International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 2
ISSN: 2088-5334

# Modeling the Cart Method on the Factors that Cause Stunting

Mimmy Sari Syah Putri [a,*], Aghni Syahmarani [a], Putri Khairiah Nasution [a], Nenna Irsa Syahputri [b], Alberto Tondang [c]

[a] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatra Utara, Indonesia*
[b] *Department of Informatics Engineering, Faculty of Engineering and Computing, Universitas Harapan Medan, Indonesia*
[c] *Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatra Utara Medan, Indonesia*
*Corresponding author: *mimmysarisyah@usu.ac.id*

***Abstract*— Modeling the Classification and Regression Tree (CART) method to identify and address the factors causing stunting can significantly aid in reducing stunting rates. Stunting, a chronic malnutrition problem, is influenced by various factors, including poor nutrition, repeated infections, inadequate psychosocial stimulation, and low socioeconomic status. Key determinants such as maternal education, exclusive breastfeeding, low birth weight, and environmental sanitation play crucial roles in stunting prevalence. Effective interventions include improving antenatal care, contraceptive use, and handwashing practices, which have been shown to reduce stunting rates by significant margins. Community empowerment through education and socialization, mainly targeting mothers and young women, is essential for stunting prevention and management. The CART method can effectively model these multifaceted factors by identifying the most significant predictors and their interactions, guiding targeted interventions to reduce stunting rates. By leveraging comprehensive data and community-based strategies, the CART method can provide a robust framework for reducing stunting and improving child health outcomes. The findings of this study are based on data variables, including gender, exclusive breastfeeding, height, and age of toddlers, which are determinants of the accuracy of stunting cases. Analysis of the CART method shows results that can be used in decision-making related to stunting data. In 200 cases, 75.5% accuracy was obtained with this method. The results of this study provide an overview so that future research can use other, more accurate methods and add more variables to the cause of stunting.**

***Keywords*— Stunting; CART; regression tree; hierarchy; malnutrition.**

*Manuscript received 12 Sep. 2024; revised 27 Dec. 2024; accepted 26 Mar. 2025. Date of publication 30 Apr. 2025.*


## I. Introduction

Stunting is a persistent nutritional issue defined by a child's height being below the normative standard for their age, adversely affecting long-term growth and cognitive development. Numerous studies indicate that stunting results from the interplay of multiple factors, including maternal nutritional status, parenting practices, environmental health, and family socio-economic conditions[1]. The frequency of stunting in Indonesia remains significantly high and is a primary concern in the national health program, given its impact on the future quality of human resources[2]. This study seeks to examine the determinants of stunting through a quantitative methodology and to develop a robust predictive model utilizing the CART (Classification and Regression Tree) technique. This method was selected due to its capacity to manage diverse kinds of variables, including categorical and numeric, and its facilitation of straightforward interpretation using a decision tree framework. Consequently, the elements influencing children's nutritional status can be determined more thoroughly. The data was gathered from several sources, including anthropometric measures at Posyandu and affiliated Health Offices, maternity and child health records at Puskesmas, and socio-economic and environmental surveys. The project aims to enhance comprehension of the primary characteristics contributing to stunting and forecast the risk of stunting at individual and group levels through CART analysis. The results acquired will serve as the foundation for developing more effective and targeted policy and intervention measures to reduce stunting. The spread of stunting in Medan City is increasingly worrying [3].

### *A. Data Mining Process*

Data mining is a component of KDD (Knowledge Discovery in Database), which involves converting unprocessed data into usable and comprehensible information or novel insights [4]. Data mining is systematically extracting valuable information from vast quantities of data. Data mining approaches typically serve two primary objectives: prediction and description. Data mining is employed to describe an object by initially identifying similarities among interconnected objects, to uncover patterns such as clusters, anomalies, and correlations[5]. Clustering, often known as cluster analysis, is a data mining algorithm commonly employed to describe an item [6]. Clustering can be categorized into two main types: hierarchical and non-hierarchical. The K-Means technique is widely recognized as the most renowned clustering method. Data mining in prediction aims to construct a model that can accurately forecast the unknown or future value of a specific attribute of interest. The attribute to be predicted is generally referred to as the dependent variable, class, or target [7]. The factors utilized for making predictions are called explanatory or independent variables. Two categories of algorithms are commonly employed in constructing models: classification and regression[8].

Classification is employed to forecast target variables with discrete data types, whilst regression is utilized to forecast target variables with continuous data types. The objective of classification and regression is to construct a model that minimizes the discrepancy between the predicted value and the actual value of the target variable [9]. The data mining process is a systematic approach to discovering valuable patterns, relationships, and insights from large datasets [10]. It involves various steps and techniques to extract useful information and knowledge from raw data. The data mining process is iterative, and it may include going back to previous steps to refine the method based on new findings or changing goals [11].

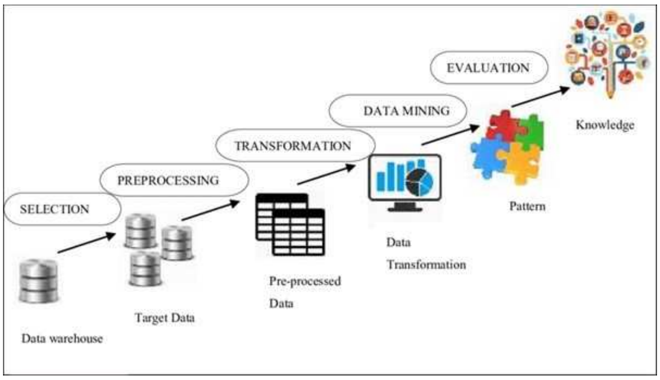


Fig. 1 Steps of data mining process

## B. Decision Tree

Decision trees significantly enhance the accuracy of diagnosing and treating stunting by providing a structured approach to classify and predict cases based on various factors. Their ability to handle complex datasets and identify key variables contributes to more effective public health strategies[12]. Classification Accuracy Decision trees, notably the C4.5 algorithm, have shown promising results in classifying stunting cases[13]. In a comparative study, a decision tree achieved an impressive accuracy of 99% for detecting stunting in toddlers, outperforming other algorithms like Naïve Bayes and SVM[14]. Integration with Other Techniques The stacking ensemble method, which includes decision trees, demonstrated a high accuracy when combined with XGBoost and Gradient Boosting[15]. Particle Swarm optimization applied to decision trees improves accuracy and demonstrates the potential of hybrid approaches[16]. Addressing Data Imbalance Decision trees can effectively manage data imbalance issues, which is crucial in stunting classification, as evidenced by the Random Forest algorithm achieving 96.33% accuracy after optimization[17]. While decision trees provide a robust framework for stunting diagnosis, their effectiveness can be further enhanced through integration with other machine learning techniques and optimization methods, indicating a promising direction for future research and application[18].

## C. Classification and Regression Trees

While not explicitly detailed in the provided papers, the CART (Classification and Regression Trees) algorithm can be compared to other machine learning models used for predicting stunting cases based on their performance metrics. The studies indicate that Random Forest (RF) consistently outperforms other algorithms' accuracy and reliability across various contexts. Performance Comparison:

1) *Random Forest*: Achieved the highest accuracy rates, with 99.95% in one study[19] and 79% in another[13]. It is noted for its robustness in handling complex datasets and feature selection.

2) *Support Vector Machine* (*SVM*): Demonstrated high accuracy (98%) in stunting detection[20], but generally less effective than RF in other studies.

3) *Extreme Gradient Boosting* (XGBoost): Identified as effective in PPH prediction models with a focus on feature importance and flexibility. It showed lower misclassification rates compared to other models[21].

4) *Logistic Regression and Naïve Bayes*: These models performed poorly relative to RF and SVM, with accuracies of 76% and lower, respectively [22].

While CART could be a viable option, Random Forest and SVM are currently preferred due to their superior performance metrics [23]. However, the choice of algorithm may depend on specific dataset characteristics and computational resources. In contrast, some studies highlight the potential of simpler models like Logistic Regression for interpretability, despite their lower accuracy. This suggests a trade-off between model complexity and ease of understanding, which may influence practical applications in public health interventions [24].

## D. Key Findings and Comparative Performance of CART

Previous studies have utilized Classification and Regression Trees (CART) modeling to investigate factors affecting stunting, mainly focusing on nutritional status and growth patterns in children [25]. These studies have provided insights into the variables influencing stunting and the effectiveness of CART in analyzing such data. The findings from these studies highlight the importance of maternal education and nutritional behavior, as well as the potential of CART in stunting research. A study conducted in Padang City, West Sumatra, used CART to analyze factors affecting the nutritional status of children aged 6-23 months. The study identified mother's education level, knowledge, and nutritional behavior as significant factors. It found that children with mothers having lower education levels and poor nutritional knowledge were more likely to be underweight [26].

Conversely, in children with mothers having higher education levels, poor nutritional behavior was linked to a higher prevalence of overweight [27]. CART in Stunting Classification. Another study applied CART, among other machine learning models, to classify stunting in children under five. Although CART was one of the models tested, the study found that the Support Vector Machine with an RBF kernel achieved the highest accuracy in classifying stunting, indicating that while CART is useful, other models may outperform it in certain datasets [28]. Comparative Analysis of Machine Learning Models. While CART was not the primary focus, a comparative study of machine learning models for stunting detection included CART as one of the models. The study highlighted the effectiveness of K-Nearest Neighbors (KNN) over CART, suggesting that while CART is a viable option, other models might offer better performance in specific contexts[29]. While CART has been effectively used in analyzing factors related to stunting, it is important to consider that other machine learning models, such as Random Forest and KNN, have shown superior performance in some studies. This suggests that the choice of model may depend on the specific dataset and research objectives[30].

## II. Materials and Method

The data used is a sample of learning data that is still heterogeneous. The sample will be selected based on the selection rules and goodness-of-split criteria and the selection of disaggregates depending on the type of dependent variable [31]. The resulting set of parts from the selection process must be more homogeneous than the previous selection.

### A. Classification Tree Formation

Before selecting the separators using the Gini index, it is better to find the gain information of each node using the formula as follows:

$$GI(t) = \sum_{j=1}^{a} P(j|t) log_2 P(j|t) \quad (1)$$

The method of selecting sorters in the CART method uses the Gini index i(t) which is a measurement of the level of diversity of a class of a particular node in the classification so that it can help in finding the optimal sorting function. The diversity function used is the Gini index, where the selection of this attribute will produce a binary tree. The Gini index measures the diversity between the probabilities of the target attribute values. The Gini Index function is as follows:

$$i(t) = 1 - \sum_{j=1} p^2(j|t) \quad (2)$$

The characteristics that were chosen will make up a group of classes known as nodes. Recursive selection will always happen on nodes until final nodes are found. The next step is to find the goodness-of-split criterion, which is a measure of how well the s-splitter chose t, which is also known as reducing variability [32]. This can be done with the following formula:

$$\phi(s,t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (3)$$

The tree formed by the parsing rule and the goodness of split criterion is huge because the tree termination is based on the number of observations at the terminal node or the degree of homogeneity. A large tree size can lead to overfitting, but if the tree observations are limited to a certain boundary precision, it will lead to underfitting. A feasible tree size can be obtained by pruning the tree based on the minimum cost complexity measure using the following formula:

$$g_m(t) = \frac{R(t) - R(T_k)}{|T_k| - 1} \quad (4)$$

As long as the goodness-of-split criterion doesn't show a big drop in heterogeneity, there is only one observation at each child node, or the minimum number of cases in a final terminal observation is less than or equal to five (n≤5), then node t is a terminal node [33]. The process of making a tree will also stop when it reaches a certain number of levels or the most profound level possible in the tree. And the process of splitting the tree at node t into t_R and t_L applies:

$$R(t) > R(t_R) + R(t_L) \quad (5)$$

### B. Class Label Marking

Class labelling is the process of identifying each node to a particular class. Class labelling is done on terminal, nonterminal, and root nodes. However, label tagging is most needed at terminal nodes because these nodes are essential for predicting an object in a particular class located at this terminal node. Marking the class label on the terminal nodes is done based on the most significant number rule, i.e. if:

$$P(j|t) = max_j \frac{N_j(t)}{N(t)} \quad (6)$$

### C. Determination of Optimal Classification Tree

Because the data structure described is usually complicated, a big tree size will lead to a high complexity value. Because of this, it is important to choose an ideal tree that is small but can give a small enough replacement estimator value. A frequently used estimator is the cross-validation V-fold estimate. The cross-validation V-fold estimate is used on data that is not large enough < 3000 (less than 3000 data). In the cross-validation V-fold estimate, the data in L are randomly split into V parts that are not connected to each other and are about the same size for each class. A T^v tree is formed from the v-th learning sample where v v=1,2,3,…,V. The test sample estimator for $R(T_t^{(v)})$ is as follows:

$$R(T_t^{(v)} = \frac{1}{N_v} \sum^{v} X\left(d^{(v)}\right) \quad (7)$$

### D. Research Design

This research employs a quantitative methodology with a cross-sectional design. Data was gathered from a population of toddlers, encompassing socio-economic variables, maternal health status, parenting style, and environmental conditions. The primary objective is to ascertain the most influential factors that contribute to stunting. Multiple independent variables were acquired by structured interviews, anthropometric assessments, and health documentation. Following the data preprocessing phase, the analysis was conducted utilizing a decision tree method to forecast the stunting status or risk level. The model is produced using a training and validation procedure, focusing on accuracy, sensitivity, and specificity. The final outcome is a map of essential characteristics accompanied by a predictive model that aids in the prevention of stunting. These findings are crucial for the development of effective public health policy.

## III. Results and Discussion

### A. Data Classification

In this work, we have considered public datasets of four categories of causes of stunting, namely, gender, exclusive breastfeeding, height, and age of the baby.

TABLE I
INDEPENDENT VARIABLES

| Variable | Name Variable | Scale | Category |
|---|---|---|---|
| **$X_1$** | Gender | Nominal | 1. Male<br>2. Female |
| **$X_2$** | Exclusive ASI | Nominal | 1. Yes<br>2. No |
| **$X_3$** | Height | Nominal | 1. Normal<br>2. Short<br>3. Very Short |
| **$X_4$** | Toddler Age | Ordinal | 1. 0-12 months<br>2. 13-24 months<br>3. 25-36 months<br>4. 37-48 months<br>5. 49-60 months |

In Table 1, we can see the independent variables consisting of gender, exclusive breast milk, height, and age of toddlers. The dependent variable is the Incidence of Stunting, as shown in Table 2.

TABLE II
DEPENDENT VARIABLES

| Variable | Name Variable | Scale | Category |
|---|---|---|---|
| **Y** | Incidence of Stunting | Nominal | 1. Yes<br>2. No |

### B. Descriptive Data

The data used in this study is a sample of stunting data in Medan City in 2023. The population in this study is toddlers in the Medan City Health Office area, with a random sample of up to 200 toddlers who experience symptoms of the factors that cause stunting. The sampling technique used was purposive. Based on Table 3, it was found that the number of stunting incidents was 48 cases (24%) and the number of non-stunting incidents was 152 cases (76%), where the factors causing stunting studied were exclusive breastfeeding and toddler height, which were classified into standard, short, and very short.

TABLE III
STUNTING INCIDENCE RATE

| Incidence of Stunting | Frequency | Percentage |
|---|---|---|
| Yes | 48 | 24% |
| No | 152 | 76% |
| **Toddler Age** | | |
| 0-12 month | 42 | 21% |
| 13-24 month | 37 | 18,5% |
| 25-36 month | 35 | 17,5% |
| 37-48 month | 44 | 22% |
| 49-60 month | 42 | 21% |
| **Exclusively ASI** | | |
| Yes | 137 | 68,5% |
| No | 63 | 31,5% |
| **Height** | | |
| Normal | 56 | 28% |
| Short | 68 | 34% |
| Very Short | 76 | 38% |

### C. Data Analysis

From the data in Table 4, the left and right branch candidates were determined based on the variables.

TABLE IV
BRANCH CANDIDATE

| Branch Candidate number | Left Branch Candidate | Right Branch Candidate |
|---|---|---|
| 1 | Gender = Female | Gender = Male |
| 2 | Exclusive breastfeeding = No | Exclusive breastfeeding = Yes |
| 3 | Height = {Normal, Short} | Height = Very Short |
| 4 | Toddler age < average age | Toddler age > average age |

Based on Table 4, it can be explained that the left branch candidates are female gender, not exclusively breastfed, of normal and short height, and age below the average data. In comparison, the right branch candidates are male gender, exclusive breastfeeding, very short height, and toddler age above the average data. To determine the quality of the accuracy percentage of the data analysis results, this research uses the Weka v 3.9.6 data mining application, as shown below.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         151               75.5    %
Incorrectly Classified Instances        49               24.5    %
Kappa statistic                          0.248
Mean absolute error                      0.2827
Root mean squared error                  0.4004
Relative absolute error                 77.1494 %
Root relative squared error             93.7148 %
Total Number of Instances              200

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.333    0.112    0.485      0.333    0.395      0.255    0.756     0.447     Yes
                 0.888    0.667    0.808      0.888    0.846      0.255    0.756     0.906     No
Weighted Avg.    0.755    0.534    0.731      0.755    0.738      0.255    0.756     0.796

=== Confusion Matrix ===

   a   b   <-- classified as
  16  32 |   a = Yes
  17 135 |   b = No
```

Fig. 2 Accuracy of data analysis

Based on this figure, it can be concluded that the data used to form this decision tree is 75.5% accurate (Correctly Classified Instances). Furthermore, the results are obtained as in the following figure in the training data set process.

```
Height = Normal: No (56.0)
Height = Short: No (68.0/10.0)
Height = Very_Short
|   Gender = Male
|   |   Age <= 17: Yes (12.0/5.0)
|   |   Age > 17: No (20.0/5.0)
|   Gender = Female
|   |   Exclusive_breastfeeding = Yes: No (22.0/10.0)
|   |   Exclusive_breastfeeding = No: Yes (22.0/6.0)

Number of Leaves  :     6

Size of the tree :      10
```

Fig. 3 Training data set

The data obtained is then classified to find the correlation of each candidate branch to be formed. This process then continues to determine the CART tree which is visualized as shown below:

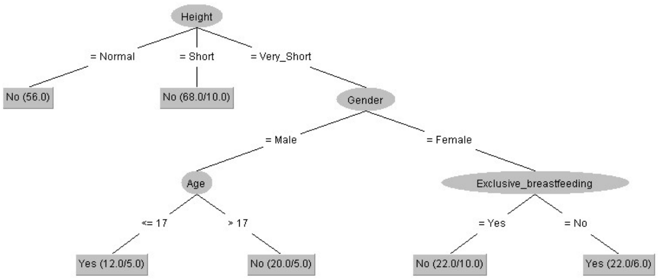


Fig. 4 Decision tree result

It then makes 10 nodes, with one starting node, three inner nodes, and 6 terminal nodes based on the classification tree that was made. The parser with the highest improvement value in the CART Tree is the initial node. In this case, the Height Variable becomes the starting node (node 0), where Normal height at node one and Short height at node 2 do not produce sorters, while Very Short height produces sorters based on Male and Female Gender. Then the disaggregation based on Male gender produces node 5, which can be disaggregated based on Toddler Age, and the disaggregation based on Female gender produces node 6, which can be disaggregated based on Exclusive breastfeeding to toddlers. Toddlers under the age of 17 months who are stunted are depicted in node 7, while toddlers over 17 months who are not stunted are depicted in node 8. Toddlers who are exclusively breastfed and do not experience stunting cases are depicted in node 9, and toddlers who are not exclusively breastfed and experience stunting are depicted in node 10.

## IV. Conclusion

From the decision tree, the main distinguishing factor for the final classification (stunting vs. not stunting) appears to be a child's Height. If the child's height is Normal, the outcome is No, indicating no stunting. Likewise, if the child's height is Short, the classification remains No. However, if the child' s height is Very Short, the model then splits on Gender: Male: Next split is on Age. If Age ≤ 17 months, the final classification is Yes (indicating stunting). If Age > 17 months, the classification is No (not stunting). Female: Next split is on Exclusive_breastfeeding. If exclusive_breastfeeding = Yes, the classification is No (not stunting). If exclusive_breastfeeding = No, the classification is Yes (stunting). Thus, the tree highlights that among Very Short children, being male and younger (≤ 17 months) is more likely associated with stunting, whereas for girls, not receiving exclusive breastfeeding is linked to stunting. Overall, height is the foremost predictor, followed by gender, age, and exclusive breastfeeding in determining stunting status.

## Acknowledgment


TALENTA Universitas Sumatera Utara, Indonesia, funded this research under contract number: 221/UN5.2.3.1/PPM/KP-TALENTA/R/2023.


## References


[1] A. D. Laksono, N. E. W. Sukoco, T. Rachmawati, and R. D. Wulandari, "Factors Related to Stunting Incidence in Toddlers with Working Mothers in Indonesia," *International Journal of Environmental Research and Public Health*, vol. 19, no. 17, 2022, doi: 10.3390/ijerph191710654.

[2] A. Kustanto and O. R. Hasbullah, "The Prevalence of Stunting in Indonesia: An Examination of the Health, Socioeconomic Status, and Environmental Determinants," *Journal of Iranian Medical Council*, 2025, doi: 10.18502/jimc.v8i1.17062.

[3] H. Andriani, E. Friska, M. Arsyi, A. E. Sutrisno, A. Waits, and N. D. Rahmawati, "A multilevel analysis of the triple burden of malnutrition in Indonesia: trends and determinants from repeated cross-sectional surveys.," *BMC public health*, vol. 23, no. 1, p. 1836, Sep. 2023, doi: 10.1186/s12889-023-16728-y.

[4] A. Zia, M. Aziz, I. Popa, S. A. Khan, A. F. Hamedani, and A. R. Asif, "Artificial Intelligence-Based Medical Data Mining.," *Journal of personalized medicine*, vol. 12, no. 9, Aug. 2022, doi:10.3390/jpm12091359.

[5] M. Drogkoula, K. Kokkinos, and N. Samaras, "A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management," *Applied Sciences*, vol. 13, no. 22, 2023, doi:10.3390/app132212147.

[6] M. L. Kolling *et al.*, "Data Mining in Healthcare: Applying Strategic Intelligence Techniques to Depict 25 Years of Research Development.," *International journal of environmental research and public health*, vol. 18, no. 6, Mar. 2021, doi: 10.3390/ijerph18063099.

[7] A. Gondeau, Z. Aouabed, M. Hijri, P. Peres-Neto, and V. Makarenkov, "Object Weighting: A New Clustering Approach to Deal with Outliers and Cluster Overlap in Computational Biology.," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 2, pp. 633–643, 2021, doi:10.1109/TCBB.2019.2921577.

[8] I. Omar, M. Khan, and A. Starr, "Suitability Analysis of Machine Learning Algorithms for Crack Growth Prediction Based on Dynamic Response Data," *Sensors*, vol. 23, no. 3, 2023, doi:10.3390/s23031074.

[9] Y. Matsuzaka and Y. Uesawa, "Ensemble Learning, Deep Learning-Based and Molecular Descriptor-Based Quantitative Structure-Activity Relationships.," *Molecules (Basel, Switzerland)*, vol. 28, no. 5, Mar. 2023, doi: 10.3390/molecules28052410.

[10] W.-T. Wu *et al.*, "Data mining in clinical big data: the frequently used databases, steps, and methodological models.," *Military Medical Research*, vol. 8, no. 1, p. 44, Aug. 2021, doi:10.1186/s40779-021-00338-z.

[11] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: a systematic literature review.," *PeerJ. Computer science*, vol. 6, p. e267, 2020, doi: 10.7717/peerj-cs.267.

[12] J. Minardi, "Enhancing Stunting Detection Accuracy in Children Using SVM with Advanced Data Balancing Techniques," *Journal of Information Systems Engineering and Management*, vol. 10, pp. 501–509, 2025, doi: 10.52783/jisem.v10i6s.748.

[13] O. N. Chilyabanyama *et al.*, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, 2022, doi:10.3390/children9071082.

[14] H. Shen, H. Zhao, and Y. Jiang, "Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea," *Children*, vol. 10, p. 1638, Sep. 2023, doi:10.3390/children10101638.

[15] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," *International Journal of Distributed Sensor Networks*, vol. 18, p. 155013292211069, Jun. 2022, doi:10.1177/15501329221106935.

[16] F. Shen, X. Yan, and Y. Shang, "A novel hybrid PSO-MIDAS model and its application to the U.S. GDP forecast," *PLOS ONE*, vol. 19, Dec. 2024, doi: 10.1371/journal.pone.0315604.

[17] G. Pisoni, B. Molnár, and Á. Tarcsi, "Data Science for Finance: Best-Suited Methods and Enterprise Architectures," *Applied System Innovation*, vol. 4, no. 3, 2021, doi: 10.3390/asi4030069.

[18] T. Sugihartono, B. Wijaya, Marini, A. Alkayes, and H. Anugrah, "Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN," *Journal of Applied Data Sciences*, vol. 6, pp. 667–682, Jan. 2025, doi: 10.47738/jads.v6i1.494.

[19] P. Manorom, U. Detthamrong, and W. Chansanam, "Comparative Assessment of Fraudulent Financial Transactions using the Machine Learning Algorithms Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbor, and Random Forest," *Engineering, Technology*

*& Applied Science Research*, vol. 14, no. 4, pp. 15676–15680, 2024, doi: 10.48084/etasr.7774.

[20] R. Qasrawi *et al.*, “Machine Learning Approach for Predicting the Impact of Food Insecurity on Nutrient Consumption and Malnutrition in Children Aged 6 Months to 5 Years,” *Children*, vol. 11, no. 7, 2024, doi: 10.3390/children11070810.

[21] S. Holcroft, I. Karangwa, F. Little, J. Behoor, and O. Bazirete, “Predictive Modelling of Postpartum Haemorrhage Using Early Risk Factors: A Comparative Analysis of Statistical and Machine Learning Models.,” *International journal of environmental research and public health*, vol. 21, no. 5, May 2024, doi:10.3390/ijerph21050600.

[22] E. van der Heide, R. Veerkamp, M. Pelt, C. Kamphuis, I. Athanasiadis, and B. Ducro, “Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle,” *Journal of Dairy Science*, vol. 102, Aug. 2019, doi: 10.3168/jds.2019-16295.

[23] D. Kirk, E. Kok, M. Tufano, B. Tekinerdogan, E. J. M. Feskens, and G. Camps, “Machine Learning in Nutrition Research,” *Advances in Nutrition*, vol. 13, no. 6, pp. 2573–2589, 2022, doi:10.1093/advances/nmac103.

[24] L. Farah, J. M. Murris, I. Borget, A. Guilloux, N. M. Martelli, and S. I. M. Katsahian, “Assessment of Performance, Interpretability, and Explainability in Artificial Intelligence–Based Health Technologies: What Healthcare Stakeholders Need to Know,” *Mayo Clinic Proceedings: Digital Health*, vol. 1, no. 2, pp. 120–138, 2023, doi:10.1016/j.mcpdig.2023.02.004.

[25] L. Atamou, D. C. Rahmadiyah, H. Hassan, and A. Setiawan, “Analysis of the Determinants of Stunting among Children Aged below Five Years in Stunting Locus Villages in Indonesia,” *Healthcare*, vol. 11, no. 6, 2023, doi: 10.3390/healthcare11060810.

[26] G. Rezaeizadeh *et al.*, “Maternal education and its influence on child growth and nutritional status during the first two years of life: a systematic review and meta-analysis.,” *EClinicalMedicine*, vol. 71, p. 102574, May 2024, doi: 10.1016/j.eclinm.2024.102574.

[27] S. G. Bawa and L. Haldeman, “Fathers Nutrition Knowledge and Child Feeding Practices Associated with Childhood Overweight and Obesity: A Scoping Review of Literature From 2000 to 2023.,” *Global pediatric health*, vol. 11, p. 2333794X241263199, 2024, doi:10.1177/2333794X241263199.

[28] X. Ding, J. Liu, F. Yang, and J. Cao, “Random Radial Basis Function Kernel-based Support Vector Machine,” *Journal of the Franklin Institute*, vol. 358, Oct. 2021, doi: 10.1016/j.jfranklin.2021.10.005.

[29] S. Buttrey and C. Karo, “Using k-nearest-neighbor classification in the leaves of a tree,” *Computational Statistics & Data Analysis*, vol. 40, pp. 27–37, Jul. 2002, doi: 10.1016/S0167-9473(01)00098-6.

[30] H. Shi *et al.*, “Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with congenital heart disease,” *Clinical Nutrition*, vol. 41, Nov. 2021, doi:10.1016/j.clnu.2021.11.006.

[31] D. Rajput, W.-J. Wang, and C.-C. Chen, “Evaluation of a decided sample size in machine learning applications.,” *BMC bioinformatics*, vol. 24, no. 1, p. 48, Feb. 2023, doi: 10.1186/s12859-023-05156-9.

[32] M. Banerjee, E. Reynolds, H. B. Andersson, and B. K. Nallamothu, “Tree-Based Analysis.,” *Circulation. Cardiovascular quality and outcomes*, vol. 12, no. 5, p. e004879, May 2019, doi:10.1161/circoutcomes.118.004879.

[33] X. Zhao and X. Nie, “Splitting Choice and Computational Complexity Analysis of Decision Trees.,” *Entropy (Basel, Switzerland)*, vol. 23, no. 10, Sep. 2021, doi: 10.3390/e23101241.