# Classifying Stunting Status In Toddlers Using K-Nearest Neighbor And Logistic Regression Analysis

1st Alvin Tolopan Armando Sibuea
Faculty of Informatics
Telkom University
Bandung, Indonesia
Email: alvintsibuea26@gmail.com

2nd Putu Harry Gunawan
*CoE HUMIC, School of Computing*
*Telkom University*
Bandung, Indonesia
phgunawan@telkomuniversity.ac.id

3rd Indwiarti
*CoE HUMIC, School of Computing*
*Telkom University*
Bandung, Indonesia
indwiarti@telkomuniversity.ac.id

***Abstract*—*Stunting* is a critical child growth disorder, characterized by a height below the norm for one's age group. Despite a notable decrease in stunting prevalence in Indonesia from 37% in 2014 to 21.6% in 2022, achieving the targeted reduction to 14% by 2024 remains imperative. This study contributes to this national health goal by developing a robust predictive model for stunting in toddlers using machine learning. The research employs two models, K-Nearest Neighbors (KNN) and Logistic Regression. The dataset used for this research shows a big gap of data imbalance, for the majority of its data is significantly higher than the minority. Both of the chosen method is focusing on mitigating data imbalance through oversampling and undersampling techniques. The KNN model is particularly suited for this study due to its effectiveness in handling the complex, non-linear patterns often found in multifaceted health data like stunting indicators. It consistently demonstrated high accuracy, averaging 0.980, and reaching 0.987 for F1-Score. Logistic Regression, chosen for its ability to provide clear interpretability, especially useful in understanding the impact of various health indicators, also performed well with an average accuracy of 0.877 and an F1-Score of 0.894. This study highlights the significance of machine learning in addressing child stunting, providing effective tools for prediction. The combination of KNN's ability to handle complex data and Logistic Regression's interpretability, along with data balancing, contributes to the goal of reducing stunting prevalence. In summary, this research tackles child stunting, a pressing issue in Indonesia. By leveraging machine learning techniques, it develops predictive models to aid in stunting prevention. KNN excels in capturing complex patterns, while Logistic Regression offers interpretability. These models offer promise in reaching the vital goal of reducing stunting to 14% by 2024.**

***Index Terms*—machine learning, stunting, k-nearest neighbor, logistic regression**

## I. Introduction

### A. Background

Stunting is still considered to be one of the most well-known and significant child health disorder in Indonesia. Stunting is a condition where a child's height appears to be shorter than their peers [1]. Stunting affects child's physical and cognitive development as well as impacting future productivity of the child [2], [3]. According to Indonesian Health Ministry data, there are approximately one-third of nine million children in Indonesia experienced stunting in 2018, with a prevalence of 30.8%, which is a slight increase from the previous year's prevalence of 29.6% [4], [5]. In 2020, Indonesia still ranked 115th out of 151 on stunting prevalence category. Despite being on the lower ranks, Indonesia is still classified as a country with the most stunting prevalence according to JME, UNICEF, and World Bank [6]. While this increase might not seem significant, the government still keeps a close monitor on the issue at hand.

To handle the rising issue of stunting in toddlers, there needs to be an accurate and rapid measurement method [7]. Currently, the most common method is anthropometric measurement, which measures height, weight, and calculating child's growth index (IPT). However, this method of measurement has drawbacks which requires extra time, efforts, and also specialized knowledge for calculating and interpreting the results. In order to achieve higher accuracy, a machine learning-based system is needed to predict the likelihood of stunting status in children [8]. This system will be able to provide an extra hand to health workers, from efficiency to the handling process of stunting cases.

Numerous studies in Indonesia, such as conducted by Otong Saeful Bachri and Raden M. H. Bhakti in 2021. They implemented the K-Nearest Neighbor (KNN) algorithm for stunting classification in Brebes. The result revealed that KNN achieved the highest accuracy at 83% and the lowest error rate at 0.142 [9].

Building machine-learning systems requires training with data that provide an integrated view and information about the subject. A model can learn what happens to patients only if the outcomes are included in the data set that the model is based on [10]. Therefore, machine learning implementation can help healthcare system to overcome stunting issue by providing the ability to efficiently and accurately identify stunting status in children especially toddlers. This research proposal aims to determine stunting status in children under five using K-Nearest Neighbor (KNN) and Logistic Regression Machine learning techniques. The selection of these two algorithms aims to provide classification variations for the dataset. With this predictive model, the research hopes to offer a more effective alternative in determining stunting status in children.

### B. Related Works

*1) Determining Children's Stunting Status With KNN Algorithm:* O. Bachri and R. Bhakti. (2021) [9] observed a similar cases of stunting on children in Brebes, Central Java where it was found that around 3.77% of toddlers are still experiencing malnutrition and stunting, and 13.20% of toddlers are still malnourished and tend to be stunted. The study findings gave a resourceful insight in determining stunting status with KNN algorithm. By tuning the model parameters as method, setting

979-8-3503-6535-1/24/$31.00 ©2024 IEEE

three as the K-value parameter appeared to be the best option among all of the parameters as it performed best with the datase. Achieving the smallest error rate of 0.142 and 83% in accuracy.

*2) Machine Learning in Stunting Prediction – A Case Study from Rwanda:* S. Ndagijimana et al. (2022) [11] conducted an insightful study on stunting among under-5 children in Rwanda, where stunting prevalence was 33% in 2020. By employing various machine learning algorithms, they developed a more accurate model for predicting stunting. Among the methods tested, the Gradient Boosting Classifier emerged as the most effective, underlining the value of sophisticated algorithms in tackling complex health issues like stunting.

*3) Machine Learning for Undernutrition Prediction in Ethiopia:* In another significant study, F. H. Bitew et al. (2021) [12] focused on undernutrition among children under five in Ethiopia. They reviewed previous research findings on stunting causes, including wasting and underweight factors, and demonstrated how machine learning surpasses traditional statistical models in classifying nutrition-related problems. This study highlights the growing relevance of advanced data analytics in understanding and addressing undernutrition and its consequences.

The insights from these studies underscore the transformative potential of machine learning and data-driven approaches in addressing stunting. By leveraging advanced algorithms and technology, we can achieve a deeper understanding and more accurate predictions of stunting, which are crucial for effective interventions and policy-making. Therefore this research is going to be focusing on testing and creating an accurate prediction model to help health workers determine stunting status for toddlers especially in Bojongsoang, Bandung Region as a follow up to recent studies that has been conducted. By experimenting with two models, this study aims to help on increasing the variety of machine learning approaches used in the healthcare system, to respectively enriches and continue the findings from previous studies, and to achieve the best model possible to help predict and prevent more risks to occur.

## II. Methodology

### A. Data Collection

The dataset used for this research is obtained from Children's Growth Data gathered in August 2022 by The Community Health Center (Puskesmas) in Bojongsoang. It contains 6.677 data regarding children's growth and some other variety of personal identifiers of the children. Some samples of the dataset can be seen as follows:

TABLE I
Child's Growth Table

| Name | Gender | Age | Weight | Height | Status |
|---|---|---|---|---|---|
| Faisal | M | 2,4 | 4,9 | 58 | Not Stunting |
| Hafidz | M | 58,4 | 17 | 109 | Not Stunting |
| Aulia | F | 19,3 | 5,7 | 72 | Stunting |

### B. Exploratory Data Analysis (EDA)

After categorizing the variables, a detailed statistical analysis was conducted. For numerical variables, including 'Age', 'Weight', and 'Height', examined from their distribution, central tendencies, and variability. The summary is visualized as such.

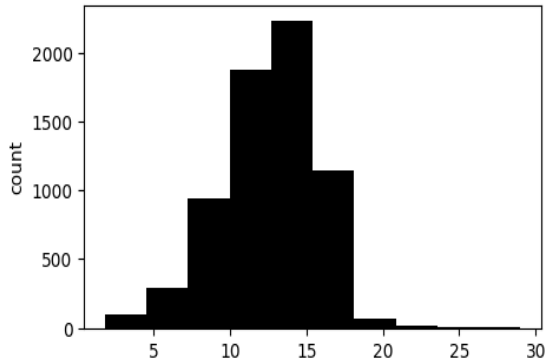


Fig. 1. Toddlers' Weight Chart

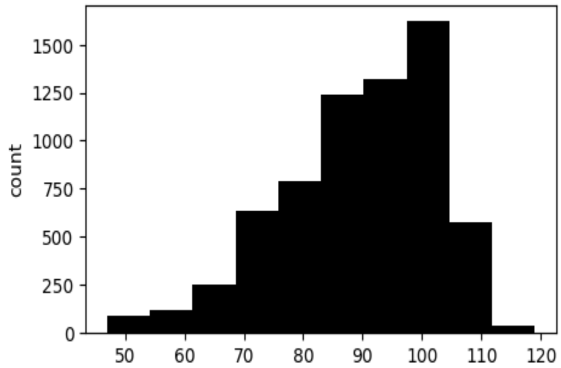


Fig. 2. Toddlers' Height Chart

The findings based on Figure 3 indicate that there are some likeliness on some of the kids' ages, ranging from less than one month old to 60 months of age. Figure 3 also shows that the majority of toddlers are getting measured by the age of 30-months-old to 50-months indicating that the majority of toddlers are getting the attention that they deserve by getting measured in time to take precautions. While Figure 1 and Figure 2 indicates that children, in the region where the measurement taken place, are growing well. Though these two categories didn't indicate some significant gap.

The categorical variables 'Gender' and 'Status' were analyzed to understand their distribution within the dataset. As seen on Figure 4, this analysis revealed significant data imbalance on the 'Status' category. The majority of the data

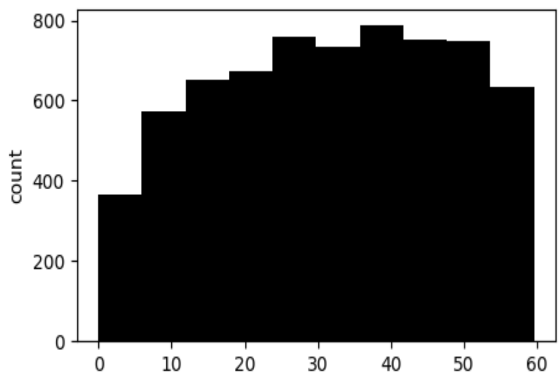


Fig. 3. Toddlers' Age Chart

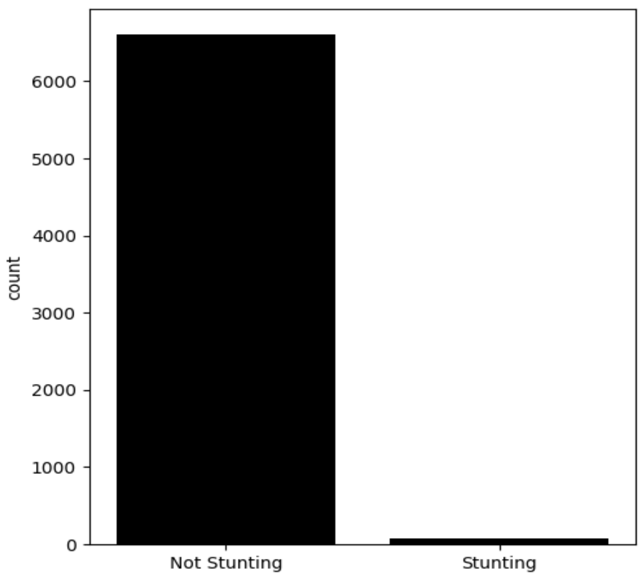


Fig. 4. Stunting Status Bar Chart

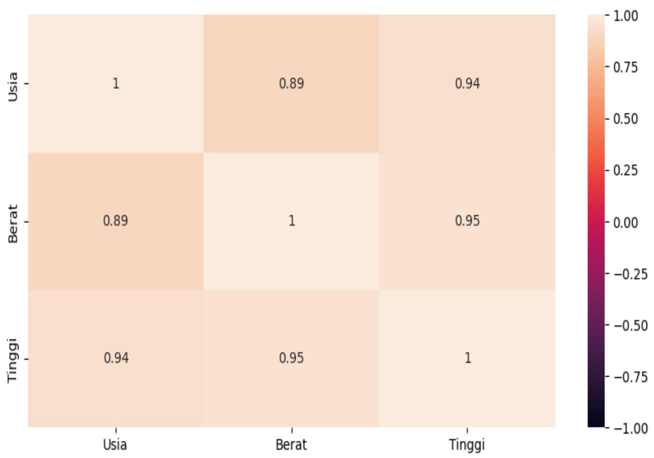


Fig. 5. Heatmap Correlation Between Age, Height, and Weight

appears to be not stunting while there's only 10% of the data which appears stunting. It would cause bias in the analysis part.

Histograms and box plots provided a clear picture of the distribution and variance in numerical data. These visuals highlighted the range of values and potential outliers. For categorical data, bar charts were employed to showcase the frequency of different categories within 'Gender' and 'Status', offering insights such as the most common category or imbalances between categories.

A correlation analysis was also performed to identify relationships between the numerical variables. The correlations is shown in Figure 5. This analysis was crucial in uncovering like positive or negative correlations. And based on the analysis, it's shown that there is no negative correlations and the majority of the data shows positive correlations between parameters and mostly close to one.

The preliminary findings from the EDA have provided valuable insights into the characteristics and relationships within the dataset. These insights will be instrumental in guiding further analysis, particularly in handling the data imbalance.

### *C. Data Preprocessing*

The preprocessing phase of this research was a critical step in preparing the dataset for accurate and effective modeling. This phase involved several key procedures, each designed to optimize the data for the machine learning algorithms employed later in the study.

Initially, this phase focused on feature transformation and encoding. To achieve this, I used the ColumnTransformer from scikit-learn, which allowed the application of different transformations to specific columns of the dataset. Categorical variables were particularly addressed using OneHotEncoder, ensuring that these variables were suitably encoded for analysis. This transformation process was essential for converting categorical data into a format that could be effectively utilized by our machine learning models. Following the transformation, the feature matrix, denoted as x, was converted into a NumPy array, facilitating seamless integration into subsequent analytical steps.

In addition to feature transformation, the preprocessing included a critical step of label encoding. Utilizing scikit-learn's LabelEncoder to transform the target variable y from categorical to a numerical format. This conversion was crucial for aligning the dataset with the requirements of the machine learning algorithms, particularly because many algorithms are designed to work optimally with numerical inputs.

Lastly, the preprocessing phase involved the standardization of the numerical features in the dataset. Employing the StandardScaler from scikit-learn to standardized the features, particularly focusing on columns from the third column onwards in both the training (x_train) and testing (x_test) subsets. Standardization is a vital process that normalizes the range of independent variables or features of data, which is particularly important when using algorithms that are sensitive to the scale of input data, like K-Nearest Neighbors (KNN) and Logistic Regression.

Each of these preprocessing steps played a vital role in ensuring that the dataset was not only in an optimal format for analysis but also aligned with the methodological requirements of this research. This thorough preparation of the data set a strong foundation for the subsequent stages of our research.

### *D. Handling Imbalance*

In the course of my study, I encountered the prevalent issue of data imbalance, a common challenge in machine learning where there's a significant disparity in the distribution of classes within the dataset. Such imbalance often leads to models that are biased towards the majority class, undermining their effectiveness in predicting instances of the minority class. To minimize this problem, I implemented two key strategies: undersampling and oversampling.

Undersampling was the first technique I implemented. This approach involves reducing the size of the majority class to align it with the minority class. The aim is to create a balanced dataset by selectively removing instances from the overrepresented class. In applying undersampling, I carefully ensured that the reduced dataset still retained the essential characteristics of the majority class. This method proved particularly beneficial, as the dataset was large enough that reducing the majority class did not significantly compromise the information content necessary for model training.

Conversely, oversampling focuses on augmenting the minority class by adding more instances. I applied oversampling to enhance the representation of the minority class. This was crucial in ensuring that the predictive model developed a better understanding of the minority class's characteristics, thus improving its ability to accurately predict underrepresented categories.

Both undersampling and oversampling have their advantages and limitations. While undersampling can lead to the loss of potentially valuable data, oversampling, particularly through synthetic data generation, may introduce the risk of overfitting. Therefore, my choice between undersampling and oversampling was made after careful consideration of the dataset's size and the nature of the imbalance.

By strategically applying these techniques, my aim was to create a more balanced dataset, enhancing the robustness and reliability of the machine learning models in predicting stunting in children. This approach was crucial in ensuring that the models were not biased towards the majority class and were sensitive to the nuances of the minority class, which is essential in a study with significant public health implications.

### E. Training Models

The main objective of this research is being able to train machine learning models to classify stunting status in children using available dataset. Machine learning requires a huge amount of data for classifying or predicting diseases. Supervised machine learning algorithms need annotated data for classifying the text or image into different categories [13]. There are two distinct machine learning models applied for this research, K-Nearest Neighbors (KNN) and Logistic Regression, chosen for their effectiveness in classification tasks and their applicability to the dataset under study.

The K-Nearest Neighbors (KNN) algorithm is a supervised algorithm which will produce new instance queries by classifying the majority of the category. Finding new patterns from existing data by connecting the patterns to the new data [14]. This algorithm is employed for its simplicity and proven efficacy in classification tasks. KNN classifies data based on the feature similarity, where the 'K' in KNN signifies the number of nearest neighbors considered to determine the class of a data point [14], [15]. A critical aspect of using KNN was the selection of an optimal 'K' value [16], which I determined through experimentation to enhance the model's performance.

On the contrary, Logistic Regression, a statistical method for predicting binary outcomes, was selected for its suitability in binary classification problems [17]. This model was particularly relevant for determining stunting in children, as it operates by estimating the probabilities of binary outcomes using a logistic function [14], [18]. Such an approach not only facilitates binary classification but also provides insights into the model's confidence in its predictions through class probabilities. This is particularly useful not only to understand the independent relationship of each variable with the outcome, but also, to adjust the estimates for the effects of confounding variables in observational research [19], [20]. The tuning of model parameters in Logistic Regression was undertaken with a focus on balancing the bias-variance tradeoff, which is crucial for the model's generalization capability on new data.

Both models were applied to the same dataset, which had been subjected to thorough preprocessing and balancing. The deployment of KNN and Logistic Regression allowed for a comprehensive comparative analysis, exploring the potency of a distance-based algorithm versus a probabilistic model in tackling the classification challenge posed by the study.

## III. Results And Discussion

The analysis of the experimental results from the KNN and Logistic Regression models, applied to the dataset with both oversampling and undersampling techniques, offers insightful implications for this research in predicting stunting in toddlers. By conducting seperate learnings and applying two different imbalance handlings for each models.

TABLE II
KNN Model Accuracy Results

| Oversample | Undersample |
|---|---|
| 0.980 | 0.695 |
| 0.973 | 0.692 |
| 0.969 | 0.730 |
| 0.990 | 0.769 |
| 0.989 | 0.807 |
| 0.993 | 0.807 |
| 0.995 | 0.769 |
| 0.992 | 0.807 |
| 0.994 | 0.884 |
| 0.993 | 0.653 |

TABLE III
Logistic Regression Model Accuracy Results

| Oversample | Undersample |
|---|---|
| 0.882 | 0.801 |
| 0.900 | 0.846 |
| 0.883 | 0.961 |
| 0.904 | 0.769 |
| 0.981 | 0.730 |
| 0.904 | 0.846 |
| 0.896 | 0.769 |
| 0.909 | 0.730 |
| 0.897 | 0.846 |
| 0.914 | 0.576 |

Based on the results shown in Table II and Table III gathered by conducting several trials on both models, the observable results are evaluated and summarized into some main points. As shown in Table IV the experimental results are gathered and averaged for analysis.

### A. Effectiveness of Oversampling in Handling Imbalance

By using two different handling methods for imbalanced dataset, undersampling and oversampling, oversampling has proven to be more effective to increase the model's ability to learn. The results for the very few first attempt using undersampling for both models are not satisfying. Although for the first try, KNN reached 0.92 in accuracy but still showed significant difference in result than the previous research done by O. Bachri and R. Bhakti [9] which achieved 0.83 in accuracy with similar method and dataset. Logistic Regression reached 0.86 but by repeating the same experiment, the results for both models took a significant downfall to below 0.70 for each models. Since the gap between experiments' results

TABLE IV
AVERAGE OBSERVATION RESULTS

| Resampling Method | Category | LR | KNN |
|---|---|---|---|
| Undersampling | Accuracy | 0.769 | 0.780 |
| | F1-Score (0) | 0.789 | 0.709 |
| | F1-Score (1) | 0.800 | 0.742 |
| Oversampling | Accuracy | 0.877 | 0.981 |
| | F1-Score (0) | 0.890 | 0.987 |
| | F1-Score (1) | 0.894 | 0.987 |

seemed wide, switching handling methods turned out to affect both models greatly.

The consistently high accuracy achieved by the KNN model, shown in Table II, with oversampling (mostly above 0.980) shows the effectiveness of this technique in addressing the class imbalance issue prevalent in stunting data. Oversampling, by augmenting the minority class, seems to have provided a more representative and diverse training dataset, enabling the model to learn more effectively. This is particularly relevant in the context of stunting prediction, where the minority class (stunted children) holds significant importance. Similarly, the Logistic Regression model, shown in Table III, also exhibited performance improvements with oversampling, reinforcing the idea that a balanced dataset through oversampling enhances model accuracy and reliability in public health datasets. Although it was not much different than S. Ndagijimana's study [11] of using Gradient Boosting Classifier method to increase the performance of each model, this study encourage the idea of having balanced dataset could also significantly increase the models' performances.

### B. Variability in Performance with Undersampling

In Table II, the varied performance of both models with undersampling, especially the wide range in accuracy (as low as 0.653 to as high as 0.884 for KNN), indicates a potential loss of critical information. In the context of stunting prediction, this suggests that undersampling might lead to the exclusion of vital patterns or characteristics inherent in the larger class. Such loss of information can be fatal, as it may lead to overlooking key predictors of stunting.

### C. Comparative Analysis of Models

According to the results shown in Table IV, while both models performed better with oversampling, the KNN model displayed the highest consistency and accuracy. This suggests that for datasets similar to the one used in this study, KNN might be more adept at capturing the complexities involved in stunting prediction when provided with a well-balanced dataset. It is also seen by observing the average F1 score obtained by each models with oversampling which suggests a good overall performances of the model in terms of both precision and recall. The Logistic Regression model, though slightly less accurate than KNN in this study, still showed strong performance with oversampling. This indicates its potential applicability in stunting prediction, especially in scenarios where a probabilistic output is desired for understanding the likelihood of stunting.

### D. Implications for Stunting Prediction

The findings from these experiments are crucial in the context of public health, where accurately predicting stunting in children can lead to timely and effective interventions. The superior performance of the KNN model with oversampling suggests its suitability for deployment in real-world scenarios. However, the Logistic Regression model's strong performance also indicates that it can be a viable alternative, particularly where the ability to interpret results is a key factor.

## IV. CONCLUSION

The findings of this study using K-Nearest Neighbors (KNN) and Logistic Regression models for predicting stunting in toddlers have provided significant insights. The KNN model, when applied with oversampling techniques, demonstrated exceptional accuracy, averaging 0.981, and an F1-Score of 0.987. This indicates its robustness in correctly identifying instances of stunting. The Logistic Regression model also exhibited commendable performance under the same conditions, achieving an average accuracy of 0.877 and an F1-Score of 0.894. These results underscore the effectiveness of these models in a balanced dataset, highlighting the crucial role of data sampling methods in predictive accuracy. However, the application of undersampling revealed variability in performance, with KNN accuracy ranging from 0.653 to 0.884 and Logistic Regression from 0.576 to 0.961. This variation emphasizes the impact of dataset composition on model outcomes and the importance of choosing appropriate data handling techniques, especially in public health research.

The effectiveness of oversampling in improving model performance emerged as a key takeaway, highlighting the need for careful consideration of data preprocessing techniques in machine learning applications. This approach not only enhanced the predictive accuracy but also ensured a more equitable representation of the minority class, which is vital in public health contexts.

The implications of this research extend beyond the technical domain, offering practical tools for healthcare professionals and policymakers. The development of an accurate, reliable, and efficient predictive model for stunting is a step forward in enabling early detection and intervention in toddlers, thereby potentially reducing the prevalence and long-term impacts of stunting.

In conclusion, this research not only contributes to the existing body of knowledge in machine learning applications in healthcare but also presents a promising avenue for tackling the critical issue of stunting in children. By harnessing the power of data and advanced analytics, it is hoped that strides can be made towards mitigating this public health challenge and improving the lives of children worldwide. Overall, the study contributes valuable insights into the application of machine learning in the field of child health, demonstrating the potential of these models in aiding the early detection and intervention of stunting, thereby supporting public health efforts and research.

## REFERENCES

[1] D. Sudigyo, A. A. Hidayat, R. Nirwantono, R. Rahutomo, J. P. Trinugroho, and B. Pardamean, "Literature study of stunting supplementation in indonesian utilizing text mining approach," *Procedia Computer Science*, vol. 216, pp. 722–729, 2023, 7th International Conference on Computer Science and Computational Intelligence 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050922022670

[2] W. Wahyudinna, F. M. Hanab, and A. Prihandono, "Prediksi stunting pada balita di rumah sakit kota semarang menggunakan naive bayes," *Jurnal Ilmu Komputer dan Matematika*, pp. 32–36, 2023.

[3] R. Rahutomo, G. N. Elwirehardja, N. Dominic, B. Caesario, and B. Pardamean, “Database management system design improvement for child stunting data collection in multiple observation areas,” in *2022 International Conference on Information Management and Technology (ICIMTech)*, Semarang, Indonesia, 2022, pp. 149–154.

[4] World Health Organization (WHO), “The unicef/who/wb joint child malnutrition estimates (jme) group released new data for 2021,” https://www.who.int/, May 2021.

[5] A. D. N. Yadika *et al.*, “Pengaruh stunting terhadap perkembangan kognitif dan prestasi belajar,” *Journal Name Here*, vol. 8, no. 2, pp. 273–282, 2019.

[6] A. Z. Widniah, T. Hidayat, and A. Febriana, “Optimalisasi pencegahan dan penanggulangan stunting di desa sungai tuan ilir,” *INDRA: Jurnal Pengabdian Kepada Masyarakat*, vol. 3, no. 1, pp. 13–18, 2022.

[7] S. Lonang and D. Normawati, “Klasifikasi status stunting pada balita menggunakan k-nearest neighbor dengan feature selection backward elimination,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 01, pp. 49–6, Jan. 2022.

[8] A. Yusuf, “K-means clustering based on distance measures: Stunting prevalence clustering in south kalimantan,” *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2023.

[9] O. S. Bachri and R. M. H. Bhakti, “Penentuan status stunting pada anak dengan menggunakan algoritma knn,” *j.ilm.intech*, vol. 3, no. 02, pp. 130–137, Nov. 2021.

[10] A. Rajkomar, J. Dean, and I. Kohane, “Machine learning in medicine,” *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019, pMID: 30943338. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMra1814259

[11] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, “Prediction of stunting among under-5 children in rwanda using machine learning techniques,” *Preventive Medicine and Public Health*, vol. 56, pp. 41–49, Dec. 2022.

[12] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, “Machine learning algorithms for predicting undernutrition among under-five children in ethiopia,” *Public Health Nutrition*, Oct. 2021.

[13] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan *et al.*, “Machine learning based approaches for detecting COVID-19 using clinical text data,” *International Journal of Information Technology*, vol. 12, pp. 731–739, 2020. [Online]. Available: https://doi.org/10.1007/s41870-020-00495-9

[14] M. R. Romadhon and F. Kurniawan, “A comparison of naive bayes methods, logistic regression and knn for predicting healing of covid-19 patients in indonesia,” in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, Surabaya, Indonesia, 2021, pp. 41–44.

[15] L. Farokhah, “Implementasi k-nearest neighbor untuk klasifikasi bunga dengan ekstraksi fitur warna rgb,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 6, pp. 1129–1136, 2020.

[16] D. Zou, Y. Xiang, T. Zhou, Q. Peng *et al.*, “Outlier detection and data filling based on knn and lof for power transformer operation data classification,” *Energy Reports*, vol. 9, pp. 698–711, 2023.

[17] M. Ohyver, J. V. Moniaga, K. R. Yunidwi, and M. I. Setiawan, “Logistic regression and growth charts to determine children nutritional and stunting status: A review,” Oct. 2017.

[18] W. A. Shantika Martha and S. W. Rizki, “Perbandingan metode k-nearest neighbor, regresi logistik biner, dan pohon klasifikasi pada analisis kelayakan pemberian kredit,” *EULER: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 10, no. 02, pp. 262–273, Dec. 2022.

[19] P. Schober and T. R. Vetter, “Logistic regression in medical research,” *Anesthesia & Analgesia*, vol. 132, no. 2, pp. 365–366, Feb 2021.

[20] T. N. Joshi and P. M. Chawan, “Logistic regression and svm based diabetes prediction system,” *International Journal For Technological Research In Engineering*, vol. 5, pp. 4347–4350, 2018.