Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 9, No. 2, Juni 2024 (46-55) **https://doi.org/10.32493/informatika.v9i2.40983**

Application of the K-Nearest Neighbor (KNN) Algorithm for Stunting Diagnosis in Infants Aged 1-12 Months

Moh. Abdul Kholik[1*], Cucut Hariz Pratomo[2], Sapriani Gustina[3]

[1]Sistem Komputer, Universitas Surakarta, Surakarta, Indonesia, 57731
e-mail: [1]mak240997@gmail.com

[2] Teknik Informatika, Universitas Muhammadiyah Karanganyar, Karanganyar, Indonesia, 57761
e-mail: [2]cucutharizpratomo@gmail.com,

[3] Teknologi Informasi, Fakultas Teknik, Universitas Proklamasi 45, Yogyakarta, Indonesia, 55281
e-mail: [3]sagustina@up45.ac.id
*Corresponding author

Submitted Date: June 18th, 2024 Reviewed Date: June 27th, 2024
Revised Date: July 3th, 2024 Accepted Date: July 30th, 2024

Abstract

Stunting in toddlers must be addressed immediately because it has a negative impact on their growth and development. Stunting is a disorder where toddlers experience chronic malnutrition, thus their physical growth and height do not match their age. According to the Indonesian Nutritional Status Survey (SSGI), stunting is more common among toddlers from aged 0 to 1 year than overall. Stunting can have short-term and long-term impacts. This research examines data from the Temanggung District Health Service on 3,999 toddlers aged 0 to 12 months between 2019 and 2022. Many studies have exclusively looked at stunting in children aged one to five years, especially research on stunting using the KNN method, even though stunting can actually be recognized from an early age. Therefore, researchers are more specific in using the KNN method for cases of babies 1 to 12 months so as to differentiate it from previous researchers. The aim of this research is to use the K-Nearest Neighbor (KNN) algorithm to detect stunting nutritional status in toddlers. K-Nearest Neighbor (KNN) is a classification algorithm that uses a set of K values from the closest data (its neighbors) as a reference to determine the class of incoming data. KNN classifies data based on its similarity or closeness to other data. The dataset used includes parameters of age, gender and height. The research approach is the CRISP-DM (Cross Industry Standard Process for Data Mining) method, which begins with business knowledge, followed by EDA and modeling, evaluation, testing and report preparation. The result shows that the KNN algorithm can accurately categorize children as stunted or not based on age (U) and height (TB), with the maximum level of accuracy and the lowest error rate at $k = 5$. At this optimal value (k), this algorithm has an accuracy of 99.87%, Recall 99.84%, and precision 99.73.

Keywords: Diagnose; Artificial intelligence; Detection; K-Nearest Neighbor (KNN); Stunting; Machine Learning

1. Introduction

The toddler stage is a critical period in a child's growth and development, during which their cognitive capacities progress significantly, aligning with age-related milestones (Kaesmitan & Johannis, 2017). Adequate nutrition during this time is essential for supporting these developmental processes. Stunting is a long-standing health concern, including poor nutrition, frequent illnesses, early birth, and low birth weight. However, the most common cause of stunting is malnutrition in women during pregnancy, a lack of clean water and sanitation supplies, and the fact that mothers' weight does not grow during pregnancy. (Priyono, 2020).

A child is said to be stunted if her height and body length are minus 2 from the Multicenter Growth Reference Study standard or the median standard deviation of child growth and development standards from WHO. In addition, the Indonesian Ministry of Health states that stunting is defined as children under five whose z-score is less

Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 9, No. 2, Juni 2024 (46-55) **https://doi.org/10.32493/informatika.v9i2.40983**

than -2SD/standard deviation (stunted) and less than -3SD (severely stunted) (Rahayu et al., 2018).

Toddlers are particularly vulnerable to stunting between the ages of 0 and 5 years in Indonesia. The most crucial period is from birth to 2 years old, and referred to as the first 1,000 days of life. This timeframe is characterized by rapid growth and development, making it imperative to ensure that children receive appropriate and sufficient nutrition. Proper nutrition during these first 1,000 days are fundamental to a child's overall health and future potential (Ramadhan, 2019).

Malnutrition during this critical period can result in irreparable damage to both physical growth and cognitive development. Children who suffer from malnutrition are at risk of not reaching their full potential in terms of height, cognitive abilities, and overall health. This situations can have long-term consequences, affecting their academic performance, economic productivity, and quality of life in adulthood (Galler et al., 2021).

Ensuring adequate nutrition during the toddler stage requires a multifaceted approach that includes promoting breastfeeding, providing access to nutrient-rich foods, and implementing educational programs for parents and caregivers about proper child nutrition. Addressing malnutrition effectively can help mitigate the risk of stunting and its associated consequences, ultimately contributing to the well-being and future success of children in Indonesia (Amir et al., 2020).

Effective therapies to prevent and treat stunting include increasing maternal nutrition, promoting exclusive breastfeeding, and giving access to nutrient-rich diets. Additionally, addressing food insecurity and poverty can help improve overall nutrition status and reduce the incidence of stunting in toddlers. Addressing these issues can contribute to healthy growth and development in children. Preventing stunting in toddlers involves a comprehensive approach that covers dietary, social, and economic variables (Ramadhan, 2019).

Riskesdas (2018) reported that the prevalence of stunting among toddlers in Indonesia was around 30.8%, suggesting that nearly one-third of toddlers are affected by stunting. Based on data from the Central Java Health Office in 2022, the Indonesian Nutrition Status Survey (SSGI) found that the stunting rate in Temanggung was 28.9%. This shows that out of every 100 toddlers, 29 of them were stunted (BPS, 2022). This finding exceeds the average incidence of stunting in Central Java, which is 20.8%. According to the Indonesian Nutrition Status Survey (SSGI), the rate of stunting among children aged 0 – 12 months is higher than the general prevalence of stunting. Based on this data, it is expected that the number of children aged 0 – 12 months afflicted by stunting in Temanggung Regency over the previous 3 years (2018–2022) falls between 15,000 and 20,000.

The Temanggung District Government is actively striving to minimize the prevalence of stunting, particularly among infants aged 0 – 12 months. This involves expanding the reach of exclusive breastfeeding during the first 6 months of life, offering nutritious and well-balanced complementary foods (MPASI), ensuring complete immunization for toddlers, improving access to healthcare services for toddlers, and educating parents on the significance of nutrition and child health. With continuing efforts, it is believed that the prevalence of stunting in children under 0–1 years in Temanggung District can be reduced and all children under five can grow optimally (Lukito & Setyaningsih, 2023).

Based on the aforementioned concerns, researchers will determine the nutritional state of stunted toddlers in Temanggung Regency using the K-Nearest Neighbor (KNN) algorithm approach. Currently, the determination of nutritional status is still done by physical inspection and recorded in the KIA book. Age and height are two of the factors used to identify stunted children's nutritional condition (Mustika & Syamsul, 2018). The goal of this study is to analyze the accuracy acquired by the K-Nearest Neighbor (KNN) algorithm in identifying stunted children based on these criteria (Murti et al., 2020). The KNN method was chosen because of its easy calculation and capacity to process a lot of data. In addition, KNN also yields good accuracy when the K value is chosen suitably, because KNN calculates the shortest distance from the test sample to the training sample without taking into account the distribution of each class (Purwati & Sulistyo, 2023).

Previous researchers have discussed stunting. The studies conducted by Ivo Colanus Rally Drajana et al., entitled Predicting the Status of Stunting Sufferers in Toddlers in Gorontalo Province Using K-Nearest Neighbor Based on Chi Square Feature Selection, This studies yielded that a minimal error of RMSE 1,200 (Drajana & Bode, 2022). The research undertaken by Salsabila was entitled Comparison of the K-Nearest Neighbor Algorithm with Euclidean Distance and Manhattan

Distance for Classification of Toddler Stunting. The best distance measurement research results employ k = 3 with an accuracy value of 97.39%, higher than Euclidean distance, which has an accuracy of 95.65% with a difference of 1.74% (Salsabila et al., 2024).

Stunting can develop in tall newborns with insufficient parenting, leading to lower calorie and nutritional intake (Saeful Bachri & Herdian Bhakti, 2021). Furthermore, expectant mothers who suffer from chronic energy deficit (CED) and anemia throughout pregnancy may also cause toddlers to develop stunting (Ruaida & Soumokil, 2018). In the computer sector, detecting the nutritional status of toddlers has been done using an expert system technique with a certainty factor to identify severe malnutrition such as kwashiorkor, marasmus, and marasmic-kwashiorkor (Pantaleon et al., 2016). According to experiments on 120 test data points, the certainty factor's accuracy is over 70% (Anggraeni & Syafrullah, 2023). Naïve Bayes has also been used to identify malnutrition status and the percentage probability of its evaluation through 3 types of malnutrition states and 24 types of symptoms (Wati & Sudrajat, 2022). The grouping of nutritional values of toddlers has also been computed by k-means to group toddlers into malnutrition, undernutrition, good nutrition, and obesity (Irfiani & Rani, 2018). In addition, fuzzy KNN has also been used to classify the nutritional status of children using assessments of health status, parental education, parental knowledge, genetics, and parental income. The accuracy produced by this method is 84.37% (Nugraha et al., 2017).

2. Method

The term research methodology refers to a goal-oriented, methodical framework for research. Every step and procedure in structured research must be adhered to. The primary means of carrying out research are research procedures, which enable a methodical approach to be taken from start to finish (Pratama & Damayanti, 2023). The research procedure, which was conducted in order to identify vulnerability in infants aged between 0- and 12-month-old who were affected by stunting, is presented in Figure 1.

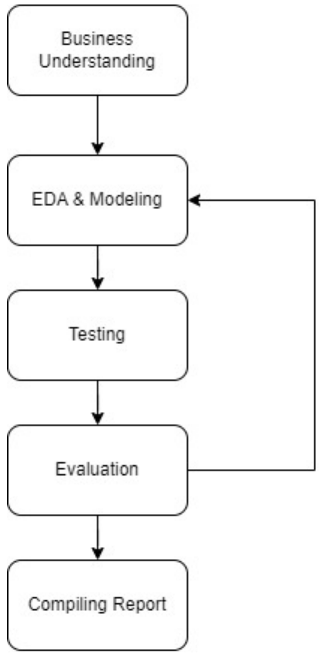

Figure. 1 Diagram of research method

2.1. Business Understanding

The first stage carried out to implement the research objectives is business understanding. This stage includes collecting the problem and determining the method to be employed. At this step, data will also be taken that is appropriate to the study object, specifically data on stunted and non-stunting toddlers for use in the analysis. After that, feature selection is carried out, and the data is reformatted to be entered into the EDA step.

2.2. EDA & Modeling

The next level is exploratory data analysis (EDA). At this step, insights are derived from the data that has been acquired by means of statistical analysis regarding hypothesis testing, finding anomalies, looking for connections between data, centralizing data, and cleaning and correcting data that has null or biased values. After that, modeling is carried out on the clean data. This data will classify K-Nearest Neighbor (KNN). K-Nearest Neighbor (KNN) is a machine learning algorithm that loads all case data and classifies it into new case data based on proximity (Sumarlin, 2015). This algorithm is included in the supervised learning type because it works by connecting the available data patterns with new data sets to find a new pattern (Mukhlis et al., 2024). To calculate the distance, the KNN algorithm generally This is accomplished by measuring the nutritional status of toddlers using TB/U, which can be classified as very short, short, normal, or high. Before

Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 9, No. 2, Juni 2024 (46-55) **https://doi.org/10.32493/informatika.v9i2.40983**

implementation, the system is tested with a confusion matrix to determine the resulting accuracy and error levels. This helps to further confirm that the pattern created is consistent with the KNN classification model. The flow of the KNN algorithm that will be applied is (Saeful Bachri & Herdian Bhakti, 2021):

a. Determine the k value.
b. Calculate the distance between training data and testing data using.

$$d_i = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

c. Sort the distances (in ascending order) and determine the closest distance to k.
d. Match the right classes.
e. Find the largest number of closest neighbors, then choose that category as the category of the data you are looking for.

.

2.3. Testing

After developing the classification model with KNN, testing is carried out to assess whether the model needs to be given more treatment or whether it can be implemented. In this scenario, the proportion of data used is 80% train data and 20% test data.

2.4. Evaluation

At this stage, it is used to see whether the model we have chosen is the optimal algorithm to apply by carrying out accuracy tests. We can determine the model's advantages, disadvantages, and performance by conducting an evaluation. After testing, it can be determined whether the model needs to be given more treatment or can be implemented. In this scenario, the proportion of data used is 80% training data and 20% test data. Next, carry out a confusion matrix test, which is a test designed to predict the possibility that an object will be declared true or false (Romadloni et al., 2022). The test sequence is presented in the form of a confusion matrix, with prediction classes presented at the top and observation classes presented on the left. Each cell contains a numeric value indicating the actual number of cases in the class of observations to be predicted predicted (Batista et al., 2021).

The calculation formula for the confusion matrix is as follows:

a. *Specificity* is used in is used in measuring the success rate of a method in correctly classifying negative classes:

$$Specificity\ \frac{TN}{FP+TN}\ x\ 100\%$$

b. *Recall* or *Sensitivity* is used to measure how good a method is at correctly classifying positive classes:

$$Recall\ \frac{TP}{FP+TN}\ x\ 100\%$$

c. *Accuracy* or Recognition rate is useful for measuring the performance of a method:

$$Accuracy\ \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\%$$

d. *Precision*

$$Precision\ \frac{TP}{Fp+TP}\ x\ 100\%$$

2.5. Compiling Report

The final stage of this research is the preparation of the final research outcomes report. The contents of this report include the backdrop of the problem, the basis, and the findings. At this stage, an evaluation of the overall research outcomes is also carried out.

3. Result and Discusion

The dataset utilized in this research is comprised of comprehensive data obtained from the Temanggung District Health Office, Central Java Province, focusing on the nutritional status of toddlers within the region. This dataset is meticulously compiled into a file with a CSV (Comma Separated Values) extension, ensuring ease of use and compatibility with various data analysis tools.

The dataset includes a substantial volume of data, encompassing a total of 3,998 data points. Within this dataset, there are 328, 766 severely stunted, 1701 tall, 1204 normal instances of stunting identified. This indicates that the dataset not only provides a broad overview of the nutritional status but also highlights specific cases of malnutrition, particularly stunting, which is a critical health concern.

Jurnal Informatika Universitas Pamulang
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang
Vol. 9, No. 2, Juni 2024 (46-55)
ISSN: 2541-1004
e-ISSN: 2622-4615
https://doi.org/10.32493/informatika.v9i2.40983

Furthermore, the data sample employed in this study specifically targets toddlers aged between 0 and 12 months. This age range is crucial for assessing early childhood nutritional status, as it is a period of rapid growth and development where proper nutrition is vital. By focusing on this age group, the research aims to provide insights into the nutritional challenges and interventions needed during the first year of life.

Figure 2 visually presents the dataset, offering a clear and concise representation of the information contained within the CSV file. This visual aid is intended to enhance the understanding of the dataset's structure and the distribution of nutritional status among the toddlers studied.

	Age (month)	Gender	Height (cm)	Nutritional status
0	0	laki-laki	44.591973	stunted
1	0	laki-laki	56.705203	tall
2	0	laki-laki	46.863358	normal
3	0	laki-laki	47.508026	normal
4	0	laki-laki	42.743494	severely stunted
...	...	...	...	...
3994	1	perempuan	51.126175	normal
3995	1	perempuan	64.374831	tall
3996	1	perempuan	68.200071	tall
3997	1	perempuan	60.757330	tall
3998	1	perempuan	58.819450	normal

Figure. 2 Data on the number of stunting patients in Temanggung Regency

Data preparation is a key step in knowledge discovery, especially when the data obtained is incomplete (contains missing values), noisy (contains errors), inconsistent, or in different formats. The toddler nutritional status test data utilized in this study does not include missing values, allowing it to progress to the next step, as illustrated in Figure 3.

```
Age (month)          0
Gender               0
Height (cm)          0
Nutritional status   0
dtype: int64
```

Figure. 3 checking missing value

Figure 3 demonstrates how to use the formula `df.isnull().sum()` in Google Colab to identify any missing data within the dataset. This formula checks for the presence of empty cells across all columns, and in this case, it returns a value of 0, indicating that there are no missing values in the dataset. This outcome confirms that the data is complete and reliable for further analysis.

As a result, the nutritional status data of toddlers in Temanggung Regency, as gathered by researchers, is confirmed to be free of any missing values. This completeness ensures that the dataset is robust and can be utilized effectively for subsequent analytical steps without the need for additional data imputation or correction.

The subsequent step involves transforming the nutritional status information into a numeric code. This transformation is essential for facilitating the processing and analysis of the data, as numeric codes are more manageable for statistical and machine learning algorithms compared to categorical data. By converting nutritional status into numeric codes, the dataset becomes more suitable for a variety of computational techniques and models.

The results of this data transformation process can be observed in Figure 4. This figure illustrates how the categorical nutritional status information has been successfully converted into numeric codes, ready for further analysis. This transformation not only enhances the dataset's usability but also ensures that the analysis can be conducted more efficiently and accurately.

	Age (month)	Gender	Height (cm)	Nutritional status
0	0	0	44.591973	1
1	0	0	56.705203	3
2	0	0	46.863358	2
3	0	0	47.508026	2
4	0	0	42.743494	0
...	...	...	...	...
3994	1	1	51.126175	2
3995	1	1	64.374831	3
3996	1	1	68.200071	3
3997	1	1	60.757330	3
3998	1	1	58.819450	2

3999 rows × 4 columns

Figure. 3 Data Transformation Result

Before the implementation of the K-Nearest Neighbour (K-NN) technique, it is essential to partition the dataset into two distinct subsets: training data and testing data. This partitioning process ensures that the model can be trained effectively and its performance can be evaluated accurately.

Jurnal Informatika Universitas Pamulang ISSN: 2541-1004
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang e-ISSN: 2622-4615
Vol. 9, No. 2, Juni 2024 (46-55) https://doi.org/10.32493/informatika.v9i2.40983

The training data comprises both independent variables, referred to as variable X, which represent the features or qualities of the data, and dependent variables, referred to as variable Y, which represent the labels or classifications. These variables are used as input for the model training process, where the K-NN algorithm learns the relationships between the features and their corresponding labels.

Concurrently, the testing data is used to evaluate the model's performance. During testing, the model categorizes the labels or classes by calculating the distance between the input features (independent variables or variable X) and the training data. This step is crucial for assessing the accuracy and reliability of the K-NN model.

Before separating the dataset into training and testing subsets, the data was randomized using Google Colab to ensure that the samples were evenly distributed and to prevent any potential biases. Randomizing the dataset is a critical step because it helps create a more robust model by ensuring that the training and testing data are representative of the overall dataset.

Subsequently, the randomized dataset was divided into two independent subsets: the training data and the testing data. The proportion of training data to testing data was fixed at 80:20. This means that 80% of the data was allocated for training the model, while the remaining 20% was reserved for testing its performance.

The outcomes of this data partitioning process, including the specific division of the training and testing data, can be observed in Figure 5. This figure illustrates how the dataset was split, ensuring that both subsets are ready for the next steps in the K-NN implementation, ultimately leading to a more accurate and reliable classification model.

```
X_train shape: (3199, 3)
y_train shape: (3199,)
X_test shape: (800, 3)
y_test shape: (800,)
```

Figure. 4 Train and Testing Data Split

Once all the essential data has been acquired, the next step is to undertake the K-Nearest Neighbour (K-NN) computation method. This step involves using the collected and preprocessed dataset to train and evaluate the K-NN algorithm. The primary purpose of this study is to discover the ideal value of (*k*), which is a crucial parameter in the K-NN algorithm. The (*k*) value determines the number of nearest neighbors to consider when making a classification decision for a given data point.

In order to achieved the objectives, the study systematically tests various (*k*) values to find the one that results in the highest accuracy and the lowest error rate in assessing the nutritional status of toddlers. Accuracy is measured by how correctly the model classifies the nutritional status based on the features provided, while the error rate indicates the proportion of misclassifications made by the model.

The process of finding the optimal (*k*) value involves:

1. Model Training and Testing: For each potential (*k*) value, the K-NN algorithm is trained using the training subset of the dataset. The trained model is then tested on the testing subset to evaluate its performance.
2. Performance Evaluation: The accuracy and error rate of the model are calculated for each (*k*) value. These metrics are used to compare the performance of the model with different (*k*) values.
3. Selection of the Best (*k*): The (*k*) value that yields the highest accuracy and the lowest error rate is considered the optimal choice. This value ensures that the model makes the most accurate predictions regarding the nutritional status of toddlers.

The outcomes of this evaluation procedure, including determination of the best (k) value from the search results for k values from the range of 5 to 20, may be shown in the table 1.

Table 1. results values k

k	**accuration**
5	0.9987
6	0.996
7	0.998
8	0.996
9	0.998
10	0.996
11	0.998
12	0.998
13	0.996
14	0.995
15	0.995

Jurnal Informatika Universitas Pamulang ISSN: 2541-1004
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang e-ISSN: 2622-4615
Vol. 9, No. 2, Juni 2024 (46-55) https://doi.org/10.32493/informatika.v9i2.40983

Figure 5 provides a visual representation of how different (k) values impact the model's performance, highlighting the (k) value that achieves the optimal balance between accuracy and error rate. The study aims to improve the reliability and effectiveness of the K-NN model in accurately classifying toddler nutritional status by selecting the best (k) value.

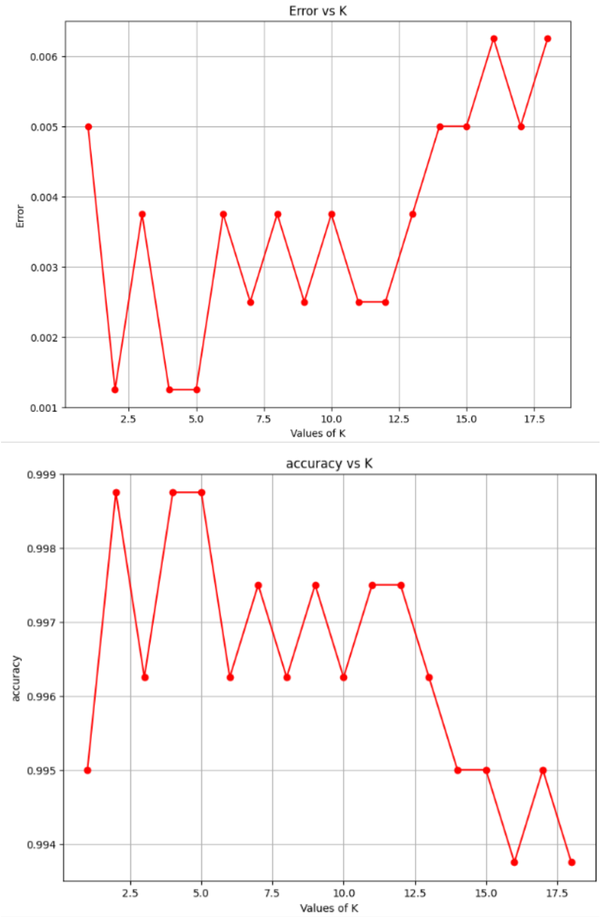

Figure. 5 Search result of the best K value

The figure above illustrates that the optimal value of (k), which yields the highest accuracy, is determined to be (k = 5). At this value, the model achieves the smallest error rate, with an error value of 0,13.

The accuracy of the model at (k = 5) is remarkably high, with an accuracy value of 0.9987. This means that the model correctly classifies 99.9% of the instances in the testing dataset. Such a high accuracy value demonstrates the effectiveness of the K-NN algorithm in this context, particularly in assessing the nutritional status of toddlers based on the given features.

The precise representation of this accuracy value is further validated through the confusion matrix, as illustrated in Figure 7. The confusion matrix provides a detailed breakdown of the model's performance by showing the number of true positive, true negative, false positive, and false negative predictions. This comprehensive view allows for a better understanding of how well the model distinguishes between different classes and where any potential errors might occur.

In summary, the optimal (k) value of 5, with an associated accuracy of 0.9987 and an error rate of 0.13, highlights the robustness of the K-NN model in this study. Figure 6's confusion matrix corroborates these findings, offering a visual and quantitative assessment of the model's performance. This optimal setting ensures that the model can be reliably used to assess the nutritional status of toddlers with a high degree of precision.

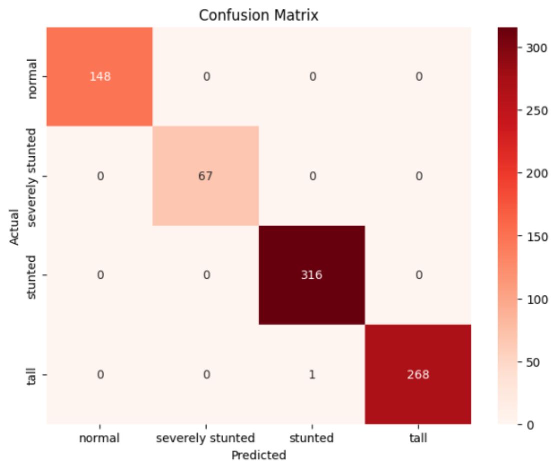

Figure. 6 Confussion Matrix

Figure 6 illustrates the K-NN test findings, which can be explained through the use of a confusion matrix. The confusion matrix gives a complete summary of model performance by providing the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. In Figure 7, it can be observed that the number of babies who have a height above the average is 268, normal babies are 148, infants affected by stunting are 316, and infants who have severe stunting are 67. This thorough representation gives an in-depth assessment of how well the K-NN model classifies the nutritional status of toddlers.

The average values of precision, recall, and accuracy are summarized in Table 2 to give the information provided by the confusion matrix and

Jurnal Informatika Universitas Pamulang ISSN: 2541-1004
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang e-ISSN: 2622-4615
Vol. 9, No. 2, Juni 2024 (46-55) https://doi.org/10.32493/informatika.v9i2.40983

to offer a more detailed analysis of the model's performance. Precision, also known as positive predictive value, measures the proportion of true positive predictions among all positive predictions made by the model. It indicates the model's ability to avoid false positives. Recall, or sensitivity, measures the proportion of true positive predictions among all actual positive cases in the dataset. It reflects the model's ability to identify all relevant cases without missing any.

Additionally, accuracy represents the overall proportion of correct predictions (both true positives and true negatives) made by the model out of all predictions. These metrics provide a nuanced evaluation of the model's performance, highlighting its strengths and potential areas for improvement. In the picture there is a confusion matrix table for the results of searching for accuracy, precision and recall values, for manual calculations it can be as below

$$Accuracy\ \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\%$$

$$:\frac{148+67+316+268}{148+67+316+268+0+0+0+0+0+0+0+0+0+0+0+0+1}\ x\ 100\%$$

$$Accuracy : \frac{799}{800}\ x\ 100\%$$

Accuracy: 0.9987 x 100% = 99.87%

$$Recall\ \frac{TP}{FP+TN}\ x\ 100\%$$

$$:\frac{148+67}{0+0+0+0+0+0+0+0+0+0+0+0+1+148+268}\ x\ 100\%$$

$$Recall : \frac{383}{417}\ x\ 100\%$$

Recall: 0.9184 x 100% = 91.84%

$$Precision\ \frac{TP}{Fp+TP}\ x\ 100\%$$

$$:\frac{148+67+316+268}{0+0+0+0+0+0+0+0+0+0+0+0+1+67+316}\ x\ 100\%$$

$$Precision\ \frac{383}{384}\ x\ 100\%$$

Precision: 0.9973 x 100% = 99.73%

Based on the results of manual calculations, the results of accuracy, recall and precision regarding the effectiveness of the K-NN model in classifying the nutritional status of toddlers can be seen in Table 2.

Parameter	Nilai
precision	99.73%
recall	91.84%
accuracy	99.87%

4. Conclusion

Stunting is a long-standing health concern, including inadequate nutrition, frequent illnesses, early birth, and low birth weight. However, the most prevalent cause of stunting is malnutrition in mothers during pregnancy, a lack of clean water and sanitation supplies, and the fact that and the fact that mothers' weight does not increase during pregnancy. The research focused solely on height and age. Test findings on 3,999 data points reveal that the K-Nearest Neighbor (K-NN) method is very good at predicting children's stunting status based on the attributes of height (TB) and age (U) correctly. Through comprehensive testing and assessment, it was revealed that the best accuracy was reached with a (k) value of 5. At this optimal (k) value, the K-NN algorithm shows excellent performance. The test results show the best accuracy rate of 99.87%, Recall of 91.84, and precision of 99.73 achieved when k = 5. This means that in 99.87% of situations, the model can correctly determine a child's stunting status based on age and height. The high level of accuracy highlights the robustness and trustworthiness of the K-NN method in this specific application. In addition, the error rate at k = 5 is quite low, only 0.13%. This minimal error rate indicates that the classification error is very negligible, further proving the efficacy of the K-NN model. A low error rate is vital for practical applications, as it ensures that the predictions generated by the model are extremely trustworthy and accurate in the decision-making process. These findings underline the capacity of the K-NN algorithm to successfully employ height variables in assessing children's stunting status. Accuracy and error rate metrics jointly illustrate the precision and dependability of the model. By determining an appropriate value (k) of 5, this research provides a firm foundation for applying the K-NN algorithm in similar circumstances, ensuring that children's nutritional

ISSN: 2541-1004
e-ISSN: 2622-4615
https://doi.org/10.32493/informatika.v9i2.40983

status can be monitored and treated effectively. Future research can be further expanded by adding stunting factors according to the WHO and also increasing the data.

References

Amir, Y., Hasneli, Y., & Erika. (2020). Hubungan Pemberian ASI Eksklusif Terhadap Tumbuh Kembang Bayi. *Jurnal Ners Indonesia*, *Vol. 1*(No. 1), 90–98.

Anggraeni, D. P., & Syafrullah, H. (2023). Sistem Pakar Diagnosa Gejala Malnutrisi pada Balita Menggunakan Metode Certainty Factor. *Jurnal Informasi Dan Teknologi*, *5*(4), 67–72. https://doi.org/10.60083/jidt.v5i4.419

Batista, A. F. de M., Miraglia, J. L., Donato, T. H. R., & Filho, A. D. P. C. (2021). Predicting mortality risk in patients with COVID-19 using machine learning. *Smart Health*, *20*(November 2020), 100178. https://doi.org/10.1016/j.smhl.2020.100178

BPS. (2022). Laporan Indeks Khusus Penanganan Stunting 2019-2020. *Badan Pusat Statistik*, 1–63. https://www.bps.go.id/publication/2021/09/08/3b622d713a80363685aef508/laporan-indeks-khusus-penanganan-stunting-2019-2020.html

Drajana, I. C. R., & Bode, A. (2022). Prediksi Status Penderita Stunting Pada Balita Provinsi Gorontalo Menggunakan K-Nearest Neighbor Berbasis Seleksi Fitur Chi Square. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *5*(2), 309–316. https://doi.org/10.32672/jnkti.v5i2.4205

Galler, J. R., Bringas-Vega, M. L., Tang, Q., Rabinowitz, A. G., Musa, K. I., Chai, W. J., Omar, H., Abdul Rahman, M. R., Abd Hamid, A. I., Abdullah, J. M., & Valdés-Sosa, P. A. (2021). Neurodevelopmental effects of childhood malnutrition: A neuroimaging perspective. *NeuroImage*, *231*, 117828. https://doi.org/https://doi.org/10.1016/j.neuroimage.2021.117828

Irfiani, E., & Rani, S. S. (2018). Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, *6*(4), 161. https://doi.org/10.26418/justin.v6i4.29024

Kaesmitan, Y., & Johannis, J. (2017). KLASIFIKASI STATUS GIZI BALITA DI KELURAHAN OESAPA BARAT MENGGUNAKAN METODE K-NEAREST NEIGBOR. *MULTITEK INDONESIA*, *11*, 42. https://doi.org/10.24269/mtkind.v11i1.506

Lukito, D. A., & Setyaningsih, A. (2023). Hubungan Pemberian Asi Eksklusif Dan Ketepatan Pemberian Mp-Asi Terhadap Kejadian Stunting Pada Balita Usia 6-23 Bulan Di Kecamatan Bansari Kabupaten Temanggung. *Jurnal Gizi Dan Dietetik*, *2*(2), 91–99. https://doi.org/10.34011/jgd.v2i2.1804

Mukhlis, I., Pipin, S., Judijanto, L., Reba, F., Mandowen, S., Al-Husaini, M., & Laili, N. (2024). *ALGORITMA PEMBELAJARAN MESIN (Dasar, Teknik, dan Aplikasi)*.

Murti, F. C., Suryati, S., & Oktavianto, E. (2020). Hubungan Berat Badan Lahir Rendah (Bblr)Dengan Kejadian Stunting Pada Balita Usia 2-5 Tahun Di Desa Umbulrejo Kecamatan Ponjong Kabupaten Gunung Kidul. *Jurnal Ilmiah Kesehatan Keperawatan*, *16*(2), 52. https://doi.org/10.26753/jikk.v16i2.419

Mustika, W., & Syamsul, D. (2018). Analisis Permasalahan Status Gizi Kurang Pada Balita di Puskesmas Teupah Selatan Kabupaten Simeuleu. *Jurnal Kesehatan Global*, *1*(3), 127. https://doi.org/10.33085/jkg.v1i3.3952

Nugraha, S. D., Putri, R. R. M., & Wihandika, R. C. (2017). Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita. *" J. Pengemb. Teknol. Inf. Dan Ilmu Komputer*, *1*(9), 925–932.

Pantaleon, M. G., Hadi, H., & Gamayanti, I. L. (2016). Stunting berhubungan dengan perkembangan motorik anak di Kecamatan Sedayu, Bantul, Yogyakarta. *Jurnal Gizi Dan Dietetik Indonesia (Indonesian Journal of Nutrition and Dietetics)*, *3*(1), 10. https://doi.org/10.21927/ijnd.2015.3(1).10-21

Pratama, F., & Damayanti, D. (2023). Implementasi MOORA untuk Seleksi Awal Pemenang Jakarta Tourism Award 2022. *Jurnal Informatika Universitas Pamulang*, *8*(4), 600–610.

Priyono. (2020). Strategi Percepatan Penurunan Stunting Perdesaan (Studi Kasus Pendampingan Aksi Cegah Stunting di Desa Banyumundu, Kabupaten Pandeglang). *Jurnal Good Governance*, *16*(2), 149–174. https://doi.org/10.32834/gg.v16i2.198

Purwati, N., & Sulistyo, G. B. (2023). Stunting

Early Warning Application Using KNN Machine Learning Method. *Jurnal Riset Informatika*, *5*(3), 373–378. https://doi.org/10.34288/jri.v5i3.550

Rahayu, A., Yulidasari, F., Putri, A. O., & Anggraini, L. (2018). Stunting dan Upaya Pencegahannya. In *Buku stunting dan upaya pencegahannya*.

Ramadhan, K. (2019). Status Gizi menurut Tinggi Badan per Umur pada Balita. *Poltekita : Jurnal Ilmu Kesehatan*, *13*(2), 96–101. https://doi.org/10.33860/jik.v13i2.38

Romadloni, P., Adhi Kusuma, B., & Maulana Baihaqi, W. (2022). Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 622–628. https://doi.org/10.36040/jati.v6i2.5238

Ruaida, N., & Soumokil, O. (2018). HUBUNGAN STATUS KEK IBU HAMIL DAN BBLR DENGAN KEJADIAN STUNTING PADA BALITA DI PUSKESMAS TAWIRI KOTA AMBON. *Jurnal Kesehatan Terpadu (Integrated Health Journal)*, *9*(2), 45–51.

Saeful Bachri, O., & Herdian Bhakti, R. M. (2021). Penentuan Status Stunting pada Anak dengan Menggunakan Algoritma KNN. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, *3*(02 SE-Articles), 130–137. https://doi.org/10.46772/intech.v3i02.533

Salsabila, Martha, S., & Andani, W. (2024). Komparasi Algoritma K-Nearest Neighbor Dengan Euclidean Distance Dan Manhattan Distance Untuk Klasifikasi Stunting Balita. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, *13*(2), 285–292.

Sumarlin, S. (2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. *Jurnal Sistem Informasi Bisnis*, *5*(1), 52–62. https://doi.org/10.21456/vol5iss1pp52-62

Wati, E. F., & Sudrajat, B. (2022). Application of Naive Bayes Method For Diagnosis of Pregnancy Disease. *International Journal of Information System & Technology*, *6*(1), 93–100.