal with stunting is to carry out preventive activities. Prevention activities that can be carried out include increasing identification, measurement, and understanding of stunting. This shows the importance of delivering information about stunting to the public. Thus, they can ensure their toddlers are not classified as stunting.

Based on the existing problems, the author makes an application that can help people know the H/A nutritional status of their toddlers and get information about stunting. With this application, the public does not need to wait for the Public Nutrition Improvement Program to be held. In building the application, the author applies the C4.5 algorithm [9] and Particle Swarm Optimization [10] on the data of Riau Province Toddlers in 2017 to classify the H/A nutritional status. The results of the algorithm will be applied to the web-based application, thus people can find out the H/A nutritional status of their toddlers and get information about stunting.

The C4.5 algorithm is known to exceed the Learning Vector Quantization (LVQ) algorithm in an average accuracy and has a faster processing time than the K-Nearest Neighbors (K-NN) method in classifying student abilities [11]. Another study that uses the C4.5 algorithm is also shown good performance for classifying the eligibility value of new prospective debtors [12] and exceeding Naïve Bayes in predicting the future behavior of the customers [13]. The C4.5 algorithm can also be optimized with PSO to select parameters/features [14]–[18]. This feature selection is done to avoid unrelated features that can reduce the performance of the classifier model. C4.5 algorithm optimized with PSO can exceed the accuracy of Support Vector Machine (SVM), Self-Organizing Map, Back Propagation Neural Network (BPNN), and C4.5 in classifying cancer [19]. Other studies also prove that the average accuracy of C4.5 combined with PSO in classifying 5 medical data sets can outperform the average accuracy of Logistic Regression, SVM, BPNN, and C4.5 [20]. Research related to the classification of nutritional status that has been done previously is the classification of W/A nutritional status of toddler using LVQ with the 80% accuracy (Euclidean) and 20% accuracy (Manhattan) [21]. Another study that classifies nutritional status by using Naïve Bayes reach 93.2% accuracy [22].

METHODS
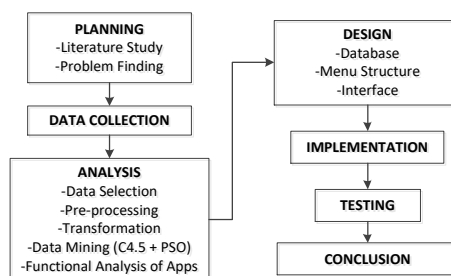The research methods of this study as in Figure 3:

Figure 3. Research methods

Planning
At this stage, the author studies the literature to find the problems that can be further researched. The problems found from this stage are the application of C4.5 algorithms and Particle Swarm Optimization (PSO) in classifying toddler nutritional status based on height-for-age indicator (H/A), the parameters influence in toddler's data on the toddler H/A nutritional status, and application development that can predict the H/A nutritional status of toddler and provide information about stunting.

Data Collection
At this stage, data collection is conducted. The data collected is toddler data from the result of Nutritional Status Monitoring in 2017 in Riau Province, Indonesia. The data obtained consisted of 3.961 records and 213 parameters.

Analysis
At this stage, the author analyzes the data of toddlers starting from the data selection, pre-processing, transformation, and data mining [23]. The best accuracy (from the optimum parameters or decision tree) would be found in the data mining step.
1) Data selection. This is the step to select data and parameters to be used. In this study, 3.961 records and 19 parameters were selected. The selected parameters are expected to have a relation with H/A nutritional status of toddler. This step is done by a discussion with the nutritionist.

Table 1. Selected parameters

No	Parameter	Information
1	ID	Individual identity
2	B1R5	Village classification
3	B2R3	Number of household members (ART)
4	B2R4	Number of the toddler in the family
5	B5R3	Iodized salt consumption
6	B4R4	Gender
7	UMUR_BULAN	Age (in months)
8	B09B	Age when given drinks or food other than breast milk for the first time
9	B21	Children 6-11 months old, have received Vitamin A (blue capsule)
10	B22	For children >12 months old, have received Vitamin A (red and/or blue capsule)
11	B23	Have a Card Towards Healthy (KMS)
12	B26A	The number of weight measurements in the last 6 months
13	B27	Have been treated for malnutrition
14	F01B	Weight
15	F02B	Height
16	F02C	Way to measure height
17	ZS_BBU	Z-Score of W/A
18	ZS_BBTB	Z-Score of W/H
19	KAT_TBU	Nutritional status of H/A

2) Pre-processing. This is the step to overcome missing values, duplicate data, inconsistent data, and outliers found.
   a) Missing Values
      Deletion of records containing the missing value is done. There are 317 records that contain missing value. After deletion, the remaining records are 3.644.
   b) Removing Duplicate
      This step is done by checking the similarity of values in the ID parameter. After checking, no duplicate data is found.
   c) Inconsistent Data
      Data should only be filled with numbers. However, there is a value of "#NULL!". The action taken at this step is to change the "#NULL!" value to "99".
   d) Outlier
      Deletion of records containing the outliers is done, thus the remaining records are 3.629. There are 15 records that contain outlier.
At the end of this step, the total record that has been deleted is 332 records. Thus, the percentage of data loss is 8.3%.

3) Transformation. This is the step to change the data format to the format needed for data mining. For example, the data mining step in this research would form decision tree based on the parameters. Thus,

the data should contain only the necessary parameter (that has relation to the class target). However, the ID parameter is not necessary for the classification process (or data mining step) as it is only needed in the pre-processing step (removing duplicate step). Thus, the ID parameter could be deleted. The total of parameters after deletion is 18 parameters.

4) Data mining. This is the step to apply the C4.5 algorithm and PSO to the data using Matlab. The following is a flowchart that describes the flow of the C4.5 algorithm and PSO in Matlab [5], [9], [10].
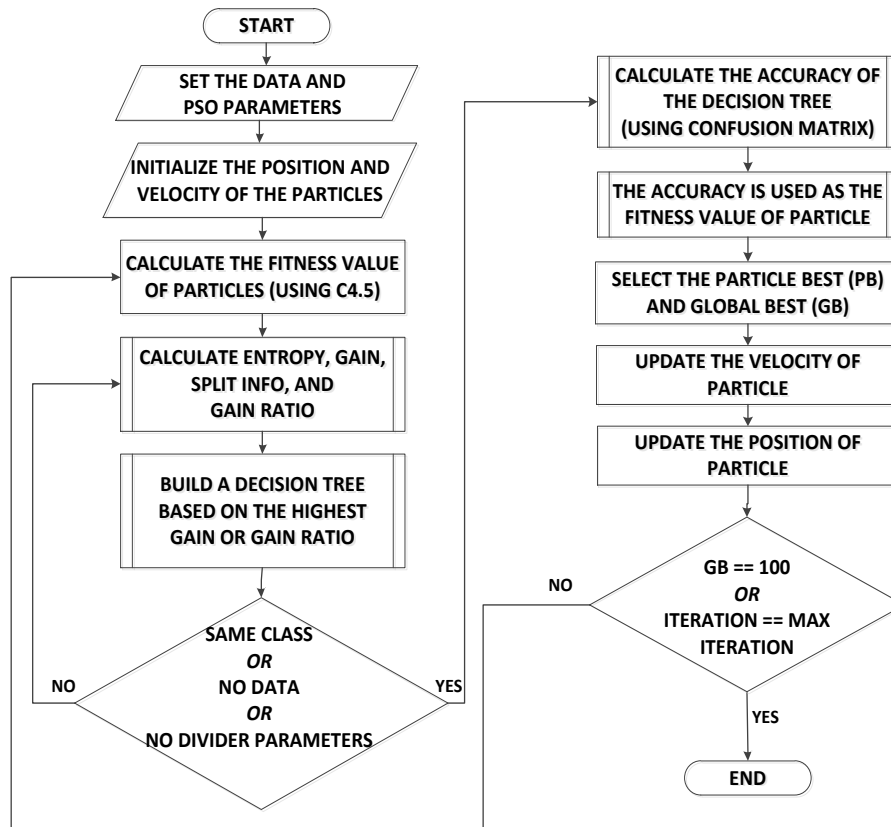
Figure 4. Flowchart of C4.5 algorithm and PSO

Based on Figure 4, the detailed process of applying the C4.5 algorithm and PSO is as follows:
1) Set the ratio of training data and test data to 90:10 and PSO parameters:
   a) Number of particles: 17
   b) Maximum iteration: 100
   c) w (inertia weight): 0,4
   d) c_1 (learning rates of particle): 2
   e) c_2 (learning rates between particles): 2
   f) v_{min} (minimum velocity): -4
   g) v_{max} (maximum velocity): 4
   h) r_1 and r_2 (random number between 0 and 1): 0,9624 and 0,9711
   The maximum iteration, w, c_1, c_2, v_{min}, and v_{max} parameters are set based on the previous research that uses C4.5+PSO on a similar domain, which is the medical problem [20]. The number of particles is set as many as the number of parameters (excluding the target or class parameters). The r_1 and r_2 are obtained from the random function in Matlab (built-in function).
2) Initialize particle position and velocity randomly. The particle position is a 17-bit binary number that represents each parameter in the data (the 18th parameter becomes the target class, thus only 17 parameters left). If the binary bit is 1, then the parameter will be used in the classification process (calculation of particle's fitness value).

3) Calculate the fitness value of each particle using C4.5 algorithm. The parameter to be used in this process (to be used in forming decision tree) is based on the particle's position. C4.5 algorithm uses gain and gain ratio to form decision trees using the following formula [11]:

$$E(s) = -\sum_{i=1}^{m} p(\omega_i|s) log_2 p(\omega_i|s) \tag{1}$$

$$Gain(s,j) = E(s) - \sum_{i=1}^{n} p(v_i|s) \times E(s_i) \tag{2}$$

$$SplitInfo(s,j) = -\sum_{i=1}^{k} p(v_i|s) \, log_2 p(v_i|s) \tag{3}$$

$$GainRatio(s,j) = \frac{Gain(s,j)}{SplitInfo(s,j)} \tag{4}$$

Explanation:

$p(\omega_i|s)$	=	proportion of i class in data processed at node s
m	=	number of different values in the data
$p(v_i|s)$	=	the proportion of the v value appears in the class in node s
$E(s_i)$	=	entropy of the v value's composition of the jth class in the i data at node s
n	=	number of different values in the node
k	=	number of divider

The particle fitness value is obtained based on the accuracy of the decision tree that produced by C4.5 algorithm. The following is an example of confusion matrix that is used to calculate accuracy [11].

Table 2. Example of confusion matrix for 3 classes

f_{ij}		Prediction Class (J)		
		Class=1	Class=2	Class=3
Actual	**Class=1**	f_{11}	f_{12}	f_{13}
Class (i)	**Class=2**	f_{21}	f_{22}	f_{23}
	Class=3	f_{31}	f_{32}	f_{33}

$$Accuracy = \frac{f_{11}+f_{22}+f_{33}}{f_{11}+f_{12}+f_{13}+f_{21}+f_{22}+f_{23}+f_{31}+f_{32}+f_{33}} \tag{5}$$

Select particle best (pb) and global best (gb) from particle's fitness values. Pb is the best fitness reached by each particle, while gb is the best fitness among its neighbors.
4) Update the velocity (v) and position (x) of particle based on pb and gb. Here is the formula used [19].

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 r_1 (pb_{id}^{old} - x_{id}^{old}) + c_2 r_2 (gb_d^{old} - x_{id}^{old}) \tag{6}$$

$$x_{id}^{new} = \begin{cases} 1, & if \; sigmoid(v_{id}^{new}) > ran[0,1] \\ 0, & otherwise \end{cases} \tag{7}$$

Explanation:

v^{new}	= new velocity of particle
i	= i^{th} particle
d	= d^{th} dimension
w	= Inertia weight
v^{old}	= current velocity of particle
c_1	= learning rates of particle
c_2	= learning rates between particles
r_1, r_2	= random number between 0 and 1
pb	= the position of the particles with the best fitness reached by each
gb	= the position of the particle with the best fitness among its neighbors

x^{old} = current position of particle

x^{new} = new position of particle

$$sigmoid(v_{id}^{new}) = \frac{1}{1 + exp^{-v_{id}^{new}}}$$

$ran[0,1]$ = random number between 0 and 1

5) Repeat steps 3-5 until the maximum iteration reached or 100% accuracy (gb is 100) obtained. Whichever reaches first.

The result of the data mining step is the fastest particle to achieve the best accuracy. The decision tree of that particle (the best decision tree) will be applied to The Application of Toddler Nutritional Status. The flowchart will be used to analyzing the flow of the application as in Figure 5.
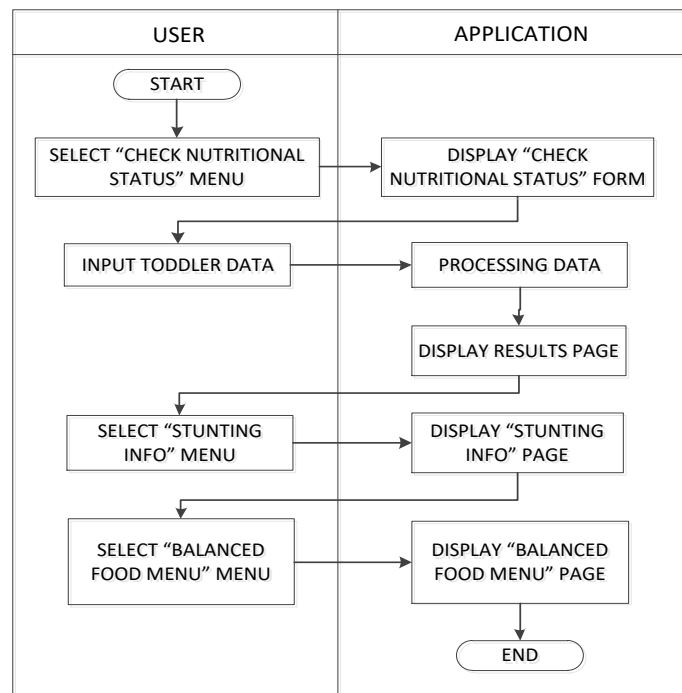
Figure 5. Flowchart of the application of toddler nutritional status

An explanation of Figure 5 is as follows:
1) The user selects the "check nutritional status" menu on the main page.
2) The application displays a form to check the H/A nutritional status of toddlers.
3) The user inputs the toddler's data into the form.
4) The application processes toddler data using a decision tree.
5) The application displays the results (H/A nutritional status of the toddler).
6) The user selects the "stunting info" menu.
7) The application displays a page about stunting info, such as the definition of stunting, causes, effects, and recommendations (one of them is a balanced food menu for toddlers).
8) The user selects the "balanced food menu" menu.
9) The application displays a page about a balanced food menu which is recommended by a nutritionist.

Design
At this stage, the author designs the concept of database, menu structure, and interface of the application. The database design will be used as a reference when creating the data storage of the application. The menu structure and interface design will be made, thus the application has a display that meets the user-friendly and usefulness aspects.

Implementation

At this stage, application development is done based on the analysis and design that has been done before. The interpretation of the best decision tree is done by changing the decision tree details from Matlab into an if-else rule in the code line for web-based applications. This if-else rule will be used to predict H/A nutritional status of toddlers.

Testing

The percentage of testing data being used is 10% or 363 records. Confusion matrix is used for the accuracy testing of C4.5 algorithm and PSO. Application (web-based) testing is done by two methods, namely black box and User Acceptance Test (UAT).

RESULT AND DISCUSSION

The result of the data mining step (C4.5 and PSO) after the 100th iteration is done is shown below.

Table 3. The 100th Iteration result

Iteration When pb Reached	i^{th} Particle	pb	Position When pb Reached	Amount of Parameters Used
14	11	94.48	00001101000101100	6
15	7	94.49	00001100000101100	5
17	6	94.49	00001101110101000	7
17	13	94.49	00011100000101100	6
26	9	94.49	00001101100001100	6
35	4	94.48	00001101000101100	6
41	5	94.49	00001101110101100	8
42	12	94.49	00011100000001100	5
43	14	93.07	00001101110111011	10
63	16	94.49	00001100100101100	6
64	15	94.49	00001101110001000	6
70	17	93.11	00011101100111100	9
77	3	92.01	00001100100111101	8
77	8	94.49	00011100000101100	6
81	2	94.49	00011101100101100	8
84	1	92.54	00001100010100011	6
93	10	94.48	00001101010101100	7

Based on the table above, gb (biggest pb or the best accuracy) is 94.49%. The fastest particle to achieve the best accuracy is 7th particle. The 7th particle obtained such accuracy only after the 15th iteration. The other obtained the same accuracy (6th particle, 13th particle, etc) after the 15th iteration. The 7th particle obtain 94.49% accuracy when the particle position is 00001100000101100 (5 parameters used). These parameters are:

1) B4R4 is gender.
2) UMUR_BULAN is age (in months).
3) B27 is about history of being treated for malnutrition (yes or no).
4) F02B is height.
5) F02C is way to measure height (stand or lie down).

In forming a decision tree (using C4.5 algorithm), the B27 parameter has never been chosen as a tree branch. Only 4 parameters are used, namely B4R4, UMUR_BULAN, F02B, and F02C. These parameters are the most important and influential in determining H/A nutritional status of toddlers, as the PSO works as feature selection to avoid unrelated parameters that can reduce the classifier performance.

Based on the interview with nutritionist, the influence of gender parameters is men are higher than women. The influence of age parameters is that the more you age, the higher your body height. The influence of height parameters is the more optimal height, the more possible for your H/A nutritional status is included

in the normal category. The influence of way to measure the height, namely the height measured while standing will be 0.7 cm shorter than lying down.

The following is the detailed calculation of the best accuracy reached using the confusion matrix.

Table 4. Confusion matrix

f_{ij}		Prediction Class (j)		
		Class=1	Class=2	Class=3
Actual Class (i)	Class=1	33	4	2
	Class=2	3	56	8
	Class=3	0	3	254

$$\text{Accuracy} = \frac{33+56+254}{33+4+2+3+56+8+0+3+254} = \frac{343}{363} = 0.9449$$

From the previous calculations, it can be concluded that the C4.5 algorithm and PSO have an accuracy of 94.49% in classifying H/A nutritional status of toddler using the 90:10 ratio. Based on the testing results of web-based applications (The Application of Toddler Nutritional Status) using the black box method, the application can run as expected and in accordance with analysis and design.

From the results of UAT testing by five nutritionists (Figure 6), it can be concluded that the application features are in accordance with the science of nutrition. These features are Nutritional Status Checking (QN2), Stunting Info (QN4), and Balanced Food Menu Info (QN5). It is also considered easier in predicting toddler nutritional status (stunting) compared to the manual method, which is manually counting Z-Score (QN3). This application also has an interface that is easy to understand (QN1), the features provided are quite in accordance with the needs (QN6), and overall the application is classified as good by nutritionists (QN7).

From the results of UAT testing by five general public (Figure 7), it can be concluded that the public can obtain the information about stunting (QP4) and the information about balanced nutritious food menu (QP5) from the application. The general public can also check the toddler nutritional status (stunting) (QP2) more easily than by using manual methods, which is visiting health center (QP3). This application also has an interface that is easy to understand (QP1), the features provided are in accordance with the needs (QP6), and overall the application is classified as good by the general public (QP7).
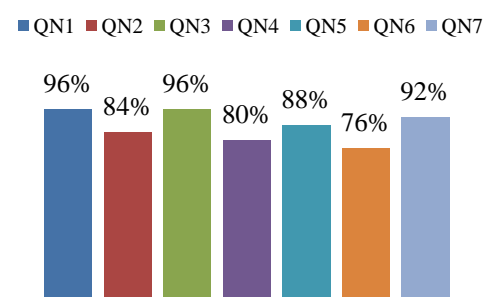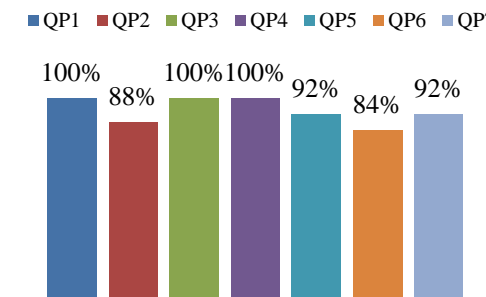
Figure 6. UAT testing by nutritionists

Figure 7. UAT testing by general public

CONCLUSION

Based on the research methodology that has been done, several conclusions can be drawn as follows. The accuracy of C4.5 algorithm and PSO in classifying H/A nutritional status of toddlers is 94.49% with the 90:10 ratio and 4 parameters. Of the 19 parameters selected, there are 4 parameters that could reach 94.49% accuracy, namely the gender, the age (in months), the height, and the way to measure the height. These parameters are the most important and influential in determining H/A nutritional status of toddlers, as the PSO works as feature selection to avoid unrelated parameters that can reduce the classifier performance. The influence of gender parameters is men are higher than women. The influence of age parameters is that the more you age, the higher your body height. The influence of height parameters is the more optimal height, the more possible for your H/A nutritional status is included in the normal category. The influence

of way to measure the height, namely the height measured while standing will be 0.7 cm shorter than lying down. The Application of Toddler Nutritional Status can predict the H/A nutritional status of toddlers according to the decision tree produced by C4.5 algorithm and PSO. This application can also provide information about stunting and recommendations for a balanced nutritious food menu for toddlers. Based on the black box testing, the application has been running as expected and in accordance with analysis and design. From the User Acceptance Test (UAT) with the nutritionists, the application is in accordance with the science of nutrition and is considered easier in predicting H/A nutritional status of toddlers compared to the manual method (Z-Score). This application also has an interface that is easy to understand, the features provided are quite in accordance with the needs, and classified as good. From the User Acceptance Test (UAT) with the general public, the application can predict H/A nutritional status of toddlers and it is easier than the manual method (visiting a health center) and the application can also provide information about stunting. This application also has an interface that is easy to understand, the features provided are in accordance with the needs, and are classified as good.

REFERENCES

[1]   Kementerian Kesehatan Republik Indonesia, "Bahan Ajar Gizi : Penilaian Status Gizi," 2017. http://bppsdmk.kemkes.go.id/pusdiksdmk/wp-content/uploads/2017/11/PENILAIAN-STATUS-GIZI-FINAL-SC.pdf (accessed Apr. 13, 2018).

[2]   O. Dwienda, L. Maita, E. M. Saputri, and R. Yulviana, *Buku Ajar Asuhan Kebidanan Neonatus, Bayi/ Balita dan Anak Prasekolah untuk Para Bidan*. Yogyakarta: Deepublish, 2015.

[3]   S. Kustiyati and L. Firrahmawati, "Optimalisasi tumbuh kembang balita dengan peningkatan pengetahuan ibu tentang gizi balita," *GASTER*, vol. 12, no. 2, pp. 86–98, 2015.

[4]   Kementerian Kesehatan Republik Indonesia, *Buku Saku Pemantauan Status Gizi Tahun 2017*. Jakarta Selatan: Direktorat Gizi Masyarakat, 2017.

[5]   Menteri Kesehatan Republik Indonesia, "Keputusan Menteri Kesehatan Republik Indonesia Nomor 1995/MENKES/SK/XII/2010 tentang Standar Antropometri Penilaian Status Gizi Anak," 2010. http://gizi.depkes.go.id/wp-content/uploads/2011/11/buku-sk-antropometri-2010.pdf (accessed Feb. 27, 2018).

[6]   World Health Organization, "WHA Global Nutrition Targets 2025 : Stunting Policy Brief," 2014. https://www.who.int/nutrition/topics/globaltargets_stunting_policybrief.pdf (accessed Feb. 28, 2018).

[7]   Sekretariat Wakil Presiden Republik Indonesia, "100 Kabupaten/Kota Prioritas untuk Intervensi Anak Kerdil (Stunting)," 2017. https://www.tnp2k.go.id/images/uploads/downloads/Buku Ringkasan Stunting.pdf (accessed Feb. 06, 2018).

[8]   World Health Organization, "STUNTING," 2014. http://www.who.int/nutrition/global-target-2025/infographic_stunting.pdf (accessed Mar. 01, 2018).

[9]   J. R. Quinlan, "Improved use of continuous attributes in C4.5," *J. Artiicial Intell. Res.*, vol. 4, pp. 77–90, 1996.

[10] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995.

[11] M. Fakhrurrifqi and R. Wardoyo, "Perbandingan algoritma nearest neighbour, C4.5 dan LVQ untuk klasifikasi kemampuan mahasiswa," *Indones. J. Comput. Cybern. Syst.*, vol. 7, no. 2, pp. 145–154, 2013.

[12] B. Hermanto and A. SN, "Klasifikasi nilai kelayakan calon debitur baru menggunakan decision tree C4.5," *Indones. J. Comput. Cybern. Syst.*, vol. 11, no. 1, pp. 43–54, 2017.

[13] M. Sadikin and F. Alfiandi, "Comparative study of classification method on customer candidate data to predict its potential risk," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, pp. 4763–4771, 2018.

[14] M. A. Muslim, S. H. Rukmana, E. Sugiharti, B. Prasetiyo, and S. Alimah, "Optimization of C4.5 algorithm-based particle swarm optimization for breast cancer diagnosis," *J. Phys. Conf. Ser.*, vol. 983, no. 1, 2018.

[15] R. S. Rohman, R. A. Saputra, and D. A. Firmansaha, "Komparasi algoritma C4.5 berbasis pso dan ga untuk diagnosa penyakit stroke," *J. Comput. Eng. Syst. Sci.*, vol. 5, no. 1, pp. 155–161, 2020.

[16] A. C. Pradana, Adiwijaya, and A. Aditsania, "Implementing binary particle swarm optimization and C4.5 decision tree for cancer detection based on microarray data classification," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, p. 012014, 2019.

[17] D. M. B. Tarigan, D. Palupi Rini, and Sukemi, "Particle swarm optimization – based on decision tree of C4.5 algorithm for Upper Respiratory Tract Infections (URTI) prediction," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, p. 012077, 2019.

[18]   I. Ariyati, S. Rosyida, K. Ramanda, V. Riyanto, S. Faizah, and Ridwansyah, "Optimization of the decision tree algorithm used particle swarm optimization in the selection of digital payments," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, p. 012090, 2020.

[19]   K. Chen, K. Wang, K. Wang, and M. Angelia, "Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data," *Appl. Soft Comput. J.*, vol. 24, pp. 773–780, 2014.

[20]   M. Tsai, K. Chen, C. Su, and H. Lin, "An application of PSO algorithm and decision tree for medical problem," in *2nd International Conference on Intelligent Computational Systems*, pp. 124–126, 2012.

[21]   E. Budianita and Novriyanto, "Klasifikasi status gizi balita berdasarkan indikator antropometri berat badan menurut umur menggunakan learning vector quantization," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 7, pp. 213–220, 2015.

[22]   S. Kusumadewi, "Klasifikasi status gizi menggunakan naive bayesian classification," *CommIT (Communication Inf. Technol. J.*, vol. 3, no. 1, pp. 6–11, 2009.

[23]   U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.

[24]   E. Prasetyo, *Data Mining - Mengolah Data menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI, 2014.