

# IMPLEMENTATION OF THE DECISION TREES C4.5 ALGORITHM ON MICROSOFT EXCEL FOR THE CASE OF DETERMINING THE NUTRITIONAL STATUS OF TODDLERS

W Susanti<sup>1</sup>, K Dwi<sup>2</sup>, S Lidya<sup>3</sup>, S Indah<sup>4</sup>, C Akiko<sup>5</sup>

Actuarial Science, School of Business, President University, Indonesia

<sup>1</sup>ni.susanti@student.president.ac.id

<sup>2</sup>kadek.citrarini@student.president.ac.id

<sup>3</sup>sellina.lidya@student.president.ac.id

<sup>4</sup>septianing.putri@student.president.ac.id

<sup>5</sup>citra.akikoputri@student.president.ac.id

## ABSTRACT

Nutritional problems in Indonesia currently need attention, especially for toddlers. Many diseases are caused by poor nutritional status, especially in toddlers nutrition. In this study, the C.45 decision tree algorithm was applied in Microsoft Excel to analyze the nutritional status of toddlers of the Posyandu in Wonorejo, Krasa District, Simalungun Regency, namely Posyandu Kenanga. The calculation results show an accuracy of 97%. Of the 50 samples, each has 5 components including gender, weight, head circumference, height, and age. We also found that weight was the most influential factor on the nutritional status of children under five. This study suggests parents pay more attention to their children's weight to get good nutritional status for toddlers.

**Keywords:** Nutrition, Trees Decision, Algorithm C4.5, Microsoft Excel

## 1. Introduction

The nutritional status of children under five is one of the health problems in Indonesia. Case nutrient poor not only be a burden for the family but also become a burden for the state. Toddlers experienced cycles of growth and development which requires substances - substances nutrients are more substantial than groups of elements other so that the toddlers most easily suffer from disorders of nutrition. The nutritional status of toddlers can be assessed by measuring the human body which is known as " Anthropometry". Measurements Anthropometric can be done by doing the measurement indicators of heavy weight and high body and pay attention to the age and type of gender the toddler was alone. The size of anthropometry to vote status of nutrition is a combination between each measure indicators of anthropometry are commonly used to assess the status of nutrition that is common is the index heavy weight against age, index heavy weight against height weight and the index high body against age. For children in general, comparison between weight and age represents an index which to measure the growth.

To be able to do the inspection and measurement of anthropometry in order to be able to know the status of nutrition toddlers can come to a place of service health community as one of them, namely Posyandu Kenanga, location at Wonorejo, Krasa District, Simalungun Regency. At posyandu this report the data toddlers are stored in book entry posyandu and digital files. Data that when excavated by right it can be known pattern or knowledge to take decisions.

The algorithm C4.5 is one of the algorithms that are used to perform a classification or segmentation or grouping and are predictive. The C4.5 algorithm itself is a development of the ID3 algorithm. Capability classification algorithm C4.5 can be used to build tree decision (Decision Tree), so that the visible tendency of the status of nutrition toddlers and can be acted upon stakeholder interests as it should.

The Decision tree method using the C4.5 algorithm has been proven to help human work in terms of the number of studies that have been conducted using this method, its ability to break down complex decision-making processes into simpler ones, so that decision makers will better interpret solutions to problems (Iykra, 2018). In (Elvicienna and Amalia, 2013), the authors conducted a study using the C4.5 algorithm to predict the results of the DKI Jakarta DPRD legislative election and found that the results were 97.84% accurate and an AUC value of 0.970 with an Excellent Classification diagnostic rate. The C4.5 algorithm is used to predict student satisfaction with the performance of TEDC Bandung Polytechnic lecturers and

found the conclusion that the decision tree method testing with the C4.5 algorithm was more accurate than using the decision stump algorithm with an accuracy rate of 94.62% (Yuliana and Pratomo, 2017). The algorithm one is also used to predict decision making in choosing time deposits and concluded that the attribute most influential on decision making is the input attribute with an accuracy rate of 84.4 % (Nawawi et. al., 2019).

In this paper, the C4.5 algorithm in the health sector is applied to predict what elements have the most influence on the nutritional status of toddlers. Our research takes data from Posyandu Kenanga. Posyandu as a forum for community participation, which organizes a service system for meeting basic needs, improving human quality, has empirically been able to equalize health services. These activities include immunization services, community nutrition education and maternal and child health services (Kesmas, 2013). This classification can help as an evaluation material for parents to pay more attention to the nutritional status of toddlers.

**2.Method**

For this research method we use the calculation algorithm. Algorithms are instructions addressed to a computer so that it can complete a given task. In order for the task to be completed without errors, the instructions were written specifically (academy). For this research we use the calculation algorithm C4.5. The algorithm C4.5 is an algorithm that is used to form the tree decisions. Medium trees decision can be interpreted one way to predict or clarify who is very strong. Decision tree to split the collection of data is large into the sets of records is small by applying a set of rules decisions. The C4.5 algorithm formula is divided into 2 formulas, the first is a formula to find the gain value. The gain formula is as follows:

$$Gain(S, A) = Entrophy(S) - \sum_{i=1}^n \left| \frac{S1}{S} \right| * Entrophy(Si) \tag{1}$$

Where S is set of cases, A is Attribute, n is number of attribute partitions A, |S1| is number of cases in participation to I, |S| is number of cases in S and the second formula is:

$$Entropy(S) = \sum_{i=1}^n \left| -Pi \right| * \log 2Pi \tag{2}$$

Where S is the set of cases, N is the number of partitions s, Pi is probability that derived from case I divided the total case.

**3.Decision Tree in Microsoft Excel**

3.1. Prepare the data to analysis

In this section, we attach the children's data that we got from the Posyandu Kenanga. In this table we include the important components to measure the nutritional status of toddlers, namely gender, age (months), body weight (kg), height (cm), head circumference (cm), and also health status nutrition at that time.

**Table 1.** Dataset of Toddlers in Posyandu Kenanga (Faridatul,2015)

Jenis kelamin	Usia(bulan)	Bb(kg)	Tb(cm)	Lka(cm)	Status
Laki-laki	9	7.2	60	37	Kurang
Laki-laki	2	4	54	37	Kurang
Laki-laki	3	5.5	58	35	Baik
Laki-laki	3	4.2	58	35	Kurang
Perempuan	2	3	53	35	Kurang
Perempuan	2	4.5	52	34	Baik
Laki-laki	6	7.8	58	36	Baik
Perempuan	6	5.2	62	36	Kurang

Perempuan	5	7	58	35	Baik
Laki-laki	2	4	54	37	Kurang
Perempuan	6	5.3	60	35	Kurang
Laki-laki	10	10.2	64	38	Baik
Perempuan	3	3.8	56	35	Kurang
Laki-laki	12	10.7	74	37	Baik
Laki-laki	11	9	72	35	Baik
Perempuan	9	8	58	37	Kurang
Laki-laki	3	4.2	58	38	Kurang
Laki-laki	9	7.2	60	35	Kurang
Laki-laki	7	7.9	68	36	Baik
Perempuan	5	6	55	35	Baik
Perempuan	7	8.2	67	37	Baik
Perempuan	3	4.8	54	35	Kurang
Laki-laki	9	7.2	60	37	Kurang
Perempuan	4	5.8	58	35	Baik
Laki-laki	3	4.2	58	35	Kurang
Perempuan	5	7	58	35	Baik
Laki-laki	9	7.2	60	37	Kurang
Laki-laki	9	6.8	73	39	Kurang
Laki-laki	8	8.8	61	38	Baik
Perempuan	4	5	53	35	Kurang
Perempuan	10	7	73	38	Baik
Laki-laki	5	7.8	60	36	Baik
Laki-laki	3	4.2	58	35	Kurang
Perempuan	6	6.2	64	35	Baik
Laki-laki	9	10	82	37	Baik
Laki-laki	8	8.8	62	38	Baik
Perempuan	2	3.4	52	35	Kurang
Perempuan	3	4	54	35	Kurang
Perempuan	9	7.5	60	38	Kurang
Laki-laki	6	8	66	36	Baik
Laki-laki	10	11.8	66	39	Baik
Perempuan	9	7.5	60	38	Kurang
Laki-laki	10	9	75	39	Baik
Perempuan	1	3	50	34	Kurang
Laki-laki	8	8.5	70	39	Baik
Perempuan	3	4	54	35	Kurang
Laki-laki	7	10.5	69	38	Baik
Perempuan	4	5	53	35	Kurang
Laki-laki	10	9.6	68	38	Baik
Laki-laki	7	7.5	70	38	Baik

3.2. Calculate the number of classifications per element

This is a data processing table consisting of data entry categories which will be processed for clarification. Amount is the total number of entries. Amount Status Baik and Amount Kurang is the number of data entries with the Yes / No criteria in the work category. we can sort easily using the COUNTIF function, then, = COUNTIF (range, criteria). For example: If amount = COUNTA (select Jenis Kelamin), If Status Baik = COUNTIF (select Status “Baik”), If Status Kurang = COUNTIF (select Status “Kurang”), If Jenis Kelamin Laki-laki = COUNTIF (select Jenis kelamin, “Laki-laki”), If Jenis Kelamin Laki-laki and Status baik = COUNTIFS (select Jenis kelamin, “Laki-laki”, select Status, “Baik”), If Jenis Kelamin Laki-laki and Status kurang= COUNTIFS (select jenis kelamin, “laki-laki, select status “Kurang”). The same way until the LKA (CM).

3.3. Calculate Entrophy and Gain

- The formula for calculating the entropy is  $((- \text{number of Baik} / \text{amount}) * \text{Log}_2 (\text{number of Baik} / \text{amount}) + (- \text{number of Kurang} / \text{amount}) * \text{Log}_2 (\text{number of Kurang} / \text{amount}))$

To calculate Log2 you can use the formula: IMLOG2

For example:

$$1. \text{ Entropy} = ((-\text{Status Baik}/\text{amount})*\text{IMLOG2}(\text{Status Baik}/\text{amount})+(-\text{Status Kurang} / \text{amount})*\text{IMLOG2}(\text{Status Kurang} / \text{amount})) = ((-17/28)*\text{IMLOG2}(17/28)+(-11/28)*\text{IMLOG2}(11/28)) = 0,966618633$$

The same way until the lingkak kepala anak (LKA (CM) entropy.

- The formula for calculating GAIN is  $(\$Entropytotal\$)-((\text{kat\_total}/\text{total}) * \text{kat\_entropy})-.....n$

For example:

$$2. \text{ Gain Jenis Kelamin} = (0,966618633) - ((28/50) * 0,966618633) - ((22/50) * 0.945660305) = 0.04260303$$

The same way until the lingkak kepala anak (LKA (CM) GAIN. The result will be like this:

**Table 2.** Result of classification per element

	Amount	Status baik	Status kurang	Enthophy	Gain
	50	25	25	1	
Jenis kelamin					0.04260303
Laki-laki	28	17	11	0.966618633	
Perempuan	22	8	14	0.945660305	
Usian(bulan)					0.59444138
1	1	0	1	0	
2	5	1	4	0.721928095	
3	9	1	8	0.503258335	
4	3	1	2	0.918295834	
5	4	4	0	0	
6	5	3	2	0.970950594	
7	4	4	0	0	
8	3	3	0	0	
9	9	1	8	0.503258335	
10	5	5	0	0	
11	1	1	0	0	
12	1	1	0	0	

Bb (kg)					1.05509775
3	2	0	2	0	
3.4	1	0	1	0	
3.8	1	0	1	0	
4	4	0	4	0	
4.2	4	0	4	0	
4.5	1	1	0	0	
4.8	1	0	1	0	
5	2	0	2	0	
5.2	1	0	1	0	
5.3	1	0	1	0	
5.5	1	1	0	0	
5.8	1	1	0	0	
6	1	1	0	0	
6.2	1	1	0	0	
6.8	1	0	1	0	
7	3	3	0	0	
7.2	4	0	4	0	
7.5	3	1	2	0.918295834	
7.8	2	2	0	0	
7.9	1	1	0	0	
8	2	1	1	0	
8.2	1	1	0	0	
8.5	1	1	0	0	
8.8	2	2	0	0	
9	2	2	0	0	
9.6	1	1	0	0	
10	1	1	0	0	
10.2	1	1	0	0	
10.5	1	1	0	0	
10.7	1	1	0	0	
11.8	1	1	0	0	
Tb (cm)					0.59302969
50	1	0	1	0	
52	2	1	1	1	
53	3	0	3	0	
54	5	0	5	0	
55	1	1	0	0	
56	1	0	1	0	
58	10	5	5	1	
60	8	1	7	0.543564443	

61	1	1	0	0
62	2	1	1	1
64	2	2	0	0
66	2	2	0	0
67	1	1	0	0
68	2	2	0	0
69	1	1	0	0
70	2	2	0	0
72	1	1	0	0
73	2	1	1	1
74	1	1	0	0
75	1	1	0	0
82	1	1	0	0
Lka (cm)				0.10772629
34	2	1	1	1
35	20	7	13	0.934068055
36	5	4	1	0.721928095
37	9	3	6	0.918295834
38	10	7	3	0.881290899
39	4	3	1	0.811278124

### 3.4. Eliminate Data

After we calculate the entropy and gain values, we can start eliminating data by first, we see the highest gain value, in our data the highest gain value falls to body weight (BB (KG)). Second, in the BB (KG) section we see the highest entropy value, in our data the highest entropy value falls on body weight 7.5 kg. Third, we eliminate the dataset and only take data whose element weighs (BB (KG)) 7.5 kg.

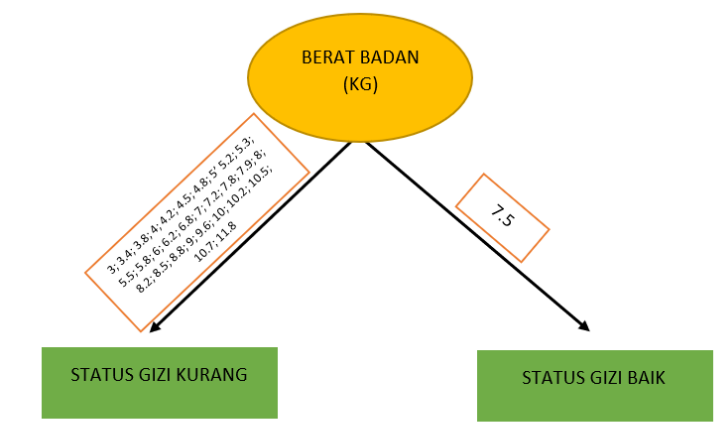
**Table 3:** Data after eliminated

Jenis kelamin	Usia(bulan)	Bb(kg)	Tb(cm)	Lka(cm)	Status
Perempuan	9	7.5	60	38	Kurang
Perempuan	9	7.5	60	38	Kurang
Laki-laki	7	7.5	70	38	Baik

In our data, we can see what qualifications are needed to have a good nutritional status, namely Jenis kelamin "Laki-Laki", Usia(bulan) "7", BB(kg) "7.5", Tb(cm) "70", Lka(cm) "38".

### 3.5. Create Decision Tree

From the calculation above, first we see that body weight gets the highest gain value while the entropy falls at "7.5" while the other entropy value for body weight is 0 (depleted) so we can immediately draw a decision tree as below



**Figure 1.** Decision Tree of classification of nutritional status at Posyandu Kenanga

#### 4. Conclusion

Based on the results of calculations which we achieve by using the methods of algorithm C4.5 in excel can we summarize the status of both poor nutrition toddlers influenced by five (5) component, namely: the type of gender, weight (kg), age of child (months), high body (cm), the child's head circumference (cm), of these 5 (five) components, it is seen that the most influencing component is the internal weight component (KG). Heavy weight of children is affecting a component other and have proven to be seen from the calculations that have us do.

#### References

- Yuliana and Pratomo,D.B. (2017). Algoritma Decision Tree (C4.5) untuk memprediksi kepuasan mahasiswa terhadap kinerja dosen politeknik TEDC Bandung. *Seminar Nasional Inovasi Teknologi*, UN PGRI Kediri. 377-384. Accessed June 29, 2021.  
<https://adoc.pub/algoritma-decision-tree-c45-untuk-memprediksi-kepuasan-mahas.html>
- Elvicienna and Amalia,H. (2013). Algoritma C4.5 untuk prediksi hasil pemilihan legislatif DPRD DKI Jakarta. *Techno Nusa Mandiri* Vol. X No.1. 48-56. Accessed June 29, 2021  
<https://core.ac.uk/reader/229771595>
- Nawawi et. al., 2019. (2019). Algoritma C4.5 untuk Memprediksi Pengambilan Keputusan Memilih Deposito Berjangka. *Techno Nusa Mandiri* Vol.16, No.1. 65-72. Accessed June 29, 2021  
<https://core.ac.uk/download/pdf/229771467.pdf>
- Iykra. (2018.). *Mengenal Decision Tree dan Manfaatnya*. Accessed June 28, 2021,  
<https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d>
- Kesmas. (2013). *Pengertian, Tujuan, dan Kegiatan Posyandu*. Accessed June 29, 2021.  
<http://www.indonesian-publichealth.com/manajemen-posyandu/>
- Faridatul, Choiriyah. (2015). *Implementasi Metode Naive Bayes Sebagai Penentu Status Gizi Balita (Studi Kasus Puskesmas Dukuh Kupang Surabaya)*. Accessed April 15, 2021.  
<http://eprints.umg.ac.id/2028/4/BAB%20III.pdf>