International Journal of Engineering and Computer Science Applications **(IJECSA)**
Vol. 1, No. 2, September 2022, pp. 77~82
DOI: 10.30812/IJECSA.v1i2.2387
e-ISSN : 2828-5611 *https://journal.universitasbumigora.ac.id/index.php/IJECSA*

Web-Based Application for Toddler Nutrition Classification Using C4.5 Algorithm

Hairani Hairani[1*], Lilik Nurhayati[2], Muhammad Innuddin[3]

[1,2,3] *Department Engineering, Bumigora University, Jl. Ismail Marzuki, Cilinaya, Cakranegara, Mataram, Indonesia*
[*]*Corresponding author: Hairani@universitasbumigora.ac.id*

(Received Sep 19 2022 ; Revised Sep 27, 2022 ; Accepted Sep 28, 2022)

***Abstract*— Health is something that is important for everyone, from year to year various efforts have been developed to get better and quality health. Good nutritional status for toddlers will contribute to their health and also the growth and development of toddlers. Fulfillment of nutrition in children under five years old (toddlers) is a factor that needs to be considered in maintaining health, because toddlerhood is a period of development that is vulnerable to nutritional problems. There are more than 100 toddler data registered at the Integrated Healthcare Center in Peresak Village, Narmada District, West Lombok Regency. The book contains data on toddlers along with the results of weighing which is carried out every month. However, to classify the nutritional status of toddlers, they are still going through the process of recording in a notebook by recording the measurement results and then looking at the reference table to determine their nutritional status. This method is still conventional or manual so it takes a long time to determine the nutritional status. Therefore, the solution in this study is to develop a web-based application for the classification of the nutritional status of children under five using the C4.5 method. The stages of this research consisted of problem analysis, collection of 197 instances of nutritional status datasets obtained from Integrated Healthcare Center Presak, analysis of system requirements, use case design, implementation using the C4.5 method, and performance testing based on accuracy, sensitivity, and specificity. The results of this study are a website-based application for the classification of the nutritional status of children under five using the C4.5 method. The performance of the C4.5 method in the classification of the nutritional status of toddlers using testing data as much as 20% gets an accuracy of 95%, sensitivity of 100%, and specificity of 66.6%. Thus, the C4.5 method can be used to classify the nutritional status of children under five, because it has a very good performance.**

Keywords : **Toddler Nutritional Status, C4.5 Algoritm, Data Mining.**

I. INTRODUCTION

Health is something that is important for everyone, from year to year various efforts have been developed to get better and quality health. Good nutritional status for toddlers will contribute to their health and also to the growth and development of toddlers. Fulfillment of nutrition in children under five years old (toddlers) is a factor that needs to be considered in maintaining health, because toddlerhood is a period of development that is vulnerable to nutritional problems.

From the results of the Nutrition Status Monitoring (PSG) data in. 2017, giving the prevalence of stunting in NTB Province of 37.2%, higher than the national average of 29.6%. That number also increased when compared to 2016 which was 29.9% or an increase of 7.29%. For the highest prevalence of stunting in Sumbawa Regency, namely 41.9%, Central Lombok 39.9%, Dompu 38.3%, Mataram City 37.8%, North Lombok 37.6%, Bima 36.6%, Bima City 36 , 3%, West Lombok 36, 1% , and East Lombok 35.1% [1]. There are more than 100 data on children under five registered at the Integrated Healthcare Center in Peresak village, Narmada sub-district, West Lombok regency. The book contains data on toddlers along with the results of weighing which is carried out every month. However, to classify and classify the nutritional status of toddlers, they are still going through the process of recording in a notebook by recording the measurement results and then looking at the reference table to determine their nutritional status. This method is still conventional or manual so it takes a long time to determine the nutritional status.

According to research [2] Assessment of the nutritional status of children under five can be determined through measurements of the human body known as Anthropometry. In order to get the right results, a benchmark is given as a guide, namely the Z-Score. Z-Score is an anthropometric index used internationally for determining nutritional status and growth, which is expressed as a population standard deviation (SD) population. Z-Score is used to calculate nutritional status anthropometrically on body weight for age, height for age, weight for height [3]. Anthropometric examinations and measurements are used to determine the nutritional status of children under five by visiting public health services such as Integrated Healthcare Center.

Previous research that uses the C4.5 method to solve problems such as research [4][5] case of classification of scholarship recipients, research [6] for the case of credit recipient banking, research [7] for the case of prediction of student graduation. The previous research that focused on the problem of nutritional status classification was carried out by researchers such as researcher [8] using the k-nn method for the classification of the nutritional status of children under five with an accuracy of 77.8%. The dataset used is 72 instances with 4 attributes, namely gender, weight, height, and nutritional status. Researcher [9] using the C4.5 method for the classification of the

nutritional status of children under five with an accuracy of 90.93%. The dataset used is 853 instances with 4 attributes, namely gender, weight, age, and nutritional status.

Researcher [10] using the Linkage Agglomerative Hierarchical Clustering method for determining nutritional status by producing 4 clusters. Researcher [11] using the K-Nearest Neighbor, Naïve Bayesian Classification, and Decision Tree methods to identify nutritional status. The results showed that the nave Bayes method had better accuracy than the K-Nearest Neighbor and Decision Tree methods. Researcher [12] used the K-Means method to determine nutritional status with 57 instances of data and resulted in 4 clusters.

Based on previous research, there are differences that will be carried out by this study, namely the dataset used by this study is different from previous research and also develops a web-based application for classifying the nutritional status of toddlers using the C4.5 method.

II. MATERIALS AND METHODS

This research consists of problem analysis, data collection, needs analysis, design, implementation, and testing as shown in Figure 1.

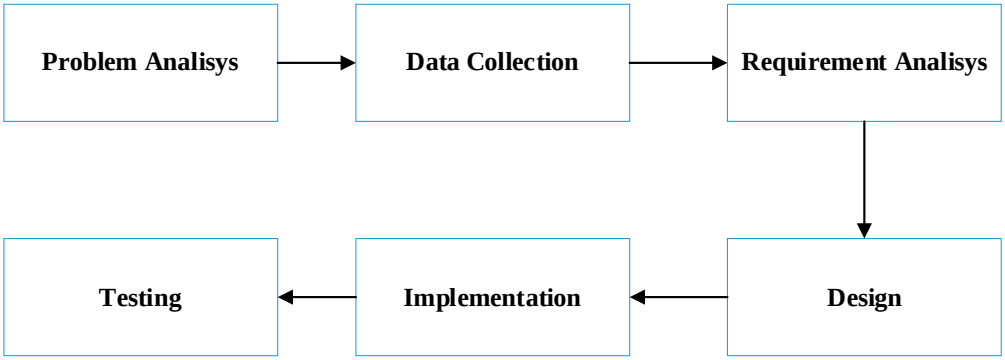

Figure 1. Research Stages

A. Problem Analisys

The problem in this study is that the Integrated Healthcare Center in Peresak Village still uses books to record the results of checking the nutrition of toddlers who visit the Integrated Healthcare Center. This method is less effective and takes a long time, because the data is not stored properly and is easily lost. Not only that, checking the nutritional status of toddlers still uses manual calculations.

B. Data Collection

In this section, 197 instances of under-five nutrition data were collected at Integrated Healthcare Center in Presak village with the attributes of Gender, Age, Weight (BB), Height (TB), and Nutritional Status. The dataset used is as shown in Table 1.

Table 1. Dataset of Toddler Nutritional Status Based on BB/TB

No	Gender	Age (Week)	Wight	Height (cm)	Nutritional Status
1	L	53	13.6	102.1	Good Nutrition
2	L	57	17.7	108.7	Good Nutrition
3	P	51	14.8	100	Good Nutrition
4	P	51	13.7	91.5	Good Nutrition
5	L	59	15.4	102	Good Nutrition
…	……	……	……	…..	……..
194	L	9	9.1	81	Malnutrition
195	L	6	4.3	64.5	Malnutrition
196	p	8	7.1	69	Good Nutrition
197	P	6	8.7	73	Good Nutrition

C. System Functional Requirements Analysis

Functional requirements analysis describes the process of activities that will be implemented in a system and explains the requirements needed for the system to run properly and as needed. The functional requirements of the system built are shown in Table 2.

Table 2. System Functional Requirements Analysis

No	User	Role

1	Nutritionist As Admin	1. login 2. Manage training data menu (Create, Update, Delete) 3. Manage data testing data menu (Create, Update, Delete) 4. Manage Testing of data testing 5. Showing classification results 6. Do a toddler data search 7. Logout
2	Head of Integrated Healthcare Center as Super Admin	1. login 2. Manage training data menu (Create, Update, Delete) 3. Manage user data (Create, Update, Delete) 4. Manage data testing data menu (Create, Update, Delete) 5. Showing classification results 6. Import Toddler Data 7. Can print classification results report 8. Do a toddler data search 9. Can add admin data 10. Logout

D. Design System

At this stage use case diagrams are used to show the functionality of a system about how the system interacts with the outside world or actors. The use case diagram for the application of the nutritional status classification of children under five using the WEB-based C4.5 method is shown in Figure 2.

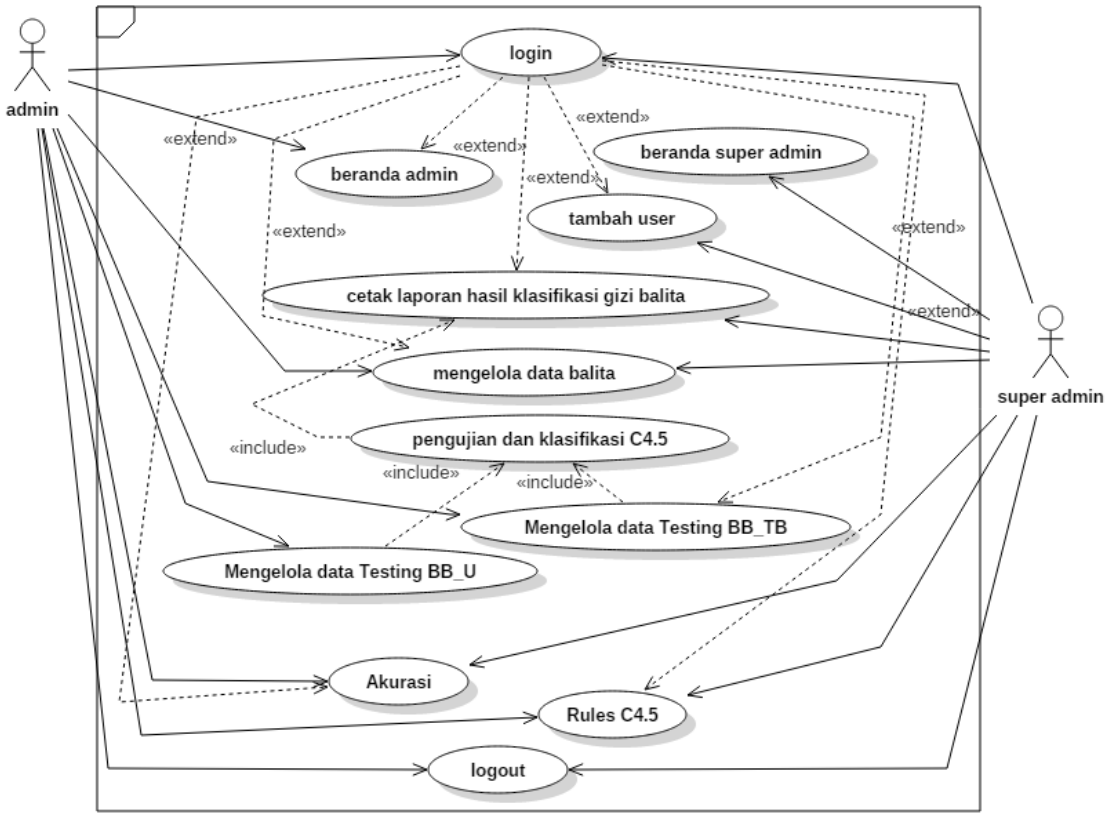

Figure 2. Use Case Design for Toddler Nutritional Status Classification Applications

E. Implementation System

In this section, the implementation of the C4.5 method is used for web-based classification of nutritional status of children under five with PHP and SQL programming languages. The stages of the C4.5 method in the classification of the nutritional status of children under five are shown in Figure 3 [13].

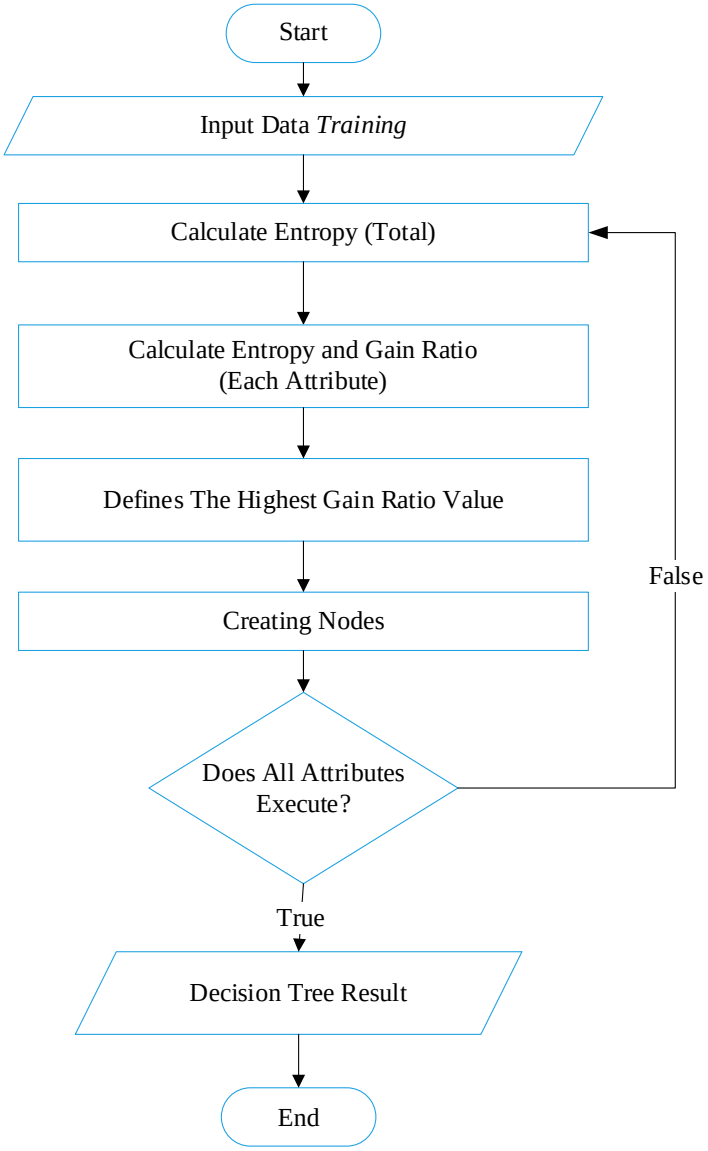

Figure 3. Flowchart of C4.5 Method

The calculation of the entropy value is by equation (1), while the search for the gain value uses equation (2).

$$Entropy(S) = \sum_{i=1}^{n} -\text{pi}.log_2\text{pi} \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} * Entropy(S_i) \quad (2)$$

S is the number of cases, A is the attribute, N is the number of partitions of the attribute A. While |Si| is the number of cases on the I-th partition, and |S| is the number of cases in S. Pi is the proportion of Si to S.

F. Testing

At the stage of testing the performance of the C4.5 method in classifying the nutritional status of toddlers using a confusion matrix table. In this study, performance testing was carried out based on accuracy (3), sensitivity (4), and specificity (5). The confusion matrix table can be seen in Table 3.

Table 3. Confusion Matrix Table

Actual Class	Predicted Class	
	Good Nutriton	Malnutriton
Good Nutriton	*True Positive* (TP)	*False Negative* (FN)
Malnutriton	*False Positive* (FP)	*True Negative* (TN)

The formulas used in the calculation of accuracy (3), sensitivity (4), and specificity (5).

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (3)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

III. RESULTS AND DISCUSSION

This section describes the research results that have been obtained based on the previous stages. The results of the research are an application that has been made along with an evaluation of the performance of the C4.5 method based on accuracy, sensitivity, and specificity. In Figure 4 there is a training data page that is used to train the C4.5 method to recognize the pattern of the dataset used. The formed C4.5 model can be used to classify the nutritional status of children under five based on testing data. The testing data page used is shown in Figure 5.

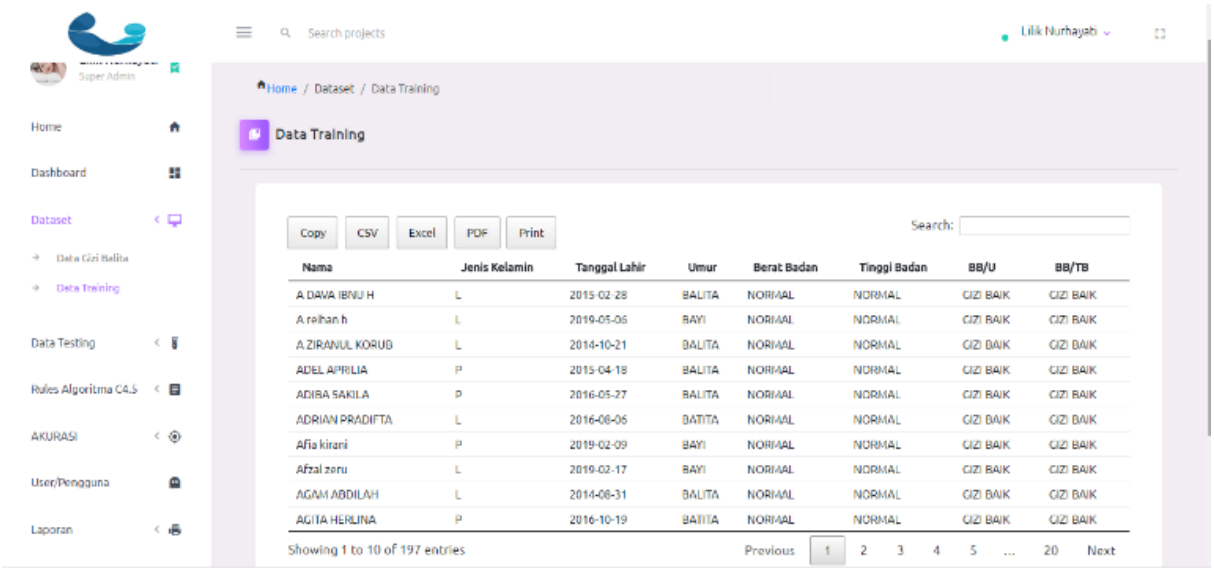

Figure 4. Dataset Training Page

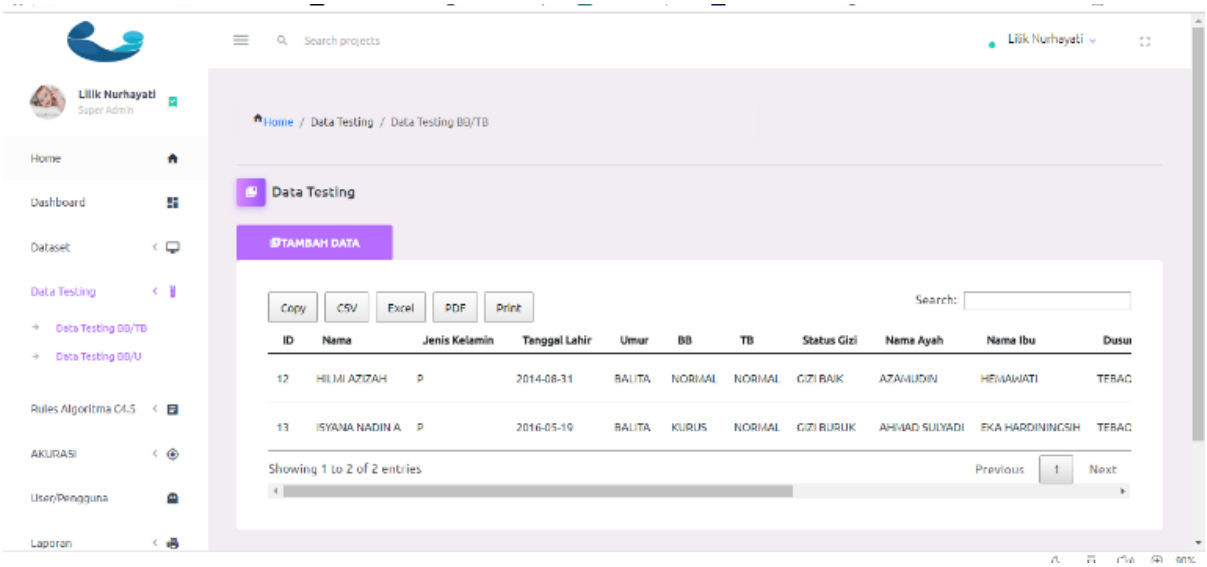

Figure 5. Dataset Testing Pages Based on BB/TB

After carrying out the training process, then testing data testing using a model that has been created based on the results of training data based on accuracy, sensitivity, and specificity Using the confusion matrix table shown in Table 4.

Table 4. Testing Results of Data Testing with Confusion Matrix Table

Actual Class	Predicted Class	
	Good Nutrition	Malnutrition
Good Nutrition	17	0
Malnutrition	1	2

Based on the results of the testing data in Table 4, the accuracy, sensitivity, and specificity of the C4.5 method can be calculated as follows.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} = \frac{19}{20} = 0.95*100 = 95$$

$$Sensitivity = \frac{TP}{TP+FN} = \frac{17}{17} = 1*100 = 100\%$$

$$Specificity = \frac{TN}{TN+FP} = \frac{2}{3} = 0.666*100 = 66.6\%$$

The C4.5 method obtained an accuracy of 95%, a sensitivity of 100%, and a specificity of 66.6%. Thus the C4.5 method can be used to classify the nutritional status of children under five, because it has a very good performance.

IV. CONCLUSION

This study aims to develop an application for classifying the nutritional status of children under five using the web-based C4.5 method. The stages of this research consist of problem analysis, data collection, system requirements analysis, design, implementation, and testing. The results of this study are a website-based application for the classification of the nutritional status of children under five using the C4.5 method. The performance of the C4.5 method in the classification of the nutritional status of children under five gets an accuracy of 95%, a sensitivity of 100%, and a specificity of 66.6%. Further research can handle the problem of unbalanced data in this dataset based on oversampling and undersampling.

REFERENCES

[1] I. dan S. P. N. Dinas Komunikasi, "Angka Kurang Gizi di NTB Masih Tinggi," 2018. https://old.diskominfotik.ntbprov.go.id/content/angka-kurang-gizi-di-ntb-masih-tinggi (accessed Sep. 17, 2022).

[2] Y. R. Kaesmitan and J. A. Johannis, "Klasifikasi Status Gizi Balita Di Kelurahan Oesapa Barat Menggunakan Metode K-Nearest Neigbor," *Multitek Indonesia*, vol. 11, no. 1, pp. 42–50, Aug. 2017, doi: 10.24269/mtkind.v11i1.506.

[3] A. S. Hutasoit, P. Tarigan, and E. R. Siagian, "Implementasi Data Mining Klasifikasi Status Gizi Balita Pada Posyandu Medan Timur Dengan Menggunakan Metode C4.5," *Jurnal Pelita Informatika*, vol. 7, no. 2, pp. 120–125, 2018.

[4] H. Hairani, A. S. Suweleh, and D. Susilowaty, "Penanganan Ketidak Seimbangan Kelas Menggunakan Pendekatan Level Data," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 109–116, 2020, doi: 10.30812/matrik.v20i1.846.

[5] A. S. Suweleh, D. Susilowati, and Hairani Hairani, "Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5," *Jurnal BITe*, vol. 2, no. 1, pp. 12–21, 2020.

[6] H. Hairani, N. A. Setiawan, and T. B. Adji, "Metode Klasifikasi Data Mining dan Teknik Sampling SMOTE Menangani Class Imbalance Untuk Segmentasi Customer Pada Industri Perbankan," in *Prosiding SNST Fakultas Teknik*, 2016, vol. 1, no. 1, pp. 168–172, doi: 978-602-99334-5-1.

[7] D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System On Students Using C4.5 Algorithm," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 358–365, 2020, doi: 10.30812/matrik.v19i2.685.

[8] M. I. Ula M., Ulva A. F., Saputra I., Mauliza M., "Implementation of Machine Learning Using the K-Nearest Neighbor Classification Model in Diagnosing Malnutrition in Children," *Multica Science And Technology (Mst) Journal*, vol. 2, no. 1, pp. 94–99, 2022, doi: 10.47002/mst.v2i1.326.

[9] K. Pinaryanto, R. A. Nugroho, and Y. Basilius, "Classification of Toddler Nutrition Using C4.5 Decision Tree Method," *International Journal of Applied Sciences and Smart Technologies*, vol. 3, no. 1, pp. 131–142, 2021, doi: 10.24071/ijasst.v3i1.3366.

[10] M. Ula, A. Ulva, Mauliza, I. Sahputra, and Ridwan, "Implementation of Machine Learning in Determining Nutritional Status using the Complete Linkage Agglomerative Hierarchical Clustering Method," *Jurnal Mantik*, vol. 5, no. 3, pp. 1910–1914, 2021.

[11] T. M. Mulyono, F. Natalia, and S. Sudirman, "A Study of Data Mining Methods for Identification Undernutrition and Overnutrition in Obesity," in *Proceedings of the 2019 3rd International Conference on Software and E-Business*, 2019, pp. 6–10, doi: 10.1145/3374549.3374565.

[12] S. S. Nagari and L. Inayati, "Implementation of Clustering using K-Means Method to Determine Nutritional Status," *Jurnal Biometrika dan Kependudukan*, vol. 9, no. 1, pp. 62–68, 2020, doi: 10.20473/jbk.v9i1.2020.62.

[13] W. Wiguna and D. Riana, "Diagnosis Of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5 Algorithm," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 71–80, Mar. 2020, doi: 10.33480/pilar.v16i1.1293.