p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
13(1) 2025: 45-60

ORDINAL XGBOOST FOR MULTICLASS NUTRITIONAL STATUS CLASSIFICATION WITH IMBALANCED DATA

Luthfia Hanun Yuli Arini [1*], Lutfiah Maharani Siniwi[2]
[1] Department of Mathematics Education, Universitas Negeri Yogyakarta, Indonesia
[2] Department of Mathematics, Universitas Jenderal Soedirman, Indonesia

***e-mail**: *luthfiahanunyuliarini@uny.ac.id*

Article Info:
Received: August 14th , 2025
Accepted: September 5th , 2025
Available Online: September 5th , 2025

Keywords:
Binary decomposition; Imbalanced data; Stunting; XGBoost

Abstract: Stunting is a critical global health challenge that undermines children's physical growth and cognitive development, particularly in developing countries. Accurate classification of toddlers' nutritional status is essential for early intervention but is complicated by two challenges: the imbalanced distribution of classes, where stunted and tall categories are underrepresented, and the ordinal nature of nutritional status labels. This study employs XGBoost combined with a Binary Decomposition approach and Enhanced Instance Weighting to address these issues. Secondary data from 100 respondents in Sumberputih Village, Wajak District, were analyzed using four predictors: economic status, health services, children's diet, and environment. The dataset was divided into 80% training and 20% testing portions, and model performance was assessed with metrics suitable for imbalanced ordinal data. Results showed that the model achieved 75% accuracy, an ordinal MAE of 0.25, a QWK of 0.22, and a Macro-F1 score of 0.39. Variable importance analysis highlighted health services as the primary determinant for stunting detection, while environmental factors were most influential in identifying tall status. These findings suggest that XGBoost with Binary Decomposition Enhanced Instance Weighting is effective for handling imbalanced ordinal data and provides valuable insights for supporting stunting prevention and targeted public health interventions.

1. INTRODUCTION

The nutritional status of toddlers, particularly regarding stunting, is one of the most pressing global health issues, affecting both physical and cognitive development, especially in developing countries [1], [2], [3], [4], [5]. Stunting is defined as a failure to grow experienced by toddlers due to chronic malnutrition, resulting in a height significantly below the WHO standard for their age [3], [6], [7]. According to the latest WHO data, in 2024 an estimated 150.2 million children under five worldwide were experiencing stunting. Global data from 2022 also indicated that approximately 148.1 million children (22.3%) were stunted[3], [8], [9], [10]. The stunting problem is not only a national health concern but also part of the global agenda within the Sustainable Development Goals (SDGs), particularly SDG 2 (Zero Hunger) and SDG 3 (Good Health and Well-being).

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

One of the main challenges in addressing stunting is the classification of toddlers' nutritional status when dealing with imbalanced data. In this context, the stunting class is much smaller in number compared to toddlers with normal or tall nutritional status. Such imbalance creates challenges for classification techniques, as models tend to favor the majority class and overlook the minority class, leading to biased and inaccurate predictions, especially for minority data[11], [12], [13], [14].

Furthermore, toddlers' nutritional status data are ordinal, meaning the classes have an inherent order (stunting < normal < tall). However, most standard classification algorithms treat each category as an independent nominal class without considering their order. This ordinal nature requires a more suitable approach that can leverage the order information for more accurate and relevant classification results [15], [16].

While ordinal logistic regression is a popular method for ordinal data classification, it may not perform as well when handling imbalanced data, as it doesn't provide explicit mechanisms to address the class imbalance [17], [18]. In contrast, XGBoost offers greater flexibility through weighting mechanisms that can mitigate imbalance and improve sensitivity to minority classes [19], [20]. Furthermore, when combined with the binary decomposition approach, XGBoost has been extended to effectively capture the ordinal structure of data [21].

Building on these strengths, this study develops an Ordinal XGBoost framework that integrates binary decomposition with Enhanced Instance Weighting to better handle imbalanced ordinal data. By improving the accuracy and interpretability of nutritional status classification, the proposed approach contributes to public health monitoring and supports the achievement of SDG targets through stronger evidence for targeted stunting interventions.

2. LITERATURE REVIEW

2.1. Ordinal Classification and Class Imbalance Challenges

Ordinal classification is a specific variant of multi-class classification in which the labels possess a natural order, but the distance between classes is not necessarily numerically defined [22]. For example, the categories very low < low < medium < high < very high have an inherent order, yet the "distance" between categories is not constant.

Formally, if $C = \{c_1, c_2, \ldots, c_K\}$ a is the set of ordinal classes, there exists a total order relation as expressed in Equation (2.1) that must be considered [23], [24].

$$c_1 < c_2 < \cdots < c_K \tag{2.1}$$

On the other hand, class imbalance occurs when the distribution of data among classes is not uniform [25]. In ordinal classification, imbalance can appear in several forms, such as: a) minority at the lower extreme (extreme low), e.g., the number of stunted toddlers being significantly smaller than normal; b) minority at the upper extreme (extreme high), e.g., children with heights far above average being rare; and c) minority in the middle class, which, although rare, can occur in transitional categories.

Class imbalance in ordinal data has a dual impact: a) bias toward the majority class – the model tends to predict the majority class, sacrificing accuracy for minority classes; and b) degradation of order information. If extreme classes are rare, the model struggles to learn the differences between ordinal thresholds, leading to inconsistent order in predictions [16], [26].

These conditions demand the use of classification methods that can both preserve the order information between classes and remain robust to class imbalance. One algorithm

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

meeting both criteria is Extreme Gradient Boosting (XGBoost) with the binary decomposition approach, which applies the principles of gradient boosting with loss function optimization and decision tree regularization. Additionally, to further address the challenge of class imbalance, Enhanced Instance Weighting (EIW) can be incorporated into the XGBoost framework. EIW dynamically adjusts instance weights during training, giving more importance to underrepresented or misclassified instances, thereby ensuring that the model gives adequate focus to minority class samples and improving overall model performance in imbalanced datasets.

2.2. XG Boost

Extreme Gradient Boosting (XGBoost) is an advanced ensemble learning algorithm based on gradient boosting decision trees [27]. Given a dataset $D = \{(\boldsymbol{x}_i, y_1)\}$, where $\boldsymbol{x}_i \in \mathbb{R}^m$ denotes the feature vector and $y_i \in \mathbb{R}$ the corresponding label, XGBoost constructs a strong learner as an additive ensemble of regression trees shown in Equation (2.2).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i)\,, f_k \in \mathcal{F} \tag{2.2}$$

Where $\mathcal{F}$ is the functional space of regression trees and K is the number of trees. Since $\mathcal{F}$ represents the functional space of regression trees, which encompasses all the possible functions that can be formed by the decision trees in the model, the learning objective in XGBoost aims to find the optimal combination of these functions. Specifically, XGBoost strives to identify the best set of trees $f_k \in \mathcal{F}$ where $k = 1,2,3, \ldots, K$ that minimize a combined loss function and regularization term. By doing so, XGBoost not only improves the model's predictive performance but also controls the complexity of the functional space, ensuring that the model is both accurate and capable of generalizing well to new, unseen data.

The learning objective combines a differentiable convex loss function $l(y_1, \hat{y}_i)$ that measures the difference between the prediction and the true label, and a regularization term $\Omega(f_k)$ that penalizes model complexity shown in Equation (2.3).

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_{\mathrm{i}}, \hat{y}_i) + \sum_{t=1}^{K} \Omega(f_k) \tag{2.3}$$

The regularization term is defined in Equation (2.4).

$$\Omega(f) = \gamma \mathrm{T} + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2.4}$$

where T denotes the number of leaves, w_j is the score on leaf j, and γ, λ are regularization parameters that control model complexity. To optimize the objective, XGBoost employs an additive training strategy. At the t-th iteration, a new tree f_t is added to minimize the second-order Taylor approximation of the objective in Equation (2.5).

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{2.5}$$

where

$$g_i = \frac{\partial l\left(y_{\mathrm{i}}, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}}, h_i = \frac{\partial^2 l\left(y_{\mathrm{i}}, \hat{y}_i^{(t-1)}\right)}{\partial \left(\hat{y}_i^{(t-1)}\right)^2}$$

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

represent the first and second order gradients, respectively. The use of second-order information allows XGBoost to achieve both accuracy and computational efficiency, distinguishing it from traditional gradient boosting algorithms. The loss function $l(y_1, \hat{y}_i)$ is modified to account for ordinal distances, such as through ordinal logistic loss or proportional odds penalty, where errors between adjacent classes receive lower penalties than errors involving large jumps [21]. Considering ordinal distances can also be achieved using the binary decomposition approach.

2.3. XGBoost with Binary Decomposition

Frank dan Hall [16] proposed the Binary Decomposition approach, which transforms an ordinal classification problem with M classes into a series of $M-1$ binary classification tasks. Suppose the ordinal classes are $Y \in \{c_1, c_2, \dots, c_M\}\}$ with order $c_1 < c_2 < \cdots < c_M$, so for every threshold k subproblem is created as in Equation (2.6),

$$f_k\text{: } \{Y \leq c_k\} \text{ vs } \{Y > c_k\},\ k = 1, \dots, M-1 \tag{2.6}$$

where:

f_k: the binary model for the k-th threshold

$\{Y \leq c_k\}$: the positive class (1),

$\{Y > c_k\}$: the negative class (0).

Then, cumulative binary labels are created for each data point (x_i, y_i) as shown in Equation (2.7).

$$y_{i,k} = \begin{cases} 1, \text{if } y_i \leq c_k \\ 0, \text{if } y_i > c_k \end{cases} \tag{2.7}$$

with:

$y_{i,k} = 1$ if the original class of y_i is at or bel the threshold c_k.

$y_{i,k} = 0$ if the original class of y_i is above the threshold c_k.

This transforms the ordinal regression problem into $M-1$ binary classification problems. For each binary classification task k, the prediction model shown in Equation (2.8).

$$\hat{y}_{i,k} = \sum_{t=1}^{T} f_{t,k}(\boldsymbol{x}_i) \tag{2.8}$$

Where $f_{t,k}(\boldsymbol{x}_i)$ represents the t-th tree in the k-th binary classifier, dan T is the number of trees. The goal is to minimize the following objective function in Equation (2.9) for each binary classifier k.

$$\mathcal{L}_k(\phi_k) = \sum_{i=1}^{n} \mathcal{L}_{bin}(y_{i,k}, \hat{y}_{i,k}) + \sum_{t=1}^{T} \Omega(f_{t,k}) \tag{2.9}$$

Where $\mathcal{L}_{bin}(y_{i,k}, \hat{y}_{i,k})$ is the binary cross entropy loss for k-th classifier, and $\Omega(f_{t,k})$ is the regularization term. The binary cross entropy liss is defined as Equation (2.10)

$$\mathcal{L}_{bin}(y_{i,k}, \hat{y}_{i,k}) = -[y_{i,k} \log(\hat{y}_{i,k}) + (1 - y_{i,k}) \log(1 - \hat{y}_{i,k})]$$

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

where $y_{i,k}$ is the binary target for the k-th classifier, and $\hat{y}_{i,k}$ is the predicted probability that $y_i \leq c_k$. For the overall XGBoost objective function with binary decomposition, we sum the individual binary classification losses across all $M-1$ classifiers shown in Equation (1.10)

$$\mathcal{L}(\phi) = \sum_{k=1}^{M-1}\left(\sum_{i=1}^{n} \mathcal{L}_{bin}(y_{i,k}, \hat{y}_{i,k}) + \sum_{t=1}^{T} \Omega(f_{t,k})\right) \quad (2.10)$$

This combined objective is minimized to find the optimal set of trees $f_{t,k}$ for each of the binary tasks k. To address the challenges of imbalanced data, Enhanced Instance Weighting (EIW) can be used in XGBoost by adjusting the instance weights for each data point based on its class distribution or difficulty level. This adjustment increases the importance of minority class instances, allowing the model to focus more on them during training. The objective function for those shown in Equation (2.11).

$$\mathcal{L}(\phi) = \sum_{k=1}^{M-1}\left(\sum_{i=1}^{n} w_i\mathcal{L}_{bin}(y_{i,k}, \hat{y}_{i,k}) + \sum_{t=1}^{T} \Omega(f_{t,k})\right) \quad (2.11)$$

w_i is the instance weights, where $w_i = \frac{n}{n_{y_i}}$, n is the number of samples and n_{y_i} is the number of samples in the class of y_i. This ensures that the minority class receives a larger weight, helping the model to pay more attention to the minority class during training. For imbalanced datasets, the majority class will have smaller weights.

2.4. Evaluation Metrics for Ordinal Classification

Evaluating an ordinal classification model requires metrics that not only measure prediction correctness but also respect the inherent class order and the impact of errors across classes [24]. In this study, four primary metrics were used: Accuracy, Ordinal Mean Absolute Error (MAE), Quadratic Weighted Kappa (QWK), and Macro-F1.

a. Accuracy

measures the proportion of correct predictions over total test data Equation (2.12).

$$Accuracy = \frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = y_i)}{N} \quad (2.12)$$

where:

N: number of samples,

y_i: actual label/category,

$\hat{y}_i$: predicted label/category,

$\mathbf{1}(\cdot)$: indicator function, which equals 1 if the condition is true and 0 otherwise.

b. Ordinal Mean Absolute Error (MAE)

The MAE for ordinal data measures the average absolute distance between the actual and predicted labels. MAE accounts for the distance between classes, so errors across distant classes (e.g., 0 → 2) receive a larger penalty compared to closer errors (e.g., 0 → 1), as shown in Equation (2.13) [28].

$$\mathrm{MAE}_{\mathrm{ord}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|\left(1 - t_i^{(k)}\right)\log(1 - p_i^{(k)}) \quad (2.13)$$

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

where:

N: number of samples,

y_i: actual label/category,

$\hat{y}_i$: predicted label/category,

$t_i^{(k)} = 1$ if the original class of y_i is at or below the threshold c_k

$p_i^{(k)} = \sigma\left(F^{(k)}(x_i)\right) = \frac{1}{1 + e^{-F^{(k)}(x_i)}}$: cumulative probability $P(Y \leq c_k \mid x_i)$

c. Quadratic Weighted Kappa (QWK)

QWK measures the level of agreement between predictions and actual labels, applying a quadratic penalty for distant errors [29], [30], [31], [32]. QWK ranges from −1 (complete disagreement) to 1 (perfect agreement), with 0 indicating performance equivalent to random guessing. QWK is calculated using Equation (2.14)

$$QWK = 1 - \frac{\sum_{ij} W_{ij} O_{ij}}{\sum_{ij} W_{ij} E_{ij}} \tag{2.14}$$

where:

$$W_{ij} = \frac{(i-j)^2}{(K-1)^2}$$

O_{ij}: number of observations with actual label/category i and predicted label/category j

E_{ij}: number of observations expected under random distribution.

d. Macro-F1 Score

Macro-F1 calculates the average F1-score per class without considering class proportions [33], [34]. is more representative for imbalanced data because it assigns equal weight to each class, ensuring minority class performance is monitored. Macro-F1 is calculated using Equation (2.15).

$$\mathrm{F1}_k = \frac{2 \cdot \mathrm{Precision_k} \cdot \mathrm{Recall_k}}{\mathrm{Precision_k} + \mathrm{Recall_k}} \tag{2.15}$$

$$Macro\ F1 = \frac{1}{K} \sum_{k=1}^{K} \mathrm{F1}_k$$

where

$$\mathrm{Precision_k} = \frac{TP_k}{TP_k + FP_k},$$

$$\mathrm{Recall_k} = \frac{TP_k}{TP_k + FP_k}$$

TP_k: True Positive for the k-th class

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

FP_k: False Positive for the k-th class

FN_k: False Negative for the k-th class

3. METHODOLOGY

3.1. Research Data

The data utilized in this study is secondary data sourced from previous research [35].The study's population consisted of mothers with young children in Wajak district, while the sample was drawn from mothers with young children in Sumberputih village. A stratified random sampling method was applied, which is a probability sampling technique where the population is divided into subgroups or strata based on specific characteristics. A random sample is then selected from each stratum to represent the entire population. The sample size was calculated using the Slovin formula, with a 10% margin of error and an 80% anticipated response rate, considering the total population of 389.

$$n = \frac{N}{1 + N(e^2)} \frac{1}{rr} = \frac{389}{1 + 389(0{,}8^2)} \frac{1}{0{,}8}$$
$$= 99{,}437 \approx 100$$

The sample size for this study was 100 respondents. The research instrument underwent testing for validity and reliability. These two concepts are essential in questionnaire design, ensuring that the instrument measures what it is intended to and yields consistent results. Validity refers to the degree to which the questionnaire accurately measures the intended construct, while reliability refers to the consistency of the results across time, items, or raters. The data structure in this study is presented in Table 1.

Table 1. Data Structure

No.	X_1	X_2	X_3	X_4	Y
1	3.447	3.262	3.502	3.213	2
2	3.712	3.931	3.459	3.231	3
3	2.738	2.774	2.905	2.532	1
4	3.436	3.093	3.324	3.150	2
5	3.472	3.216	3.483	3.562	2
⋮	⋮	⋮	⋮	⋮	⋮
100	2.978	3.152	2.829	3.715	2

Based on Table 1, The predictor variables in this study included Economic Status (X_1), Health Services (X_2), Children's Diet (X_3), and Environment (X_4). The response variable was the Nutritional Status of Toddlers (Y), categorized into three groups: stunted, normal, and tall, represented by categories 1, 2, and 3, respectively.

3.2. Research Stages

The research was conducted in several systematic stages to ensure methodological rigor and reproducibility. The research stage shown in Figure 1.

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

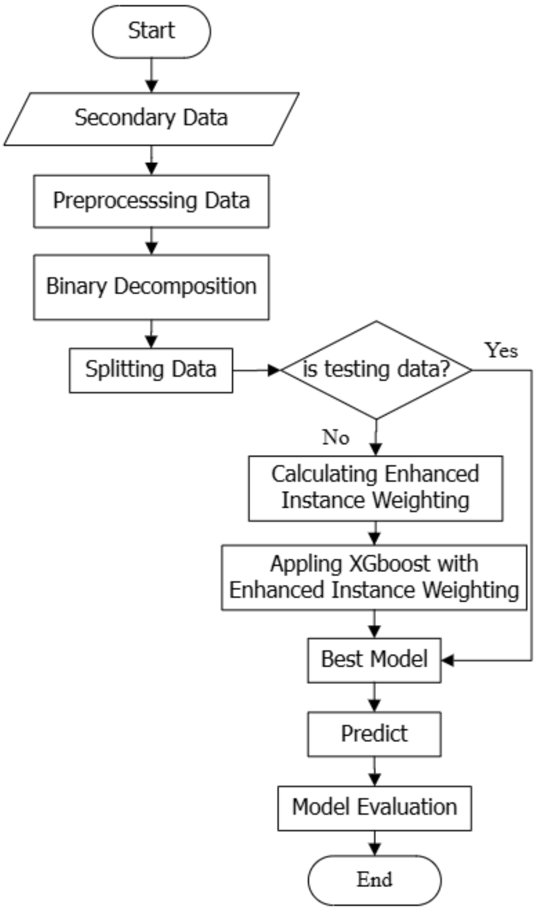


Fig 1. Research Flowchart

4. RESULTS AND DISCUSSION

4.1. Descriptive Statistics

The percentage of toddlers with stunted, normal, and tall nutritional status is presented in Figure 2.

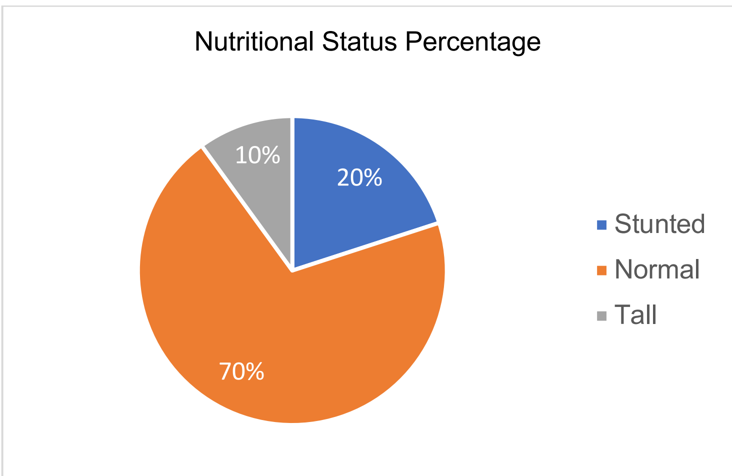


Fig 2. Percentage of Toddler Nutritional Status Categories

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

Based on Figure 2, the percentage of toddlers with stunting is 20%, those with normal height is 70%, and those with tall status is 10%. This condition indicates that the nutritional status of toddlers in Sumberputih Village is imbalanced.

4.2. XGboost Result

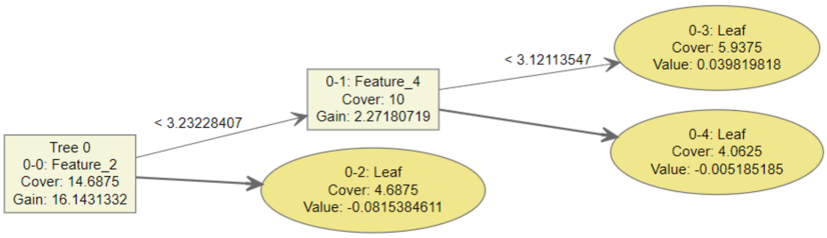
After the data were transformed through binary decomposition, they were divided into training and testing sets, followed by the calculation of Enhanced Instance Weighting. The values of Enhanced Instance Weighting are presented in Table 2.

Table 2. Data Structure and Instance Weighting Training Data

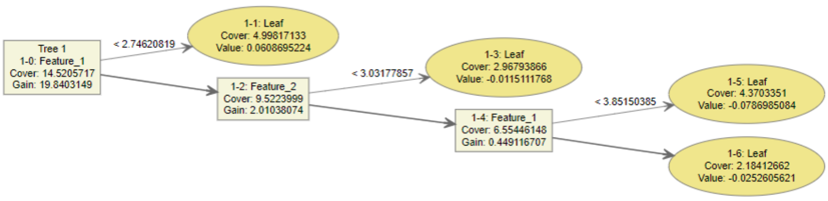
No.	X_1	X_2	X_3	X_4	Y	Y_1	Y_2	W_1	W_2
1	2.309	3.068	2.748	2.913	0	1	1	2.5	0.563
2	2.116	2.708	3.261	2.585	1	0	1	0.625	0.563
3	3.565	3.217	4.188	3.663	1	0	1	0.625	0.563
4	3.447	3.761	3.649	3.685	1	0	1	0.625	0.563
5	3.189	3.316	3.087	3.147	1	0	1	0.625	0.563
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	2.776	3.219	3.137	3.403	0	1	1	2.5	0.563

Table 2 presents the data structure after binary decomposition in the training data along with the Instance Weighting values. The variables include Economic Status (X_1), Health Services (X_2), Children's Diet (X_3), and Environment (X_4), dan Nutritional Status of Toddlers (Y) categorized into three groups: stunted (0), normal (1), and tall (2). Y_1 and epresent the outcomes of the binary decomposition, where Y_1 denotes the response group for stunted (1) versus normal and tall (0), and Y_2 denotes the response group for stunted and normal (1) versus tall (0). W_1 corresponds to the Instance Weighting for Y_1 and W_2 corresponds to the Instance Weighting for Y_2.

XGBoost results from the training data for binary decomposition subtask 1 (Y_1) are presented as follows.

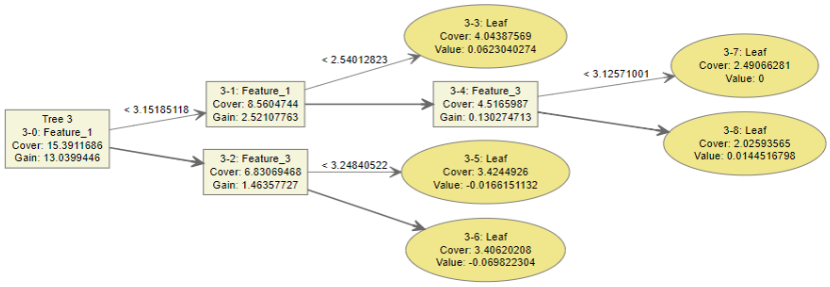


(a)



(b)

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

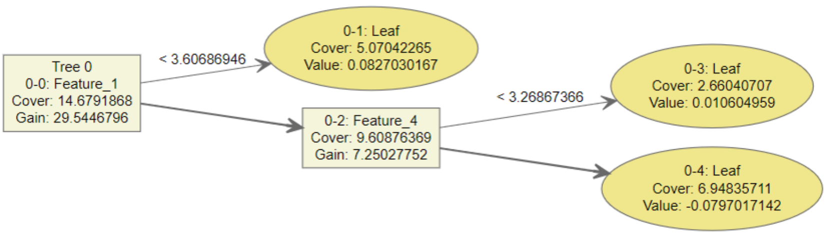


(c)

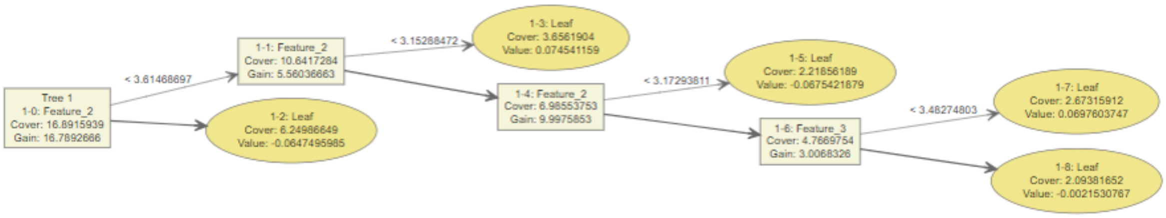
Fig 3. Decision Tree for Subtask 1 (Y_1) (a) first iteration, (b) second iteration, (c) third iteration

Figure 3 presents the visualization of the decision tree for the first three iterations. In subtask 1, a total of 79 iterations were performed. The most important variables obtained from the decision tree for subtask 1, in order of important, were Health Services (X2), Economic Status (X1), Environment (X4), dan Children's Diet (X3).

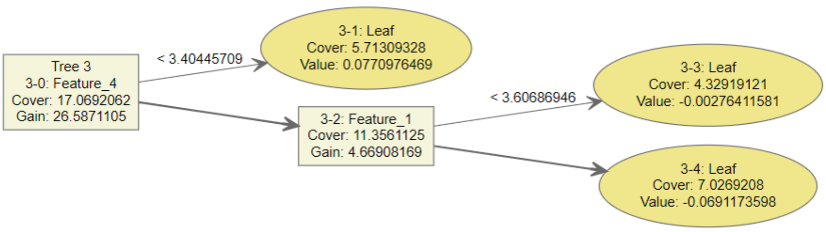
he XGBoost results from the training data for the binary decomposition of subtask 2 (Y_2) are presented as follows.



(a)



(b)



(c)

Fig 4. Decision Tree for Subtask 2 (Y_2) (a) first iteration, (b) second iteration, (c) third iteration

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

Figure 4 illustrates the visualization of the decision tree during the first three iterations. In subtask 2, a total of 187 iterations were performed. The most important variables obtained from the decision tree for subtask 2, in order of importance, were Environment (X_4), Children's Diet (X_3), Health Services (X_2), dan Economic Status (X_1).

The decision tree models from each subtask were then tested on the testing data to obtain prediction results. The predictions from subtask 1, subtask 2, and their combined outcomes are presented in Table 3.

Table 3. Data Test Predicted

No.	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}$
1	0	0	2
2	1	0	1
3	1	0	1
4	1	0	1
5	1	0	1
⋮	⋮	⋮	⋮
20	1	0	1

Table 3 presents the prediction results on the testing data for subtask 1 and subtask 2. In subtask 1, a prediction value of 1 indicates stunting, while a value of 0 indicates either normal or tall status. In subtask 2, a prediction value of 1 indicates stunting or normal, whereas a value of 0 indicates tall status. The combined prediction ($\hat{Y}$) represents the integration of results from subtask 1 and subtask 2. If both subtasks produce a prediction of 0, the combined result is 2 (tall); if both subtasks produce a value of 1, the combined result is 0 (stunting); otherwise, the combined result is 1 (normal).

4.3. Confusion Matrix

Testing the XGBoost model with the Binary Decomposition approach on the test data produced the confusion matrix shown in Table 1. The target classes consist of three ordinal categories: 0 (stunting), 1 (normal), and 2 (tall).

Table 4. Confusion Matrix of Prediction Results

Prediction	Actual		
	Stunting	Normal	Tall
Stunting	1	0	0
Normal	3	14	1
Tall	0	0	0

Based on Table 4, it can be seen that the XGBoost method with the Binary Decomposition approach was able to correctly predict 15 out of 20 toddlers' nutritional status. On the other hand, the model was most accurate in predicting the majority class (normal), while a small portion of stunting cases were classified as normal, and most tall cases were misclassified as normal.

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

4.4. Model Evaluation

The evaluation was carried out using four metrics relevant for ordinal classification on imbalanced data: Accuracy, Ordinal Mean Absolute Error (MAE), Quadratic Weighted Kappa (QWK), and Macro-F1 shown in Table 5.

Table 5. Model Evaluation Result

Metric	Value
Accuracy	0.75
MAE (Ordinal)	0.25
QWK	0.22
Macro F1	0.39

The results in Table 5 show that an Accuracy of 0.75 indicates the model was able to make correct predictions for 75% of the test data. 75%, which is a relatively high percentage for data with an imbalanced class distribution. However, as noted in [24] regarding the *Ordinal Classification* method, accuracy in ordinal data often does not fully reflect the quality of predictions, as this metric does not account for the distance between classes. Therefore, additional metrics such as ordinal MAE and QWK were used to provide a more comprehensive evaluation.

An ordinal MAE of 0.25 signifies that, on average, the model's prediction error was only 0.25 class levels from the actual value, demonstrating good preservation of class order. This finding is consistent with the results of [36], [37] on disease severity classification, which reported that ordinal regression–based approach with a specialized loss function could reduce MAE compared to conventional multi-class classification. In the context of this study, the low MAE suggests that the model effectively leveraged class order information through binary decomposition.

A QWK of 0.22 represents a low level of agreement between the model's predictions and the actual labels, with greater penalties for errors involving distant classes. This score falls within the moderate agreement category according to [38]. QWK is an important metric because it applies a quadratic penalty for errors that involve large class jumps. The study by [32] on Ordinal XGBoost also emphasizes that QWK is a sensitive indicator for assessing prediction quality in ordinal data, especially when class distribution is uneven. The QWK value in this study indicates that the model was relatively successful in maintaining class order, although it was not yet optimal for extreme classes.

A Macro-F1 score of 0.39 indicates relatively balanced performance across classes, although performance in minority classes remains lower than that on the majority class. This phenomenon is also reported by [26], [28] which state that in ordinal data with class imbalance, models tend to be biased toward the most frequent class. In this study, such bias was evident in the confusion matrix, where the normal class was predicted well, while the stunting and tall classes were more often misclassified as normal.

4.5. Discussion

This study demonstrates that the XGBoost method with a Binary Decomposition approach is able to classify toddlers' nutritional status with reasonable accuracy despite the challenges of ordinal and imbalanced data. The relatively high accuracy (0.75) and low MAE (0.25) indicate that the model successfully preserved the order of the classes. This aspect is

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

particularly important in nutritional status classification, as misclassifications between adjacent categories (e.g., stunting → normal) are more acceptable than misclassifications across extreme categories (e.g., stunting → tall).

Nevertheless, the low values of QWK (0.22) and Macro-F1 (0.39) confirm the presence of prediction bias toward the majority class (normal). This outcome reflects the limitation of the model when faced with highly skewed data distributions. Previous studies have highlighted that in ordinal classification with extreme imbalance, models often fail to establish clear decision boundaries for minority classes [13], [26]. In practice, this means that strong performance on the normal class is achieved at the cost of reduced accuracy in detecting stunting and tall cases, which are clinically more critical for early intervention.

The analysis of predictor variables provides further insights into the determinants of nutritional status. In the first subtask (stunting vs normal or tall), health services (X_2) emerged as the most influential variable. This finding is consistent with public health literature emphasizing the role of healthcare access and quality in preventing stunting through immunization, growth monitoring, and nutritional counseling [1], [2]. Limited access to healthcare directly increases children's vulnerability to stunting.

Conversely, in the second subtask (stunting or normal vs tall), environmental factors (X_4) played the most important role. A supportive and clean environment is associated with reduced risk of infectious diseases and improved nutritional outcomes. In this context, environment encompasses not only sanitation but also access to clean water, housing conditions, and the surrounding socioeconomic support. Children's diet (X3) also appeared as a key factor, underscoring the direct relationship between dietary intake and the likelihood of children achieving above-average height.

Taken together, these findings suggest that stunting prevention strategies should not be limited to improving household economic status, but should also prioritize strengthening access to primary healthcare services and improving environmental conditions. The integration of these factors into public health interventions has the potential to enhance the effectiveness of stunting prevention programs.

From a methodological perspective, this study highlights that although XGBoost with Binary Decomposition is effective in maintaining ordinal structure, additional approaches are needed to improve sensitivity to minority classes. Oversampling techniques, cost-sensitive loss functions, or hybrid methods (e.g., combining with deep learning–based ordinal classifiers) could potentially enhance detection performance for stunting and tall categories, which are relatively rare in the dataset.

5. CONCLUSION

This study addressed the challenge of classifying toddlers' nutritional status in imbalanced ordinal data, where preserving class order and detecting minority categories are crucial for meaningful predictions. Using the XGBoost method with a Binary Decomposition approach, the model achieved promising results, with an accuracy of 0.75 and an ordinal MAE of 0.25, demonstrating strong order preservation, although QWK (0.22) and Macro-F1 (0.39) indicated limited performance on minority classes. Variable importance analysis revealed that access to health services was the key determinant in distinguishing stunted children, while environmental factors played the dominant role in identifying tall status, offering practical insights for targeted public health interventions. However, the study was limited by a relatively

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

small sample size and severe class imbalance, which constrained the model's ability to establish optimal decision boundaries for minority categories. Future research should explore larger and more diverse datasets, as well as advanced techniques such as oversampling, cost-sensitive loss functions, or hybrid approaches, to improve the detection of stunting and tall status while maintaining ordinal consistency.

ACKNOWLEDGMENT

I sincerely thank all iyaparties for their valuable support and input throughout the preparation of this research, which has greatly contributed to its improvement.

REFERENCES

[1] T. T. Tamir, S. A. Gezhegn, D. T. Dagnew, A. T. Mekonenne, G. T. Aweke, and A. M. Lakew, "Prevalence of childhood stunting and determinants in low and lower-middle income African countries: Evidence from standard demographic and health survey," *PLoS One*, vol. 19, no. 4, p. e0302212, Apr. 2024, doi: 10.1371/journal.pone.0302212.

[2] A. J. Prendergast and J. H. Humphrey, "The stunting syndrome in developing countries," *Paediatr Int Child Health*, vol. 34, no. 4, pp. 250–265, Nov. 2014, doi: 10.1179/2046905514Y.0000000158.

[3] Y. Anggraini and N. F. Romadona, "Review of Stunting in Indonesia," in *Proceedings of the International Conference on Early Childhood Education and Parenting 2019 (ECEP 2019)*, Paris, France: Atlantis Press, 2020. doi: 10.2991/assehr.k.200808.055.

[4] R. R. Akbar, W. Kartika, and M. Khairunnisa, "The Effect of Stunting on Child Growth and Development," *Scientific Journal*, vol. 2, no. 4, pp. 153–160, Jul. 2023, doi: 10.56260/sciena.v2i4.118.

[5] P. Ssentongo *et al.*, "Global, regional and national epidemiology and prevalence of child stunting, wasting and underweight in low- and middle-income countries, 2006–2018," *Sci Rep*, vol. 11, no. 1, p. 5204, Mar. 2021, doi: 10.1038/s41598-021-84302-w.

[6] Sheik Mohamed S. H., "CHILD NUTRITION AND GROWTH MONITORING IN MOLDOVA: CHALLENGES AND INTERVENTIONS," *JTH: Journal of Technology and Health*, vol. 2, no. 4, pp. 187–195, May 2025, doi: 10.61677/jth.v2i4.161.

[7] M. de Onis and F. Branca, "Childhood stunting: a global perspective," *Matern Child Nutr*, vol. 12, no. S1, pp. 12–26, May 2016, doi: 10.1111/mcn.12231.

[8] B. Soetono and A. S. Barokah, "Trends in Stunting Prevalence Reduction: an Examination of Data Toward Achieving the 2024 Target in Indonesia," *Article The Social Perspective Journal*, vol. 3, no. 1, pp. 51–68, 2024, doi: 10.53947/tspj.v3i1.795.

[9] FAO, IFAD, UNICEF, WFP, and WHO, *The State of Food Security and Nutrition in the World 2024 – Financing to end hunger, food insecurity and malnutrition in all its forms*. Rome: FAO; IFAD; UNICEF; WFP; WHO;, 2024. doi: 10.4060/cd1254en.

[10] United Nations Children's Fund, "Joint child malnutrition estimates," Who. Accessed: Jul. 07, 2025. [Online]. Available: https://data.unicef.org/resources/levels-and-trends-in-child-malnutrition-2018/%0Awww.who.int/nutgrowthdb/estimates%0Ahttps://www.who.int/data/gho/data/themes/topics/joint-child-malnutrition-estimates-unicef-who-wb

[11] G. Aguiar, B. Krawczyk, and A. Cano, "A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework," *Mach Learn*, vol. 113, no. 7, pp. 4165–4243, Jul. 2024, doi: 10.1007/s10994-023-06353-6.

[12] K. De Angeli *et al.*, “Class imbalance in out-of-distribution datasets: Improving the robustness of the TextCNN for the classification of rare cancer types,” *J Biomed Inform*, vol. 125, p. 103957, Jan. 2022, doi: 10.1016/j.jbi.2021.103957.

[13] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.

[14] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, “A survey on addressing high-class imbalance in big data,” *J Big Data*, vol. 5, no. 1, p. 42, Dec. 2018, doi: 10.1186/s40537-018-0151-6.

[15] T. L. Nxumalo, R. M. Rimiru, and V. M. Magagula, “A Deep Learning Ordinal Classifier,” *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 3, 2025, doi: 10.14569/IJACSA.2025.0160330.

[16] E. Frank and M. Hall, “A Simple Approach to Ordinal Classification,” 2001, pp. 145–156. doi: 10.1007/3-540-44795-4_13.

[17] S. Kim, Y. Lym, and K.-J. Kim, “Developing Crash Severity Model Handling Class Imbalance and Implementing Ordered Nature: Focusing on Elderly Drivers,” *Int J Environ Res Public Health*, vol. 18, no. 4, p. 1966, Feb. 2021, doi: 10.3390/ijerph18041966.

[18] R. Cruz, K. Fernandes, J. F. Pinto Costa, M. P. Ortiz, and J. S. Cardoso, “Ordinal Class Imbalance with Ranking BT - Pattern Recognition and Image Analysis,” L. A. Alexandre, J. Salvador Sánchez, and J. M. F. Rodrigues, Eds., Cham: Springer International Publishing, 2017, pp. 3–12.

[19] J. Al Amien, H. Ab Ghani, N. I. Md Saleh, E. Ismanto, and R. Gunawan, “Intrusion detection system for imbalance ratio class using weighted XGBoost classifier,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 21, no. 5, p. 1102, Oct. 2023, doi: 10.12928/telkomnika.v21i5.24735.

[20] C. Wang, C. Deng, and S. Wang, “Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost,” *Pattern Recognit Lett*, vol. 136, pp. 190–197, Aug. 2020, doi: 10.1016/j.patrec.2020.05.035.

[21] F. Kahl, I. Kahl, and S. M. Jonas, “XGBOrdinal: An XGBoost Extension for Ordinal Data,” in *Stud Health Technol Inform*, 2025, pp. 462–466. doi: 10.3233/SHTI250380.

[22] A. Agresti, “Analysis of Ordinal Categorical Data - Examples of Using R for Modeling Ordinal Data Summary of R (and S-Plus),” *Analysis*, 2010.

[23] P. McCullagh, “Regression Models for Ordinal Data,” *J R Stat Soc Series B Stat Methodol*, 1980, doi: 10.1111/j.2517-6161.1980.tb01109.x.

[24] P. A. Gutiérrez and S. García, “Current prospects on ordinal and monotonic classification,” *Progress in Artificial Intelligence*, 2016, doi: 10.1007/s13748-016-0088-y.

[25] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Trans Knowl Data Eng*, 2009, doi: 10.1109/TKDE.2008.239.

[26] J. S. Cardoso and R. Sousa, “Measuring the performance of ordinal classification,” *Intern J Pattern Recognit Artif Intell*, 2011, doi: 10.1142/S0218001411009093.

[27] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. doi: 10.1145/2939672.2939785.

[28] S. Baccianella, A. Esuli, and F. Sebastiani, “Evaluation measures for ordinal regression,” in *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications*, 2009. doi: 10.1109/ISDA.2009.230.

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]

[29] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," *Psychol Bull*, 1968, doi: 10.1037/h0026256.

[30] A. Ben-David, "Comparison of classification accuracy using Cohen's Weighted Kappa," *Expert Syst Appl*, 2008, doi: 10.1016/j.eswa.2006.10.022.

[31] J. de la Torre, D. Puig, and A. Valls, "Weighted kappa loss function for multi-class classification of ordinal data in deep learning," *Pattern Recognit Lett*, 2018, doi: 10.1016/j.patrec.2017.05.018.

[32] A. Zhang, H. Hirai, M. Matsumaru, and H. Katagiri, "Corporate Rating Model using Threshold Optimization and XGBoost," in *International Conference on Industrial & Mechanical Engineering and Operations Management*, Dhaka, Bangladesh, 2020.

[33] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf Process Manag*, 2009, doi: 10.1016/j.ipm.2009.03.002.

[34] J. Brownlee, "Calculate Precision, Recall, and F-Measure for Imbalanced Classification," *Inf Process Manag*, 2020.

[35] A. A. R. Fernandes and Solimun, "Menelisik Faktor-Faktor Penyebab Stunting pada Anak di Kecamatan Wajak: Integrasi Cluster dengan Path Analysis dengan Pendekatan Statistika dan Sains Data," 2023.

[36] H. Zhu *et al.*, "Convolutional Ordinal Regression Forest for Image Ordinal Estimation," *IEEE Trans Neural Netw Learn Syst*, 2022, doi: 10.1109/TNNLS.2021.3055816.

[37] W. Li, X. Huang, J. Lu, J. Feng, and J. Zhou, "Learning probabilistic ordinal embeddings for uncertainty-aware regression," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021. doi: 10.1109/CVPR46437.2021.01368.

[38] J. R. Landis and G. G. Koch, "An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers," *Biometrics*, 1977, doi: 10.2307/2529786.

[DOI: 10.14710/JSUNIMUS.13.1.2025.45-60]