

Classification of Malnutrition Risk in Toddlers Using the C4.5 Method (Case Study: UPTD Gunung Meriah Health Center)

Fathur Rizqi Hasugian

Program Studi Teknologi Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi
Universitas Muhammadiyah Sumatera Utara
Corresponding Author e-mail : rizqifathur56@gmail.com

Article History:

Received: 14-10-2026

Revised: 26-10-2026

Accepted: 26-10-2026

**Keywords: Toddlers, Data Mining,
Decision Tree, Malnutrition,
Classification**

Abstract : *The problem of malnutrition in toddlers requires early identification through systematic anthropometric data processing. This study aims to apply the C4.5 algorithm to classify the nutritional status of toddlers, measure model accuracy, and identify influential attributes at the UPTD Gunung Meriah Health Center. This study uses a quantitative approach with a data mining classification method based on Decision Tree C4.5. The research data consists of documentation of 386 toddlers, consisting of 343 data with good nutrition, 26 data at risk of malnutrition, and 17 data with severe malnutrition. The research instruments are toddler datasets and data processing software. Predictor attributes include age, gender, weight, height, and mid-upper arm circumference (MUAC), while nutritional status is the target variable. The analysis was carried out through preprocessing, decision tree formation, functional testing Black Box Testing, and confusion matrix evaluation with a division of training data and test data of 80:20. The results showed a classification accuracy of 87.18%, with MUAC as the main attribute and weight as a supporting attribute. All main functions of the web-based system run according to the test scenario. In conclusion, the C4.5 algorithm has the potential to support early identification of toddler nutritional status and systematic data management, but the classification results still require verification by health workers.*

How to Cite : Fathur Rizqi Hasugian. (2026). The title. Ambidextrous Journal of Innovation Efficiency and Technology in Organizations. 5(3). pp 608 620
<https://doi.org/10.61536/ambidextrous.5i03.954>

 <https://doi.org/10.61536/ambidextrous.v5i03.954>

This is an open-access article under the [CC-BY-SA License](https://creativecommons.org/licenses/by-sa/4.0/).



Introduction

Malnutrition in toddlers remains a public health issue that requires serious attention, especially since nutritional status is a crucial indicator in assessing a child's health and growth in early life. Undetected malnutrition can lead to impaired physical growth and cognitive development, as well as an increased risk of disease in adulthood. The latest report on the global nutritional status of children indicates that malnutrition remains a challenge in achieving sustainable development goals, necessitating continued strengthening of monitoring and early identification (World Health Organization [WHO], 2025). Assessment of the nutritional status of toddlers is generally conducted through an anthropometric approach, including measurements of weight, height or length, and age, which are then used to assess the child's growth status based on applicable anthropometric standards (WHO, 2025). The results of this assessment serve as the basis for health workers in determining appropriate interventions, particularly to support efforts to prevent malnutrition and stunting.

The Gunung Meriah Community Health Center (Puskesmas) is a first-level health care facility responsible for monitoring the growth of toddlers in its area. The Puskesmas manages toddler data, including age, sex, weight, height, and nutritional status assessment results. The importance of this data management aligns with the implementation of the Indonesian Nutritional Status Survey (SSGI), which is used to describe the nutritional status of toddlers and support the formulation of nutritional intervention policies in Indonesia (Ministry of Health of the Republic of Indonesia, 2025). As the number of toddlers undergoing routine checkups increases, the collected data has the potential to be used to identify patterns of malnutrition risk. However, this data utilization requires systematic processing so that the available information can support the early identification process and decision-making regarding the management of toddlers at risk of experiencing nutritional problems.

The challenge faced in identifying the risk of malnutrition in toddlers at the Gunung Meriah Community Health Center (Puskesmas UPTD) is the use of a traditional approach involving manual recording and evaluation of individual parameters. This process is time-consuming and potentially results in inconsistent assessments as the data volume increases. Furthermore, the suboptimal management of historical toddler data has resulted in the underutilization of important patterns within the data as a basis for decision-making. However, research on determining policy targets based on nutritional status data in Indonesia demonstrates the importance of accurate data analysis to help identify groups requiring further attention and intervention (Laksono, 2024). This situation highlights the need for an approach capable of processing anthropometric data in a structured manner and assisting in a more systematic classification of malnutrition risks.

One approach that can be used to address these issues is data mining, the process of extracting patterns and important information from a set of data to support decision-making. In classification problems, the Decision Tree method can be used to build a model based on data attributes through a decision tree structure. The concept of decision tree induction was introduced by Quinlan (1986) through the Top-Down Induction of Decision Trees (TDIDT) approach, which builds a classification model in stages from the root to the leaves. The development of this approach resulted in the C4.5 algorithm, which is capable of handling both numeric and categorical attributes and produces relatively easy-to-understand decision rules. This algorithm determines classification attributes through iterative entropy and information gain calculations until the data is classified into a specific class. The application of the C4.5 algorithm in classifying children's nutritional status has also been studied by Hutabarat et al. (2026), who developed a model based on anthropometric data and achieved an accuracy of 96.12% in testing conducted at AN-NISA Hospital, Tangerang. These results demonstrate the potential of Decision Tree C4.5 to support nutritional status classification, although model performance still needs to be tested using different data characteristics and health service contexts.

Although previous research has demonstrated the potential of the C4.5 algorithm in classifying children's nutritional status, its application to classify the risk of malnutrition in toddlers based on local data at the Gunung Meriah Community Health Center (UPTD). Therefore, research is needed that considers the characteristics of the data available in the region. This study aims to apply the C4.5 algorithm to classify the risk of malnutrition in toddlers based on anthropometric data and basic information of toddlers at the Gunung Meriah Community Health Center (UPTD) from October 2025 to January 2026, measure the model's accuracy using a confusion matrix, and identify the most influential attributes based on the formation of a decision tree. This research is important to support

early identification of toddlers at risk of malnutrition, assist health workers in determining treatment priorities, and optimize the use of integrated health post (Posyandu) data as a basis for decision-making. The novelty of this study lies in the application of the C4.5 algorithm to classify the risk of malnutrition in toddlers using local data from the Gunung Meriah Community Health Center (UPTD), with a focus on measuring model accuracy and identifying attributes that influence the classification results. Thus, this study is expected to produce a classification model that is appropriate to the characteristics of the available data and supports the process of identifying malnutrition risks more systematically.

Method

This study uses a data mining classification approach with the C4.5 algorithm to identify the risk of malnutrition in toddlers at the Gunung Meriah Community Health Center (UPTD). The C4.5 algorithm is a decision tree-based classification method used to find patterns in a data set and generate understandable classification rules. This method builds a decision tree based on attributes that have the ability to distinguish target classes, so that the results can be represented in the form of decision structures or if-then rules (Quinlan, 1993).

The attributes used in this study include age, gender, weight, height, and mid-upper arm circumference (MUAC), while nutritional status is used as the classification target. Determination of nutritional status refers to the recording data available at the community health center. In the context of child anthropometric assessment, the interpretation of nutritional status needs to pay attention to the applicable anthropometric indicators and standards as stipulated in the Regulation of the Minister of Health of the Republic of Indonesia Number 2 of 2020 concerning Child Anthropometric Standards (Ministry of Health of the Republic of Indonesia, 2020). The application of the C4.5 algorithm is expected to produce a classification pattern based on the characteristics of toddlers contained in the research dataset.

The research instruments included documentation of toddler nutritional status data obtained from the Gunung Meriah Community Health Center's Technical Implementation Unit (UPTD) and software used to process the data and implement the web-based system. The documentation served as a source of research data, while the software supported data processing, the application of classification algorithms, and the presentation of classification results.

The data analysis technique is carried out through several stages. The first stage is data preprocessing, which includes checking data consistency, data cleaning, handling empty values, and data transformation according to analysis requirements. This stage is necessary to prepare the dataset for use in the classification model development process. Next, the C4.5 algorithm is applied by calculating the entropy, information gain, split information, and gain ratio values to determine the attributes used in forming decision tree nodes. Attribute selection is carried out based on the results of these calculations until a decision tree structure is formed that produces classification rules (Quinlan, 1993).

The next stage is evaluating model performance by comparing the predicted results to actual nutritional status. This evaluation is conducted to determine the accuracy of the model's classification results. One measure that can be used is accuracy, which is the proportion of correct predictions to the total data evaluated. However, because the research dataset contains varying amounts of data for each nutritional status category, the evaluation results also need to consider the model's performance for each class so that the ability to identify categories at risk of malnutrition and malnutrition is not solely assessed based on overall accuracy.

The research data was sourced from records of toddler nutritional status at the Gunung Meriah Community Health Center (UPTD) from 2025 to 2026. The dataset used in this study consisted of 386 toddlers, consisting of 343 children categorized as well-nourished, 26 at risk of malnutrition, and 17 with severe malnutrition. All of this data was used as analysis material to develop a classification model using the C4.5 algorithm.

In this study, toddler data served as the unit of analysis, while age, gender, weight, height, and MUAC were used as predictor variables, with nutritional status as the target variable. The use of anthropometric data needs to be adjusted to applicable child nutritional status assessment standards to ensure a clear basis for classification (Ministry of Health of the Republic of Indonesia, 2020).

The total population size and sampling technique were not explained in the research materials

provided. Therefore, this study only lists the dataset used, namely 386 toddlers, without specifying a specific sampling technique, which is not supported by the research information.

The research procedure was carried out in stages, starting with problem identification, data collection, data preprocessing, classification model development, system implementation, and evaluation of the research results. These stages were designed to ensure that the data processing and application of the C4.5 algorithm were carried out systematically.

The first stage is problem identification, which assesses the need to identify the risk of malnutrition in toddlers at the Gunung Meriah Community Health Center (UPTD). The second stage is data collection through documentation of toddler nutritional status records for the period 2025 to 2026. The third stage is data preprocessing, which includes checking for completeness, cleaning, handling blank values, and transforming the data according to analysis requirements.

The fourth stage is determining the attributes and classification targets, followed by applying the C4.5 algorithm through calculations of entropy, information gain, split information, and gain ratio. The calculation results are used to form a decision tree and generate classification rules based on the characteristics of the toddler data. This process follows the principle of decision tree formation in the C4.5 algorithm, namely selecting appropriate attributes to separate data into specified classes (Quinlan, 1993).

The fifth stage is the design and implementation of a web-based system to integrate data processing and the presentation of classification results. The sixth stage is testing the system model and functions to assess the suitability of the classification results and ensure the system functions as required. The research concludes with an evaluation of the classification results, drawing conclusions, and compiling research results based on the findings.

Results and Discussion

Model and Dataset Implementation

1. Property Dataset Implementation

The dataset used in this study is toddler data obtained from the Gunung Meriah Community Health Center (UPTD) and used as data in the malnutrition risk classification process using the C4.5 algorithm. The dataset consists of attributes such as age, gender, weight, height, and MUAC. The data is then compiled and checked before being used in the classification process. The prepared dataset is then used as a data source for processing and developing a classification model.

Figure 1. Toddler Dataset

Figure 1 shows the toddler dataset used in the study as input data for the classification process. The dataset contains several attributes used to determine the nutritional status of toddlers, namely age, gender, weight, height, and MUAC. Each toddler data item has a value. Each attribute is then processed using the C4.5 algorithm to determine the classification category: well-nourished, at risk of malnutrition, or severely malnourished. Once the dataset is compiled, it is entered into a MySQL database for management and use by the system. The database serves as a storage location for toddler data, which will later be used as a data source in the classification process.

	jenis_kelamin	umur_bulan	berat	tinggi	lila	status_gizi	indeks_balita
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000
♀	♀	2	1.1	65.5	18	Car-Buk	85.000

Figure 2. Toddler dataset in MySQL

The system then connects to the MySQL database and retrieves data from the toddler table using the Python programming language. This process ensures that the data stored in the database can be read and used in the preprocessing stage and the development of a classification model using the Decision Tree C4.5 algorithm.

```

11 # KONEKSI DATABASE
12 conn = pymysql.connect(
13     host="localhost",
14     user="root",
15     password="",
16     database="gizi_balita"
17 )
18
19 # AMBIL DATA DARI DATABASE
20 query = """
21 SELECT
22     jenis_kelamin,
23     umur_bulan,
24     berat,
25     tinggi,
26     lila,
27     status_gizi
28 FROM balita
29 """
30
31 df = pd.read_sql(query, conn)

```

Figure 3. Implementation of reading dataset from MySQL

Data Processing Implementation

In the preprocessing stage, data retrieved from the MySQL database needs to be transformed to be used in the classification model development process. This process is crucial because the Decision Tree C4.5 algorithm requires numeric input, while the dataset still contains categorical attributes. The data transformation process is performed using the sklearn.preprocessing library in Python, specifically the LabelEncoder() function. LabelEncoder converts categorical data into numeric representations. In this study, two attributes underwent the encoding process: the gender attribute and the nutritional status attribute.

```

# ENCODING DATA

encoder_jk = LabelEncoder()
encoder_status = LabelEncoder()

df["jenis_kelamin"] = encoder_jk.fit_transform(df["jenis_kelamin"])
df["status_gizi"] = encoder_status.fit_transform(df["status_gizi"])

```

Figure 4. Implementation of Data Preprocessing

Figure 4 shows the encoding process performed on the dataset. The fit_transform() method is used to adapt the encoder pattern to the data and convert it to a numeric representation. The result of this process is numeric data ready to be used as input in the classification model training stage.

Implementation of the C4.5 Decision Tree Algorithm

The clustering model used in this research applied the C4.5 Decision Tree method, operated in Python using the Scikit-learn library. This step was intended to establish a pattern of relationships between the characteristics of children under five and their nutritional status. The characteristics used as input variables included gender, age in months, weight, height, and mid-upper arm circumference

(MUAC). The output or target variable to be predicted was the nutritional status of toddlers, which was divided into three classes: Well-Nutritoned, At-Risk for Malnutrition, and Severely Malnourished.

The dataset used in the model training process consists of [number of data points] of toddler data that has undergone a previous preprocessing stage. All of this data is used to train the model to recognize patterns and produce accurate classification decisions.

```
# MENENTUKAN FITUR DAN TARGET
X = df[[
    "jenis_kelamin",
    "umur_bulan",
    "berat",
    "tinggi",
    "lila"
]]
y = df["status_gizi"]

# MEMBUAT MODEL C4.5
model = DecisionTreeClassifier(
    criterion="entropy",
    random_state=42
)
```

Figure 5. Implementation of the Decision Model Three c4.5

Figure 5 shows that the model formation process is carried out by defining the X variable as a feature set and the y variable as the target label. Next, the DecisionTreeClassifier() object is initialized with the criterion="entropy" parameter, which is a characteristic of the C4.5 algorithm in calculating the best attribute selection. The fit() process is then run to train the model using the prepared data. The result of this stage is a decision tree model ready to be used to classify new toddler data.

Rule Decision Tree C4.5

```
|--- lila <= 11.95
|   |--- berat <= 8.10
|   |   |--- class: Gizi Buruk
|   |--- berat > 8.10
|   |   |--- class: Gizi Baik
|--- lila > 11.95
|   |--- berat <= 9.15
|   |   |--- class: Beresiko Gizi Buruk
|   |--- berat > 9.15
|   |   |--- class: Gizi Baik
```

Figure 6. Rule Decision Tree

Based on Figure 6, the resulting decision tree rule shows that the LILA (Upper Arm Circumference) attribute is the root node or main node in decision making. If the LILA value is less than or equal to 11.95 cm, the system will check the weight attribute. If the weight is less than or equal to 8.10 kg, the toddler is classified as having Malnutrition status. However, if the weight is more than 8.10 kg, the toddler is classified as having Good Nutrition status. In another branch, if the LILA value is more than 11.95 cm, the system checks the weight again. If the weight is less than or equal to 9.15 kg, the toddler is classified as having Risk of Malnutrition status. Meanwhile, if the weight is more than 9.15 kg, the toddler is classified as having Good Nutrition status. From the results of the decision tree formation, it can be seen that the LILA attribute is the most dominant factor in determining the nutritional status of toddlers, followed by the weight attribute as a supporting factor. These rules are then used by the system to automatically classify new toddler data entered by the user.

System Interface Implementation

The designed web application was then implemented using PHP and MySQL. The entire interface is designed to allow users to easily execute each provided function. There are six main, interconnected pages: the login page, the home page, the toddler data management page, the analysis process page, the results display page, and the report download page. Each page has its own role in supporting the overall system workflow.

1. System Admin Interface

A. Admin Login Page

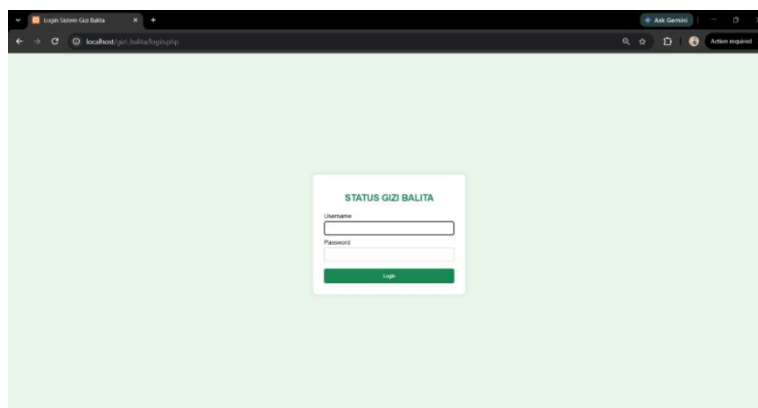


Figure 7. Admin Login Page

Figure 7. Admin Login Page: Admins are prompted to enter their username and password in the fields provided. After entering the information, the admin presses the login button to access the system. If the username and password entered are correct, the system will direct the user to the dashboard. If incorrect, the system will display a warning message and ask the user to re-enter their login details.

B. Admin Dashboard Page

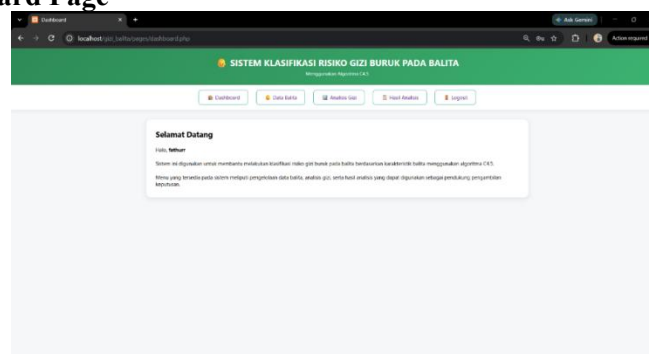


Figure 8. Admin Dashboard Page

Figure 8: The Admin dashboard page is the main page displayed after a user successfully logs into the system. On this page, the admin is greeted with a display containing several navigation menus for accessing available features. These menus include Dashboard, Toddler Data, Nutritional Analysis, Analysis Results, and Logout.

C. Toddler Data Page Admin

No	Kode Balita	Jenis Kelamin	Umur (Bulan)	Berat (kg)	Tinggi (cm)	LILA (cm)	Status Gizi	Aksi
1	007001	S	20	12.5	85.3	16	Cair Baki	Edit Delete
2	007002	S	25	9.4	82.7	12	Berisiko Gizi Buruk	Edit Delete
3	007003	S	5	6.4	61.3	14	Cair Baki	Edit Delete
4	007004	S	42	14.6	95.4	15.5	Cair Baki	Edit Delete
5	007005	S	11	9	75	12.6	Cair Baki	Edit Delete
6	007006	S	38	17	98.9	15	Cair Baki	Edit Delete
7	007007	S	46	11	94	14	Cair Baki	Edit Delete
8	007008	S	40	15.2	94.9	13	Cair Baki	Edit Delete
9	007009	S	18	11.5	78	13	Cair Baki	Edit Delete
10	007010	S	3	5.5	68.7	12	Cair Baki	Edit Delete
11	007011	S	30	15.2	101	16	Cair Baki	Edit Delete
12	007012	S	7	5.7	75.4	14	Cair Baki	Edit Delete
13	007013	S	7	5.5	68.6	15	Cair Baki	Edit Delete

Figure 9. Toddler Admin Data Page

The toddler data page displays all toddler data stored in the database. The data displayed includes the toddler code, gender, age in months, weight, height, mid-upper arm circumference (MUAC), and nutritional status of each toddler. All data is presented in tabular form, making it easy for users to view and review each stored information. At the top of the page, there is an "Add Data" button that functions to add new toddler data to the system. In addition, on each data row there are also "Edit" and "Delete" buttons to make changes or delete existing data. This page serves as the central

Classification of Malnutrition Risk in Toddlers Using the C4.5 Method (Case Study: UPTD Gunung Meriah Health Center) (Fathur Rizqi Hasugian), Page 615

management of toddler data which will later be used as training data in the classification process using the C4.5 algorithm.

D. Admin Nutrition Analysis Page

Figure 10. Admin Nutrition Analysis Page

Figure 10: The nutritional analysis page is used to classify the nutritional status of toddlers. On this page, the admin can input the toddler's data to be analyzed. The required data includes gender (selected as Male or Female), age in months, weight in kilograms, height in centimeters, and mid-upper arm circumference (MUAC) in centimeters. Once all data is filled in, the admin can press the "Nutrition Analysis" button to begin the classification process. The system will then process the entered data using the previously trained Decision Tree C4.5 model to determine the toddler's nutritional status. Additionally, a "Back" button is available to direct the admin back to the dashboard if they wish to cancel the analysis process.

E. Admin Analysis Results Page

No	Jenis Kelamin	Usia	Berat	Tinggi	LILA	Status Gizi	Tanggal Analisis	Aksi
1	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
2	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
3	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
4	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
5	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
6	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
7	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
8	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
9	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
10	P	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print
11	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17	Print

Figure 11. Admin Analysis Results Page

Figure 11. The analysis results page displays the entire history of toddler nutritional status classifications that have been performed. The data displayed includes the serial number, gender, age, weight, height, mid-upper arm circumference (MIAC), nutritional status classification results, and the date and time of the analysis. All information is presented in a neatly arranged table, making it easy for Admins to view and review each saved analysis result. At the top of the page, there is a "Print Report" button that functions to print or download all analysis results as a PDF document. Furthermore, each row of data also has a "Delete" button that can be used to delete analysis history that is no longer needed. This page allows Admins to monitor the development of toddler nutritional status over time and use it as evaluation material in decision-making.

F. Admin PDF Report Results

No	Jenis Kelamin	Usia	Berat	Tinggi	LILA	Status Gizi	Tanggal
1	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
2	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
3	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
4	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
5	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
6	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
7	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
8	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
9	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
10	P	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17
11	L	15	15.5	85	10	Gizi Baik	2024-01-11 14:11:17

Figure 12. Admin PDF Report Results

Figure 12. This page is used to print or download all analysis data in PDF document format.

2. User System Interface

A. User Login Page

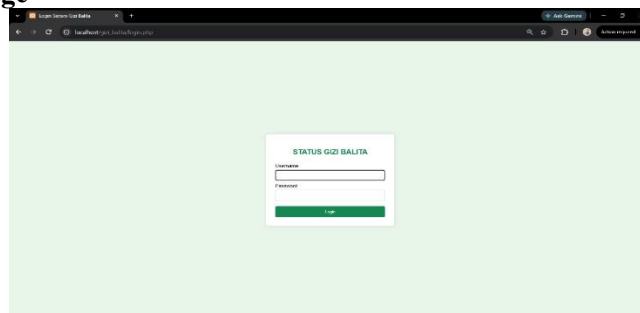


Figure 13. User login page

Figure 13 The user login page is the first page displayed when a user opens the application. On this page, the user is required to enter their username and password in the provided fields.

B. User Dashboard Page

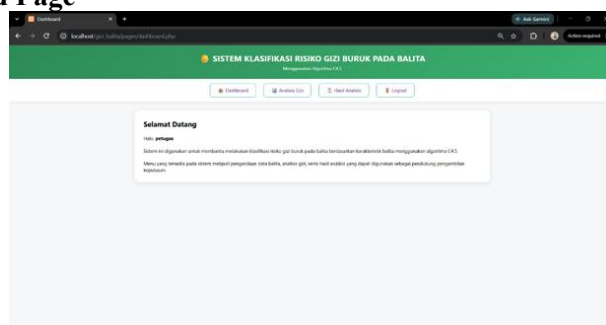


Figure 14. User Dashboard Page

Figure 14. User Dashboard page. After successfully logging in, the user will be directed to the dashboard page. This page displays menu options that the user can access based on their access authority. Options include Dashboard, Nutrition Assessment, Assessment Results, and Exit.

C. User Nutrition Analysis Page

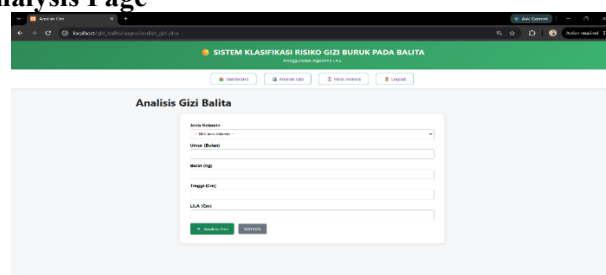


Figure 15. User Nutrition Analysis Page

Figure 15. The user's nutritional analysis page functions to classify the nutritional status of toddlers. The user is asked to fill out a form containing the toddler's data to be analyzed, including gender, age, weight, height, and LiLA. Once all the data is filled in, the user presses the "Nutrition Analysis" button to begin the classification process.

D. User Analysis Results Page

Figure 16. Nutritional Analysis Results Page X

No	Jenis Kelamin	Usia	Berat	Tinggi	LILA	Status Gizi	Tanggal Analisis	Hasil
1	L	11	15	85	20	Gizi Buruk	2023-01-10 14:00:00	Print
2	L	10	12	77	22	Gizi Buruk	2023-01-11 15:00:00	Print
3	L	10	7	77	11	Gizi Buruk	2023-01-07 16:00:00	Print
4	L	12	5	7	5	Gizi Buruk	2023-09-27 15:00:00	Print
5	L	20	6.5	55	11.5	Berisiko Gizi Buruk	2023-09-27 15:00:00	Print
6	L	2	18	80	30	Gizi Buruk	2023-09-27 15:00:00	Print
7	L	2	11	25	11	Gizi Buruk	2023-09-27 15:00:00	Print
8	L	20	6.5	55	11.5	Berisiko Gizi Buruk	2023-09-27 15:00:00	Print
9	L	20	13	25	11.5	Gizi Buruk	2023-09-27 15:00:00	Print
10	L	20	18	82	35.5	Gizi Buruk	2023-09-27 15:00:00	Print
11	L	20	12.5	81	15.5	Gizi Buruk	2023-09-27 15:00:00	Print

Figure 16. The User Nutrition Analysis page displays the entire classification history of the user. Data displayed includes the number, gender, age, weight, height, LiLA, nutritional status, and analysis date. There is a "Print Report" button to download the results in PDF format.

E. User PDF Report Results

No	Jenis Kelamin	Usia	Berat	Tinggi	LILA	Status Gizi	Tanggal Analisis	Hasil
1	L	11	15	85	20	Gizi Buruk	2023-01-10 14:00:00	Print
2	L	10	12	77	22	Gizi Buruk	2023-01-11 15:00:00	Print
3	L	10	7	77	11	Gizi Buruk	2023-01-07 16:00:00	Print
4	L	12	5	7	5	Gizi Buruk	2023-09-27 15:00:00	Print
5	L	20	6.5	55	11.5	Berisiko Gizi Buruk	2023-09-27 15:00:00	Print
6	L	2	18	80	30	Gizi Buruk	2023-09-27 15:00:00	Print
7	L	2	11	25	11	Gizi Buruk	2023-09-27 15:00:00	Print
8	L	20	6.5	55	11.5	Berisiko Gizi Buruk	2023-09-27 15:00:00	Print
9	L	20	13	25	11.5	Gizi Buruk	2023-09-27 15:00:00	Print
10	L	20	18	82	35.5	Gizi Buruk	2023-09-27 15:00:00	Print
11	L	20	12.5	81	15.5	Gizi Buruk	2023-09-27 15:00:00	Print

Figure 17. PDF Report Results

Figure 17 shows the feature for downloading nutritional analysis reports in PDF format. This feature is used to save them as a document so they can be used as report archives or documentation.

System Testing

System testing was conducted to determine whether the website could perform each function according to the predetermined design. This stage involved testing the system's main features and observing the results after users completed a process. This study utilized functional testing using black box testing and classification results using a confusion matrix. These tests were used to ensure the system's usability and to determine the C4.5 algorithm's ability to classify the risk of malnutrition in toddlers.

1. System Testing Using Black Box Testing

Table 1. Black Box Testing

No	Features tested	Testing Scenario	Expected results	Test Results
1	Login	Enter the appropriate username and password	The system validates the data and directs the user to the dashboard.	Succeed
2	Dashboard	Open the dashboard page after logging in	The system displays the main page along with the available menus.	Succeed
3	Toddler Data	Add, edit, delete toddler data	The system successfully adds, edits and deletes,	Succeed
4	Classification Process	Running the toddler data classification process	The system processes data using the C4.5 algorithm and generates a risk level.	Succeed

5	Classification Results	Opening classification results	The system displays the results of the classification of malnutrition risk.	Succeed
6	Report	Press the print/download report button	The system generates reports according to the classification results.	Succeed
7	Logout	Press the logout button	The system ends the user session and returns to the login page.	Succeed

Classification Results Testing Using Confusion Matrix

Classification testing was conducted to determine the extent to which the Decision Tree C4.5 model was able to accurately classify the nutritional status of toddlers. This testing used a confusion matrix method that compares the system's predictions with the actual data in the dataset. The dataset used consisted of 386 toddlers, then divided into training data and testing data with a ratio of 80:20, meaning 80% of the data was used to train the model and 20% of the data was used to test the model.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.tree import DecisionTreeClassifier
# Baca data
df = pd.read_csv('balitaa.csv', delimiter=';', encoding='utf-8-sig')

df['lila'] = df['lila'].astype(str).str.replace(',', '.').astype(float)

# MANUAL ENCODE status_gizi
status_map = {
    'Gizi Baik': 0,
    'Gizi Buruk': 1,
    'Beresiko Gizi Buruk': 2
}
df['status_gizi_encoded'] = df['status_gizi'].map(status_map)

df['jenis_kelamin_encoded'] = df['jenis_kelamin'].map({'L': 0, 'P': 1})

X = df[['jenis_kelamin_encoded', 'umur_bulan', 'berat', 'tinggi', 'lila']]
y = df['status_gizi_encoded']
# Split data (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Training model C4.5 (Decision Tree dengan entropy)
model = DecisionTreeClassifier(criterion='entropy', random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print("="*60)
print("KLASIFIKASI STATUS GIZI BALITA DENGAN C4.5")
print("="*60)
print(f"Akurasi: {accuracy_score(y_test, y_pred) * 100:.2f}%")
print("\nConfusion Matrix:")
print(confusion_matrix(y_test, y_pred))
print("="*60)
```

Figure 18. Accuracy and Confusion Matrix Testing Script

Figure 18 shows the test script used to test the performance of the Decision Tree C4.5 model in classifying the nutritional status of toddlers. The testing process begins by reading the dataset from a CSV file, then performing data preprocessing such as encoding categorical attributes and cleaning the LILA data. Next, the dataset is divided into training data and test data with a ratio of 80:20 using the `train_test_split` function. The Decision Tree C4.5 model is built with the `criterion='entropy'` parameter, which is a characteristic of the C4.5 algorithm. After the model is trained, predictions are made on the test data and the results are evaluated using a confusion matrix and accuracy score to determine the model's accuracy level.

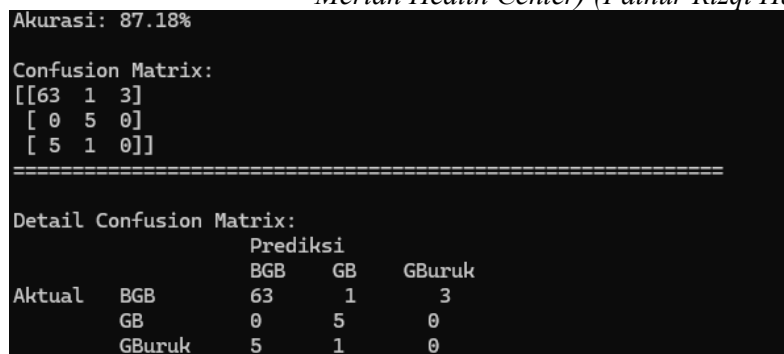


Figure 19. Confusion Matrix and Accuracy

Figure 19 shows the results of testing using a confusion matrix, which yielded a model accuracy rate of 87.18%. This result indicates that the C4.5 algorithm has good capability in classifying toddler nutritional status into three categories: Well-Nutritoned, At-Risk for Malnutrition, and Severely Malnourished.

Conclusion

This study successfully applied the Decision Tree C4.5 algorithm to classify the nutritional status of toddlers at the Gunung Meriah Community Health Center (UPTD) based on the attributes of gender, age, weight, height, and mid-upper arm circumference (MIAC). Based on the results of the decision tree formation, MIAC was the primary attribute in the classification process, followed by weight as a supporting attribute. The model produced three classification categories: good nutrition, at risk of malnutrition, and severe malnutrition, with an accuracy rate of 87.18% based on testing using a confusion matrix with an 80:20 split between training and test data. In addition, the developed web-based system was able to support toddler data management, nutritional status analysis, presentation of classification results, and printing reports in PDF format. Black Box Testing results showed that all main functions of the system ran according to the test scenario. These findings indicate that the application of the C4.5 algorithm has the potential to help health workers process toddler data more systematically and support early identification of toddlers requiring attention regarding their nutritional status. However, the classification results still need to be understood as a decision-making tool, not a substitute for examination and assessment of nutritional status by health workers based on applicable anthropometric standards.

This study was limited by the use of 386 toddler data from a single health care facility, so the model's generalizability to other regions or populations cannot be confirmed. Furthermore, the unbalanced class distribution, with 343 well-nourished children, 26 at-risk of malnutrition, and 17 severely malnourished, may have impacted the model's ability to recognize minority categories. Evaluations that focus solely on accuracy are also insufficient to ensure the accuracy of identifying each category, particularly toddlers at risk of or experiencing malnutrition. Future research is recommended using more diverse data from multiple health facilities, implementing cross-validation, and evaluating precision, recall, F1-score, and confusion matrix for each class. Comparisons with other classification algorithms are also needed to determine which method provides better performance. Practically, the developed system can be further tested through direct use by health workers to assess ease of use, consistency of results, and its usefulness in supporting toddler growth monitoring. Further development should also ensure data quality, the application of appropriate anthropometric standards, and the protection of patient information confidentiality so that the system can be used responsibly in health care.

Bibliography

Hutabarat, M. J., Simatupang, D. S., Tarigan, M., & Yulhendri. (2026). Nutritional status classification of children using a decision tree method (Case study at RS AN-NISA) based on a web application. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 16(1), 165–174. <https://doi.org/10.32664/smatika.v16i01.2238>

- Kementerian Kesehatan Republik Indonesia. (2020). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 tentang standar antropometri anak*. <https://jdih.kemkes.go.id/documents/peraturan-menteri-kesehatan-nomor-2-tahun-2020>
- Kementerian Kesehatan Republik Indonesia. (2025, May 27). *SSGI 2024: Prevalensi stunting nasional turun menjadi 19,8%*. Badan Kebijakan Pembangunan Kesehatan. <https://www.badankebijakan.kemkes.go.id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198/>
- Laksono, A. D. (2024). Determination of appropriate policy targets to reduce the prevalence of stunting in children under five years of age in urban-poor communities in Indonesia: A secondary data analysis of the 2022 Indonesian national nutritional status survey. *BMJ Open*, 14(9), e089531. <https://doi.org/10.1136/bmjopen-2024-089531>
- Matondang, T. R., Nasution, Y. R., Armansyah, & Furqan, M. (2024). Penerapan algoritma C4.5 pada klasifikasi status gizi balita. *Jurnal Fasilkom*, 14(1), 216–225. <https://doi.org/10.37859/jf.v14i1.6941>
- Nuraisana, N., Halawa, S. W., & Harun, M. (2024). Implementasi algoritma C4.5 dalam mengklasifikasi status gizi balita pada Posyandu Desa Sekip Lubuk Pakam. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, 23(1), 130. <https://doi.org/10.53513/jis.v23i1.9596>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106. <https://doi.org/10.1007/BF00116251>
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-27846-9>
- World Health Organization. (2025a). *Child growth standards*. <https://www.who.int/news-room/questions-and-answers/item/child-growth-standards>
- World Health Organization. (2025b). *Levels and trends in child malnutrition: UNICEF/WHO/World Bank Group joint child malnutrition estimates: Key findings of the 2025 edition*. <https://www.who.int/publications/i/item/9789240112308>