

Prediction of Nutritional Status of Toddlers Using C4.5 Algorithm

Sri Mujiyono¹, Novan Syaiful²

¹Faculty of Computers and education, Universitas Ngudi Waluyo, Indonesia

²Department of Informatics Engineering, Universitas Ngudi Waluyo, Indonesia

ARTICLE INFO

Article history:

Received Jun 30, 2023

Revised Jul 18, 2023

Accepted Jul 30, 2023

Keywords:

C.45 algorithm

Classification

Predictions

RapidMiner

Stunting

ABSTRACT

In developing countries, chronic malnutrition leads to stunting. Stunting will become a public health problem in Indonesia resulting in a decline in the quality of Indonesia's human resources in the future. In the modernization era, determining the nutritional status of toddlers can be simplified automatically. The author's C4.5 algorithm was used to classify and predict the nutritional status of toddlers in Bringin sub-district, in addition to testing using the system that the author built, the author also conducted manual testing and testing using RapidMiner as a comparison. Based on the analysis of the results of 186 datasets tested using this system, predictions were made of toddlers with male gender categories, very underweight and height predicted that the toddlers were in the malnutrition category. For the performance of the dataset tested obtained an accuracy of 84.221%, there was a difference of 6% from the test results using RapidMiner which obtained an accuracy of 91.40%. The use of algorithms is highly recommended to classify the nutritional status of toddlers. The application of the toddler nutritional status classification system that the author designed is feasible because it can be faster and more effective in classifying the nutritional status of toddlers based on the zscore set and there is only a difference of 6% from the same dataset prediction testing using rapid miner. Based on the dataset of system test results, it can be concluded that 89.24% of toddlers in Bringin sub-district are well nourished.

This is an open access article under the [CC BY-NC](#) license.



Corresponding Author:

Sri Mujiyono,

Fakultas Komputer dan Pendidikan,

Universitas Ngudi Waluyo,

Jl. Diponegoro No.186, Ngablak, Gedanganak, Kec. Ungaran Tim., Kabupaten Semarang, Jawa Tengah 50512

Email: ngudiwaluyo@unw.ac.id

1. INTRODUCTION

In developing countries, chronic malnutrition and infectious diseases lead to stunting, affecting 24.4% of children under five in 2022. Stunting, which usually occurs in toddlers and will become a public health problem in Indonesia resulting in a decrease in the productivity and quality of Indonesian human resources in the future. Stunting can make children more likely to experience illness, die, experience impaired motor-brain development, and be less productive. Around 35.6% of the prevalence of stunting cases in the Republic of Indonesia in 2010 increased to 37.2% in

2013. Decreased stunting prevalence to 30.8% in 2018 and 24.4% in 2022. According to SSGBI, Central Java province has a high stunting prevalence in 2019, which is 27.6%, and drops to 20.9% in 2022, with this prevalence meaning that 1 in 5 toddlers in Central Java province suffers from growth and development disorders. Babies born with low birth weight generally have difficulty catching up with growth. If malnutrition during the period inside the fetus is followed by inadequate food intake during the first two years of life, the risk of growth inhibition increases.

On this occasion, the author will test the nutritional status of toddlers on 186 toddler data in kecamatan bringin, kabupaten semarang. Based on the results of several journals that the author reviewed, the author chose the C4.5 algorithm where the C4.5 algorithm can be used to make decision trees to make predictions with ideal accuracy to predict the nutritional status of toddlers, besides that the C4.5 algorithm can also handle discrete and numeric type attributes. The C4.5 algorithm is an algorithm that functions to generate a decision tree. Decision trees are a method of classification and prediction by exploring data, finding hidden relationships between a number of prospective input variables and target variables. In this study, the author implements the standard reference for weight and height of toddlers aged 0-60 months that have been set by permenkes no. 20 tahun 2020 into the system that the author designed to categorize the nutritional status of toddlers because based on the results of literature studies from the author, no one has applied it to research, besides that the author also conducted manual testing and testing using RapidMiner as a comparison.

2. RESEARCH METHOD

The research methodology is organized into several stages. Data collection approaches, data analysis strategies, methods used, and testing methods are part of the research methodology of this final project.

PHP Programming Language

PHP or Hypertext Preprocessor is an open source scripting language commonly used by Web Developers to build a Web. Not a few projects such as GUI, Dynamic Website are created using PHP. PHP itself was invented by Rasmus Lerdorf whose first version was not released to the public in 1994, at that time Rasmus Lerdorf only used it to track who visited his online resume on his website. In 1995 PHP became available and widely used known as Personal Home Page Tools. PHP itself has disadvantages and advantages including: a) Widely used to create large websites such as Facebook, Youtube, Flickr and many more, it indicates that PHP itself has a large community that makes it easier for users to share knowledge and exchange ideas, b) There are many websites scattered using PHP language indicating that PHP is easy to learn as well as many tutorials and sources on the internet, c) PHP can be accessed by users for free or can be called open source so that users are free to use PHP without having to pay royalties, d) The installation process is not complicated, so it is more concise than other programming languages.

In addition to the advantages of PHP, of course, it has disadvantages, including: a) PHP can be easily changed and modified its function because PHP has an interpreter nature, so if there is source code that you want to hide is not suitable for making applications, b) Because it has a large community, it causes fierce competition for web developers.

Data Collection

The following are some of the author's steps to obtain and collect data for the final project: a) Interview: is the first step that the author took to obtain information and understanding about the nutritional status of toddlers from nutritionists in the study area. Interviews were conducted with resource persons from nutritionists in the family health section in Dinas Kesehatan Kabupaten Semarang, b) Observation: The next step is an observation made in the area chosen by the author, namely Kalongan village. Observations made by pustu local village to monitor the situation at the location and collect data on toddlers from the village midwife, c) Literature Study: This method is carried out to collect theories about the C4.5 algorithm and toddler nutrition taken from several experts through journals, websites, books and electronic books.

Data Analysis

At this stage, research and data collection are carried out, the data needed include data on weight, height, age and sex of toddlers. The data used were 186 data on toddlers aged 0-60 months. After the data is collected, the next step before processing the data is to clean the data, including: a) Data Understanding: The result of this stage is the collection and exploration of data. The data chosen by the authors in this study came from data on toddlers from an area. From the toddler data will be sorted and taken which has several attributes including name, address, toddler age, measurement date, weight and height, b) Data Preparation: After the previous step it is known that there are some anomalies in the data attributes obtained. In this step the data will be prepared in order of date of birth, date of measurement, NIK, Name, address, gender, age, height, weight and whether or not weight gain from previous measurements.

Data Mining

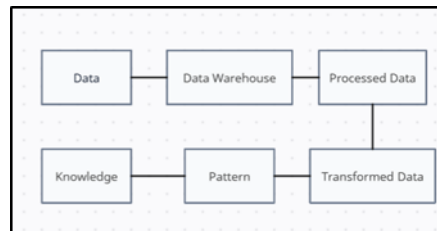


Figure 1. Data mining flow

Data mining has many methods and variations depending on the function and purpose so that the right method is used in solving it, data mining itself is a process of forming patterns to dig up information that has not been found.

Category Reference

At this stage, a categorical process is carried out where the data that was originally numerical will be changed to categorical in accordance with the child's weight and height standards, then his nutritional status will be determined based on the index determined by permenkes no. 2 tahun 2020.

```
$history = $this->db->query("select
jenis_kelamin,
CASE WHEN zscore_bb_u <= -3 THEN 'BERAT BADAN SANGAT KURANG' WHEN zscore_bb_u > -3 AND zscore_bb_u <= -2 THEN 'BERAT BADAN KURANG' WHEN zscore_bb_u > -2 AND
zscore_bb_u <= -1 THEN 'BERAT BADAN NORMAL' ELSE 'BERAT BADAN LEBIH' END as bb_u,
CASE WHEN zscore_tb_u <= -3 THEN 'SANGAT PENDEK' WHEN zscore_tb_u > -3 AND zscore_tb_u <= -2 THEN 'PENDEK' WHEN zscore_tb_u > -2 AND zscore_bb_u <= 3 THEN
'NORMAL' ELSE 'TINGGI' END as tb_u, CASE WHEN zscore_tb_bb <= -3 THEN 'GIZI BURUK' WHEN zscore_tb_bb > -3 AND zscore_tb_bb <= -2 THEN 'GIZI KURANG' WHEN
zscore_tb_bb > -2 AND zscore_tb_bb <= 1 THEN 'GIZI BAIK' WHEN zscore_tb_bb > 1 AND zscore_tb_bb <= 2 THEN 'BERESIKO GIZI LEBIH' WHEN zscore_tb_bb > 2 AND
zscore_tb_bb <= 3 THEN 'GIZI LEBIH' ELSE 'OBESITAS' END as tb_bb
From dataset")->result_array();
$column = array();
if(sizeof($history)>0){
$column = $history[0];
$column = array_keys($column);
}
```

Figure 2. Implementation of nutritional status index

For reference to weight and height standards used as a database in phpmysql and for nutritional status category indexes implemented into source code located in the C45 file in the application folder. Then the system will define the "Age", "Gender", "Height" and "Weight" attributes as support and the "NUTRITIONAL STATUS" attribute as the target.

Calculation Method

a. C4.5 algorithm

The method used to create the decision tree is the C4.5 algorithm with the following steps: Set up a dataset, Calculate total Entropy in dataset. The calculation of the entropy value uses the following formula:

$$Entropy(S) = \sum_n -p_i * \log_2 p_i$$

Information:

S: Set partition S case
 n: Total
 pi: Proportion of Si to S

Calculating Gain

The attribute with the highest value among existing attributes is selected as root. To calculate the gain is used the following formula:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i)$$

Information:

S = Sample

A = Attribute

n = Number of partitions of attribute set A

|Si| = Number of samples on partition to -i

|S| = sample in Sum S

Calculate SplitInfo

For the calculation of Gain Ratio, a new term called SplitIn formation is needed. The attribute with the highest Gain Ratio will be used as the test attribute for the node, where gain is the gain information. Normalization of information gain is applied in this approach, using what is referred to as splitIn formation. Here's the equation to calculate the SplitIn formation

$$SplitInfo(S, A) = \sum_{i=1}^n -\frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

Calculating the GainRatio

The C4.5 algorithm uses GainRatio as the attribute selection. The following equation is the equation for calculating the GainRatio. GainRatio= Gain/SplitInfo

Find the highest GainRatio value as well as create a temporary decision tree. Continue the node by repeating the calculation steps above until the decision tree can be made.

Performance

Confusion Matrix is a step to measure the effectiveness of machine learning classification accuracy, the output of which can be two or more classes.

Table 1. Confusion matrix for two classes

| | | Kelas Aktual | |
|------------------|---------|-----------------------|-----------------------|
| | | Positif | Negatif |
| Prediction Class | Positif | True Positif (TP) | False Positif (FP) |
| | Negatif | False Negatif (FN) | True Negatif (TN) |

To calculate the accuracy value of the following equation is a formula for calculating accuracy:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

TP: The value of a class is predicted as positive and the actual class is positive

TN: The value of a class is predicted as positive and the actual class is positive

FP: The value of a class is predicted as positive and the actual class is negative

FN: The value of a class is predicted as positive and the actual class is positive

3. RESULTS AND DISCUSSIONS (10 PT)

Result

The initial step carried out in this study after obtaining data is to understand and select data that has the attributes needed in data processing, namely weight, height, gender and age as data to be processed and Date of Birth, Date of Measurement, NIK, and address as identity data. The data taken amounted to 186 toddlers who will be entered into the system for zscore calculations and data mining.

a. System Testing

After the data is selected and formatted as required by the system, it is entered into the system on page C45. The system is created to automatically calculate z-score values from weight, height, gender and age data entered according to the rules permenkes no. 2. With this reference, the system detects toddlers who have good nutrition totaling 166, malnutrition 10 toddlers, at risk of nutrition more than 5 toddlers, nutrition more than 2 toddlers and obesity 3 toddlers.

| | | | | | |
|---------------------|---------------------|---------------------|----------------------|---------------|-----------|
| -4.136355279500415 | -7.244444105360243 | -1.3750001490115764 | BERAT BADAN SANGAT K | SANGAT PENDEK | GIZI BAIK |
| -2.4705884002896927 | -2.7749996185302734 | -1.5 | BERAT BADAN KURANG | PENDEK | GIZI BAIK |
| -1.8749995529652708 | -2.410255858492393 | -0.6999998092651367 | BERAT BADAN NORMAL | PENDEK | GIZI BAIK |
| -4 | -7.057141985212054 | -1 | BERAT BADAN SANGAT K | SANGAT PENDEK | GIZI BAIK |
| -1.0476192206211599 | -2.6590939041053137 | 0.9166664017573027 | BERAT BADAN NORMAL | PENDEK | GIZI BAIK |
| -1.9999994980663005 | -4.02380550151407 | 0.5 | BERAT BADAN NORMAL | SANGAT PENDEK | GIZI BAIK |
| -2.4705884002896927 | -2.6842129567352577 | -1.6666673730923856 | BERAT BADAN KURANG | PENDEK | GIZI BAIK |
| -3.2142865901093045 | -3.944445268607783 | -1.6249998509884236 | BERAT BADAN SANGAT K | SANGAT PENDEK | GIZI BAIK |

Figure 3. Example of z-score calculation result by system

For categorizing supporting attributes and target attributes, it is displayed on the initial process menu on page C45, where age, gender, weight and height as supporting attributes and nutritional status as target attributes. The nutritional status itself has been calculated based on the summation results from Zscore and categorized based on the reference threshold table determined by the Ministry of Health.

| jenis_kelamin | bb_u | tb_u | tb_bb |
|-----------------------|---------------------------|---------------|-----------|
| --Atribut Pendukung-- | | | |
| P perempuan | BERAT BADAN SANGAT KURANG | SANGAT PENDEK | GIZI BAIK |
| P perempuan | BERAT BADAN KURANG | PENDEK | GIZI BAIK |
| P perempuan | BERAT BADAN NORMAL | PENDEK | GIZI BAIK |
| L laki - laki | BERAT BADAN SANGAT KURANG | SANGAT PENDEK | GIZI BAIK |
| L laki - laki | BERAT BADAN NORMAL | PENDEK | GIZI BAIK |
| P perempuan | BERAT BADAN NORMAL | SANGAT PENDEK | GIZI BAIK |
| P perempuan | BERAT BADAN KURANG | PENDEK | GIZI BAIK |
| P perempuan | BERAT BADAN SANGAT KURANG | SANGAT PENDEK | GIZI BAIK |
| P perempuan | BERAT BADAN NORMAL | PENDEK | GIZI BAIK |
| --Label Target-- | | | |

Figure 4. Categorization of attributes on the initial process menu on page C45

After the system determines the supporting attributes and target attributes, the data will be calculated using the C4.5 algorithm in the prediction and performance menu based on the inputted dataset.

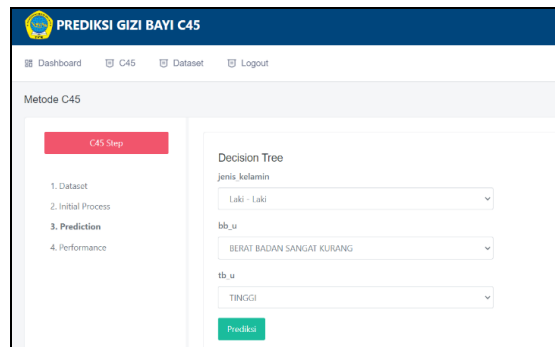


Figure 5. Prediction menu display

In figure 5 is a page for making predictions by selecting the attribute options to be entered, for example: in figure 5 selected attributes of toddlers with male gender with weight included in the category of very less and height classified in the high category then after pressing "Prediction" the results will be displayed as figure 6.

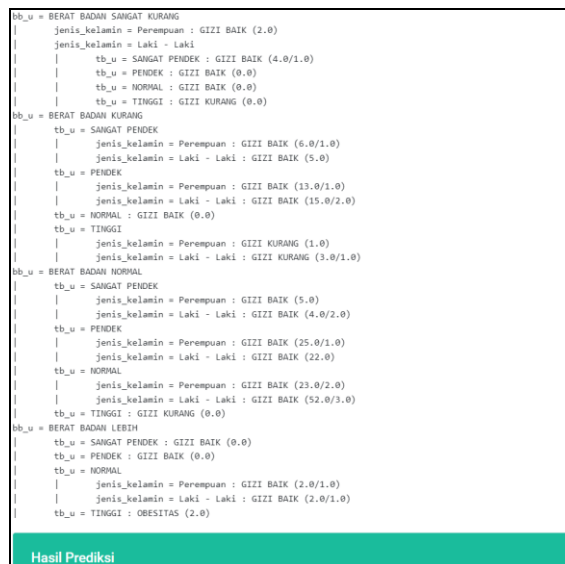


Figure 6. System prediction results

After selecting weight, height and gender categories to predict the nutritional status of toddlers, when the results come out it will be displayed poho decision. So that users can compare whether the results of the prediction match the results of reading the decision tree, decision tree located above the system's predicted results.

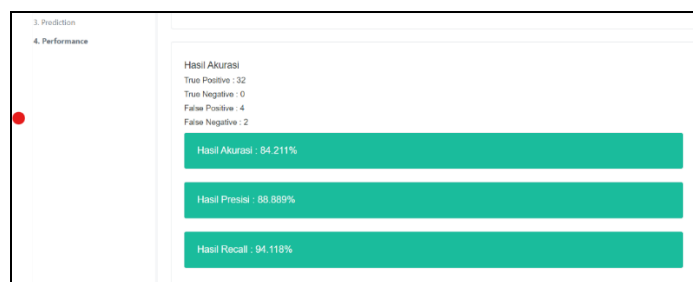


Figure 8. Results of dataset accuracy on the performance menu

The test results of the system that the author designed, on the performance menu of the tested dataset got an accuracy result of 84.211%, precision result is 88.889% and recall result is 94.118%. The performance menu can be accessed on the C45 page.

b. RapidMiner Testing

Testing using rapidminer is carried out as a comparison between the results of the designed system with the results of rapidminer. Figure 9 is a modeling design for testing using rapidminer.

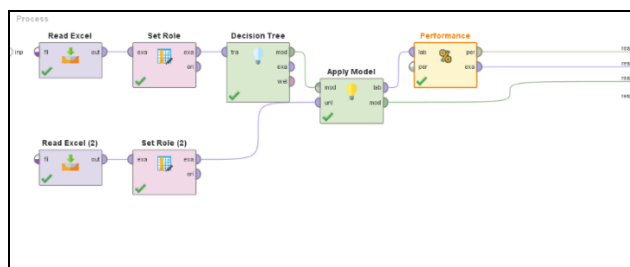


Figure 9. Modeling testing RapidMiner

This step is the initial stage in testing datasets using RapidMiner, by assembling models to derive prediction, performance and decision tree results from datasets. After running the dataset, the results of rapidminer testing will display the decision tree, prediction and performance data. For the results will be explained in the next picture.

```
Tree
bb_u = BERAT BADAN KURANG
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=13, BERESIKO GIZI LEBIH=0, GIZI KURANG=2, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=12, BERESIKO GIZI LEBIH=0, GIZI KURANG=1, OBESTITAS=0, GIZI LEBIH=0)
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=5, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=5, BERESIKO GIZI LEBIH=0, GIZI KURANG=1, OBESTITAS=0, GIZI LEBIH=0)
|
| | jenis_kelamin = Laki - Laki: GIZI KURANG (GIZI BAIK=1, BERESIKO GIZI LEBIH=0, GIZI KURANG=2, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI KURANG (GIZI BAIK=0, BERESIKO GIZI LEBIH=0, GIZI KURANG=1, OBESTITAS=0, GIZI LEBIH=0)
|
bb_u = BERAT BADAN LEBIH
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=1, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=1)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=1, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=1)
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=0, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=0, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
|
bb_u = BERAT BADAN NORMAL
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=49, BERESIKO GIZI LEBIH=1, GIZI KURANG=2, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=21, BERESIKO GIZI LEBIH=2, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=22, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=24, BERESIKO GIZI LEBIH=1, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=2, BERESIKO GIZI LEBIH=1, GIZI KURANG=0, OBESTITAS=1, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=5, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
|
bb_u = BERAT BADAN SANGAT KURANG
|
| | jenis_kelamin = Laki - Laki: GIZI BAIK (GIZI BAIK=3, BERESIKO GIZI LEBIH=0, GIZI KURANG=1, OBESTITAS=0, GIZI LEBIH=0)
| | jenis_kelamin = Perempuan: GIZI BAIK (GIZI BAIK=0, BERESIKO GIZI LEBIH=0, GIZI KURANG=0, OBESTITAS=0, GIZI LEBIH=0)
```

Figure 10. Decision tree results from RapidMiner

The decision tree from RapidMiner is generated from the decision tree operator that the author has previously assembled, based on these results it can be seen that the result of the decision tree from RapidMiner is the same as the result of the decision tree from the system, namely weight as node 1, height as node 1.2 and clash type as node 1.3.

| Row No. | tb_bb | prediction(tb_... | confidence(... | confidence(... | confidence(... | confidence(... | jenis_kelamin | bb_u | tb_u |
|---------|-----------|-------------------|----------------|----------------|----------------|----------------|---------------|------------|-------------|
| 1 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Perempuan | BERAT BADA | SANGAT PEN. |
| 2 | GIZI BAIK | GIZI BAIK | 0.923 | 0 | 0.077 | 0 | Perempuan | BERAT BADA | PENDEK |
| 3 | GIZI BAIK | GIZI BAIK | 0.960 | 0.040 | 0 | 0 | Perempuan | BERAT BADA | PENDEK |
| 4 | GIZI BAIK | GIZI BAIK | 0.750 | 0 | 0.250 | 0 | Laki - Laki | BERAT BADA | SANGAT PEN. |
| 5 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 6 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Perempuan | BERAT BADA | SANGAT PEN. |
| 7 | GIZI BAIK | GIZI BAIK | 0.923 | 0 | 0.077 | 0 | Perempuan | BERAT BADA | PENDEK |
| 8 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Perempuan | BERAT BADA | SANGAT PEN. |
| 9 | GIZI BAIK | GIZI BAIK | 0.960 | 0.040 | 0 | 0 | Perempuan | BERAT BADA | PENDEK |
| 10 | GIZI BAIK | GIZI BAIK | 0.923 | 0 | 0.077 | 0 | Perempuan | BERAT BADA | PENDEK |
| 11 | GIZI BAIK | GIZI BAIK | 0.960 | 0.040 | 0 | 0 | Perempuan | BERAT BADA | PENDEK |
| 12 | GIZI BAIK | GIZI BAIK | 0.907 | 0 | 0.133 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 13 | GIZI BAIK | GIZI BAIK | 0.907 | 0 | 0.133 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 14 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 15 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 16 | GIZI BAIK | GIZI BAIK | 0.960 | 0.040 | 0 | 0 | Perempuan | BERAT BADA | PENDEK |
| 17 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 18 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Laki - Laki | BERAT BADA | PENDEK |
| 19 | GIZI BAIK | GIZI BAIK | 1 | 0 | 0 | 0 | Laki - Laki | BERAT BADA | PENDEK |

Figure 11. Example result of RapidMiner example set prediction

Example set is the output of the apply model tool that displays the prediction results of each data attribute. On this page users can see predictions from all data based on the results of the calculation of the dataset that is used as test data.

| accuracy: 91.40% | | | | | | |
|--------------------------|----------------|-------------------------|------------------|---------------|-----------------|-----------------|
| | true GIZI BAIK | true BERESKO GIZI LEBIH | true GIZI KURANG | true OBESITAS | true GIZI LEBIH | class precision |
| pred. GIZI BAIK | 165 | 5 | 7 | 1 | 2 | 91.67% |
| pred. BERESKO GIZI LEBIH | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. GIZI KURANG | 1 | 0 | 3 | 0 | 0 | 75.00% |
| pred. OBESITAS | 0 | 0 | 0 | 2 | 0 | 100.00% |
| pred. GIZI LEBIH | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 99.40% | 0.00% | 30.00% | 66.67% | 0.00% | |

Figure 12. RapidMiner test performance results

The results of the test that the author conducted using the RapidMiner dataset inputted got a score of 91.40%, highest precision value of 100% and highest recall value of 99.40%. The results of this accuracy are generated from the performance operators that the author has designed.

c. Manual Testing

This manual test is a description of how to test the dataset if done manually, manual testing using the entropy and gain formulas implemented into excel formulas. There are 20 data samples that will be tested manually, here are the test results:

Table 2. Data to test

| Gender | BB_U | TB_U | BB_TB |
|--------|---------------|------------|---------------------------|
| Woman | UNDERWEIGHT | SHORT | GOOD NUTRITION |
| Woman | NORMAL WEIGHT | SHORT | GOOD NUTRITION |
| Male | NORMAL WEIGHT | SHORT | GOOD NUTRITION |
| Male | NORMAL WEIGHT | SHORT | GOOD NUTRITION |
| Male | UNDERWEIGHT | SHORT | UNDERNUTRITION |
| Woman | UNDERWEIGHT | SHORT | GOOD NUTRITION |
| Male | UNDERWEIGHT | SHORT | UNDERNUTRITION |
| Woman | UNDERWEIGHT | SHORT | UNDERNUTRITION |
| Woman | UNDERWEIGHT | TALL | UNDERNUTRITION |
| Male | NORMAL WEIGHT | NORMAL | UNDERNUTRITION |
| Male | NORMAL WEIGHT | VERY SHORT | AT RISK OF MORE NUTRITION |
| Woman | NORMAL WEIGHT | SHORT | AT RISK OF MORE NUTRITION |
| Male | NORMAL WEIGHT | NORMAL | UNDERNUTRITION |
| Male | UNDERWEIGHT | SHORT | UNDERNUTRITION |
| Woman | UNDERWEIGHT | TALL | UNDERNUTRITION |

| | | | |
|-------|---------------|--------|---------------------------|
| Male | NORMAL WEIGHT | NORMAL | AT RISK OF MORE NUTRITION |
| Woman | NORMAL WEIGHT | NORMAL | AT RISK OF MORE NUTRITION |
| Woman | NORMAL WEIGHT | NORMAL | AT RISK OF MORE NUTRITION |
| Woman | NORMAL WEIGHT | NORMAL | GOOD NUTRITION |
| Male | NORMAL WEIGHT | SHORT | GOOD NUTRITION |

In the data to be tested using this manual testing, the data taken has passed a data selection process and is categorized based on the standard weight and height of toddlers based on the age set by Permenkes No. 2 of 2020. After the data is ready, author performs data splitting to find out the value of each data attribute.

Table 3. Node 1 manual testing

| Node 1 | Sum | UNDERNUTRITION | GOOD NUTRITION | AT RISK OF MORE NUTRITION | ENTROPHY | GAIN |
|---------------|-----|----------------|----------------|---------------------------|-------------|-------------|
| total | 20 | 8 | 7 | 5 | 0.983538631 | |
| gender | | | | | | 0.019343614 |
| Male | 10 | 5 | 3 | 2 | 0.937230563 | |
| Woman | 10 | 3 | 4 | 3 | 0.991159471 | |
| bb_u | | | | | | 0.131912466 |
| UNDERWEIGHT | 8 | 6 | 2 | 0 | 0 | |
| NORMAL WEIGHT | 12 | 2 | 5 | 5 | 0.935893281 | |
| tb_u | | | | | | 0.17120785 |
| VERY SHORT | 1 | 0 | 0 | 1 | 0 | |
| SHORT | 11 | 4 | 6 | 1 | 0.834202289 | |
| NORMAL | 6 | 2 | 1 | 3 | 0.920619836 | |
| TALL | 2 | 2 | 0 | 0 | 0 | |

In the calculation of node 1, it is known that the attribute with the highest gain value is the tb_u attribute (height) with the highest entropy value of normal height, then in the next step creating node 1.2 by grouping normal height categories, can be seen in Table 4.

Table 4. Normal height category grouping

| Normal Height | | | |
|---------------|---------------|--------|---------------------------|
| TB_U | BB_U | Gender | BB_TB |
| NORMAL | NORMAL WEIGHT | Male | UNDERNUTRITION |
| NORMAL | NORMAL WEIGHT | Male | UNDERNUTRITION |
| NORMAL | NORMAL WEIGHT | Male | AT RISK OF MORE NUTRITION |
| NORMAL | NORMAL WEIGHT | Woman | AT RISK OF MORE NUTRITION |
| NORMAL | NORMAL WEIGHT | Woman | AT RISK OF MORE NUTRITION |
| NORMAL | NORMAL WEIGHT | Woman | GOOD NUTRITION |

After the author grouped the normal height categories and attributes related to these attributes. The author repeats the previous step, which is to find the value of each attribute to determine the highest entropy and gain values. For the calculation can be seen in table 5.

Table 5. Calculation of entropy and gain on node 1.2

| Node 1.2 | Sum | UNDERNUTRITION | GOOD NUTRITION | AT RISK OF MORE NUTRITION | ENTROPHY | GAIN |
|----------|-----|----------------|----------------|---------------------------|-------------|-------------|
| total | 6 | 2 | 1 | 3 | 0.920619836 | |
| gender | | | | | | 0.464263087 |
| Male | 3 | 2 | 0 | 1 | 0 | |

| | | | | | |
|---------------|---|---|---|---|-------------|
| Woman | 3 | 1 | 1 | 2 | 0.912713498 |
| bb_u | | | | | - |
| UNDERWEIGHT | 0 | 0 | 0 | 0 | 0.928526174 |
| NORMAL WEIGHT | 6 | 2 | 1 | 3 | 0.920619836 |

Based on the results of the calculation of node 1.2, it is known that the gender attribute has the highest gain value and the woman gender with the highest entropy value, from these results node 1.3 will be made with grouping with th woman gender category. It can be seen in table 6.

Table 6. Grouping of female gender categories

| Woman | | | |
|--------|---------------|--------|---------------------------|
| Gender | BB_U | TB_U | BB_TB |
| Woman | NORMAL WEIGHT | NORMAL | AT RISK OF MORE NUTRITION |
| Woman | NORMAL WEIGHT | NORMAL | AT RISK OF MORE NUTRITION |
| Woman | NORMAL WEIGHT | NORMAL | GOOD NUTRITION |

After the author grouped the gender categories and attributes related to these attributes. The author repeats the previous step, which is to find the value of each attribute to determine the highest entropy and gain values. For the calculation can be seen in table 7.

Table 7. Calculation of entropy and gain on node 1.3

| Node 1.3 | Sum | UNDERNUTRITION | GOOD NUTRITION | AT RISK OF MORE NUTRITION | ENTROPHY | GAIN |
|---------------|-----|----------------|----------------|---------------------------|----------|------|
| total | 3 | 0 | 1 | 2 | 0 | |
| bb_u | | | | | | 0 |
| UNDERWEIGHT | 0 | 0 | 0 | 0 | 0 | |
| NORMAL WEIGHT | 3 | 0 | 1 | 2 | 0 | |
| tb_u | | | | | | 0 |
| VERY SHORT | 0 | 0 | 0 | 0 | 0 | |
| SHORT | 0 | 0 | 0 | 0 | 0 | |
| NORMAL | 3 | 0 | 1 | 2 | 0 | |
| TALL | 0 | 0 | 0 | 0 | 0 | |

From the calculation results on node 1.3 it can be seen that the entropy and gain values of all attributes are 0 therefore, the calculation of entropy and gain cannot be continued, then the search for entropy and gain values stops at node 1.3 and will continue to implement the above calculations into a decision tree.

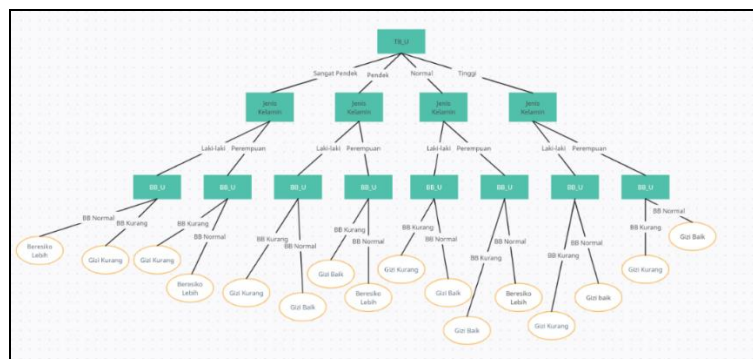


Figure 11. Decision tree from manual testing.

The decision tree is sequentially designed based on the attribute with the highest entropy value on node 1, followed by the attribute with the highest entropy value on node 1.2, and then the attribute with the highest entropy value on node 1.3. The decision trees of manual testing and testing use different systems because the amount of data tested is different.

Discussion

Based on the analysis of testing using this system, the results of 186 datasets tested, a predictive experiment was conducted on toddlers with male gender categories, very underweight and height predicted that the toddler was included in the category of malnutrition. For the performance of the dataset tested obtained an accuracy score of 84.221%, there was a difference of 6% from the test results using RapidMiner which obtained an accuracy score of 91.40%. After testing with various methods, it can be concluded that testing using a toddler nutritional status prediction system shortens the time with acceptable results compared to testing using other methods.

4. CONCLUSION

Based on the analysis of the design and testing that the author did, the author can conclude from "Classification of nutritional status of toddlers using the C4.5 algorithm method in kecamatan Bringin" as follows: The use of algorithms is highly recommended for classifying the nutritional status of toddlers, as they can produce decision trees and are very efficient in handling discrete and numerical attributes, as well as producing acceptable accuracy. The toddler nutritional status classification system application that the author designed is feasible because it can categorize the nutritional status of toddlers automatically based on the set zscore and there is only a difference of 6% from the same dataset prediction testing using a rapid miner. The system that the author designed can be faster and more effective in classifying the nutritional status of toddlers. Based on the dataset of system test results, it can be concluded that 89.24% of toddlers di kecamatan bringin well nourished. It is expected that future research will use other methods in applying case classification of nutritional status of toddlers, besides that it is also expected that in the future it can be developed into an android / ios application version.

A CKNOWLEDGEMENTS

Praise and gratitude continue for the author to offer the presence of Allah SWT who has bestowed His Mercy and Grace because it is with His permission that researchers can complete a thesis entitled "Prediction of Nutritional Status of Toddlers Using C4.5 Algorithm". On this occasion, the author would like to express his deepest gratitude to all parties who have helped the author in completing this article.

REFERENCES

- Adani, M. R. (2021). *Apa itu data mining : pengertian, fungsi, metode dan contoh*. Retrieved January 15, 2023, from <https://www.sekawanmedia.co.id/blog/data-mining/>
- Anies Irawati, S. (2014). Status Gizi Ibu Sebelum Hamil Sebagai Prediksi Berat dan Panjang Bayi di Kecamatan Bogor Tengah . *Penel Gizi Makan, XXXVI(2)*, 119-128.
- Anjar Wanto, M. N., & Agus Perdana Windarto, D. H. (2020). *Data Mining : Algoritma dan Implementasi*. Medan: Yayasan Kita Menulis.
- Erricha Paramitha Dewi, A. E. (2015). Klasifikasi Status Gizi Balita Menggunakan Metode Algoritma C4.5 Berbasis Web. *Universitas Muhammadiyah Sidoarjo*, 1-14.
- Hajar Izzatul Islam, M. K., & Enri, U. (2022). Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan, 10*, 116-125.
- Hananda Hafizan, A. N. (2020). Penerapan Metode Klasifikasi Decixion Tree Pada Status Gizi Balita di Kabupaten Simalungun. *KESATRIA*, 68-72.
- ilmuskripsi. (2016). *Algoritma C4.5*. Retrieved Desember 7, 2022, from <https://www.ilmuskripsi.com/2016/07/algoritma-c45.html>

- Irwansyah Irwansyah, D. I., & Hakimi, M. (2016). Kehamilan Remaja dan Stunting Anak Usia 6-23 bulan di Lombok Utara. *Berita Kedokteran Masyarakat*, 32(6).
- Juliardi. (2019). *C4.5 Implementation in PHP*. Retrieved 11 6, 2022, from <https://github.com/juliardi/C45>
- Miftahul Jannah, N. (2021). Riwayat Kekurangan Energi Kronis (KEK) pada Ibu dan Kejadian Stunting pada Balita di Wilayah Kerja Puskesmas Turikale. *Media Kesehatan Politeknik Kesehatan Makassar*, XVI, 343-352.
- Mutammimul Ula, A. F., & Mauliza, M. A. (2022). Application of Machine Learning In Determining The Clasification of Childrens Nutrition With Decision Tree. *Jurnal Teknik Informatika (JUSTIF)*, III, 1457-1465.
- Panji Bimo Nugroho Setio, D. R., & Winarno, B. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA*, 3, 64-71.
- RI, J. B. (2023, Januari 13). *Permenkes No. 2 Tahun 2023*. Retrieved Maret 11, 2023, from <https://peraturan.bpk.go.id/Home/Details/245563/permenkes-no-2-tahun-2023>
- Rusda Wajhillah, E. M. (2016). Penerapan Algoritma C4.5 Untuk Diagnosa Status Gizi pada Balita Berbasis Website. *Swabumi*, IV, 178-185.
- Sari Kusumawati, S. M. (2023, Juli 9). Kondisi dan penanganan stunting di kabupaten semarang.
- Slamet Ali Mashar, S., & Budiono. (2021). Faktor-Faktor yang Mempengaruhi Kejadian Stunting pada Anak : Studi Literatur. *Serambi Engineering*, VI(3), 2076-2084.