Variance
Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X

Variance: Journal of Statistics and Its Applications
Volume 6 Issue 2| October 2024
10.30598/variancevol6iss2page195-204

ORDINAL LOGISTIC REGRESSION BAGGING FOR MODELING AND CLASSIFICATION OF THE NUTRITIONAL STATUS OF TODDLERS IN SOUTHEAST PONTIANAK SUB-DISTRICT

Sekar Aulia Sista[1], Dadan Kusnandar[2], Neva Satyahadewi[3*]
[1,2,3] Statistics Study Program Mathematics Major Faculty of Mathematics and Natural Sciences Tanjungpura University, Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, West Kalimantan, Indonesia

Coresponding Author's E-mail: *neva.satya@math.untan.ac.id*

Abstract: *Although Pontianak's 2022 stunting rate of 19.7% is higher than the RPJMN's 2020–2024 target of 14%, this is still significant. The categories of stunts are very short (severely stunted), short (stunted), normal, and high, based on a high index of the body by age (TB/U). Ordinal Logistic Regression is one classification that can be used to group stunts based on the TB/U index. This approach makes the unstable parameter. Use the bagging to get stable parameters. The study aims to model and classify toddlers' nutritional status using the TB/U index. Utilizing secondary data for 150 toddlers from Pontianak Tenggara's UPT Puskesmas Parit Haji Husin II. This will monitor kids' growth from 24 to 59 months in 2022. Response factors include short, very short, normal, and high. The mother's job position, birth weight, length, and gender are the predictive variables. Due to imbalanced data utilized in the first analysis using Ordinal Logistics Regression, a decent model, and the final classification result, they used the Bagging OLR ensemble method. The study's findings are a very effective model using OLR Bagging, with an accuracy rate of 99.33%, a sensitivity value of 98.91%, and a specificity value of 98.52%. The results also revealed significant variables that influence the mother's employment status and the birth length variable.*

Keywords: *Bagging, Ordinal Logistic Regression, Toddler Nutritional Status.*

1. INTRODUCTION

Nutritional problems are still a health problem in Indonesia, and nutritional problems in certain groups can affect nutritional status in the next life cycle. Based on survey results from the Indonesian Nutrition Status Study (SSGI), the stunting rate 2022 in Pontianak City was recorded at 19.7%, which is down compared to 2021, namely 21.4%. However, this figure is still far from the prevalence rate set as the target in the 2020-2024 RPJMN, namely 14%. It was recorded in the 2018 Basic Health Research (Riskesdas) that the prevalence rate for height status per age (TB/U) for children aged 0-59 months in Pontianak City, for the Very Short category was 6.92%, the Short category was 15.34 %, and the Normal category is 77.74%.

Based on Minister of Health Regulation no. 2 of 2020 concerning Child Anthropometric Standards, Stunting (short according to age), which is measured through the height or body length index for age (TB/U or PB/U), is divided into four categories, namely very short (severely stunted), short (stunted), normal and high. You can use several classification methods to group data, including logistic regression, multivariate adaptive regression spline, and discriminant analysis. Logistic regression is one of the classification methods that is widely used; logistic regression is a data analysis technique in statistics that has the aim of seeing the relationship between several variables where the response variable (dependent variable) is categorical, in the form of nominal or ordinal with predictor variables (independent variables) which are categorical as well as continuous [1].

In this research, data was used in the form of nutritional status data for toddlers with continuous and categorical predictor variables and response variables with a multilevel categorical scale. With this data, logistic regression analysis is a good method for overcoming classification problems. However, the logistic regression classification method will provide parameter estimates that are not constant, which means that if there is a change in the dataset, it will cause a significant change in the model [2]. So, to get constant parameters in the logistic

Copyright© 2024: Author(s)
https://ojs3.unpatti.ac.id/index.php/variance/ *Submitted:2023-12-14* *Accepted: 2024-11-30*
This work is licensed with Lisensi Internasional 4.0 Atribusi Creative Commons-NonCommercial-ShareAlike 4.0.

regression model, you can use a method that provides better prediction accuracy, namely Bootstrap Aggregating (Bagging).

Bagging or "bootstrap aggregating" is used in several classification and regression methods because it can reduce the variance of predictor variables and improve the prediction process [3]. The logistic regression bagging method is used to increase classification accuracy and stabilize the estimation of model parameters from ordinal logistic regression.

Regression Bagging method was carried out by Suniantara, Putra, and Suwardika in 2019 with classification resulting in a classification error with an initial 36.63% dropping to 20.237% or increasing the classification accuracy level from 63.37% to 76.67 %. The classification accuracy of ordinal logistic regression bagging produces better numbers than ordinal logistic regression itself, so logistic regression bagging can increase the classification level when changes in the dataset occur [4].

Based on the explanation above, this research aims to apply the Ordinal Logistic Regression Bagging method in modeling and classifying indicators for assessing the nutritional status of toddlers into several categories based on the height index for age (TB/U), namely tall, normal, short (stunted) and very short. (seriously stunted).

2. METHODOLOGY

2.1. Data Sources and Data Types

The data used is secondary data obtained from the UPT of the Parit Haji Husin II Health Center, Southeast Pontianak District. In the form of growth monitoring data for toddlers for the 2022 period with a sample of toddlers aged 24-59 months with a total of 150 toddlers, with factors thought to influence the assessment of toddler nutritional status based on the TB/U index.

2.2. Research Variable

The response variables and predictors used in this study are listed in Table 1.

Table 1. Research Variable

Variable	Information	Scale
Indicators For Assessing The Nutritional Status of Toddlers Based on The TB/U Index (Y)	Tall (Y_0) Normal (Y_1) Short (Y_2) Very Short (Y_3)	Ordinal
Toddler Age (X_1)	24-59 Months	Ratio
Toddler Gender (X_2)	Female (1) Male (2)	Nominal
Toddler Body Length at Birth (X_3)	Cm	Ratio
Birth Weight (X_4)	Kg	Ratio
Mother's Employment Status (X_5)	Working (1) Housewife (IRT) (2)	Nominal

2.3. Analysis Procedure

The sequence of analysis of the research data it starts from inputting the data, carrying out descriptive statistical analysis, carried out by testing the parameters simultaneously with the G test, with the decision criteria being at least one predictor variable that can influence the response variable simultaneously and if not then a replacement will be made with other variables.

Partial parameter testing was carried out using the Wald test, with the decision criterion being that there were predictor variables that significantly affected the response variable. Then, the model obtained using the odds ratio is interpreted, and the classification accuracy is calculated in ordinal logistic regression (OLR). The next stage is to carry out analysis using Ordinal Logistic Regression Bagging, calculating the accuracy of the OLR

Bagging classification to obtain a classification value. Next, the model results are interpreted, and the accuracy of the Bagging OLR classification is achieved.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics

To obtain a general idea of the characteristics of data, descriptive statistical analysis can be used in Table 2.

Table 2. Descriptive Statistics of Qualitative Variables

Variable	Information	Category	Total	Percentage
X_2	Toddler Gender	1. Female 2. Male	84 66	56% 44%
X_5	Mother's Employment Status	1. Working 2. Housewife (IRT)	51 99	34% 66%
Y	Indicators For Assessing The Nutritional Status of Toddlers Based on The TB/U Index	0. Tall 1. Normal 2. Short 3. Very Short	1 127 17 5	0.7% 84.7% 11.3% 3.3%

From Table 2, it can be seen that female toddlers are more dominant than male toddlers by 56%. The employment status of mothers of toddlers who undergo examinations at the Community Health Center is 66% as a housewife (IRT). The indicator variable for assessing the nutritional status of toddlers based on the TB/U index was 84.7% of toddlers in the normal category. This causes an imbalance in the data used, which can cause a decrease in classification performance in minority classes or categories such as the tall, short, or very short categories. Imbalanced data is a situation where data is not balanced between one data class and another in a dataset [5].

Based on the results of the classification of toddlers based on the TB/U index with 4 categories: very short, short, normal, and tall. There is a child who is included in the category of children with tall stature; this can be a cause for concern because if later a child is very tall for his age (above normal) but the parents' height is normal, it is usually suspected of having an endocrine disorder. There are also 5 children with very short stature. This condition must be followed up so that the cause and treatment can be identified.

In assessing a child's nutritional status, the results of measuring body weight and length/height can be compared with the Children's Anthropometric Standards, which use 5 indices, including the TB/U index. This will be a reference in assessing nutritional status and growth trends in children. Underweight toddlers do not necessarily experience deficient or poor nutrition; if the child is stunted or very stunted, then his nutritional status can be said to be adequate or even over-nourished; with this, it can be said that determining nutritional status is necessary—consideration of all existing indices [6].

Table 3. Descriptive Statistics of Quantitative Variables

Variable	Information	Minimum	Average	Maximum
X_1	Toddler Age	24	35.83	59
X_3	Birth Lenght	43	49.08	58
X_4	Birth Weight	2.2	3.12	4.4

Table 3 shows that the average age of toddlers is 35.83 months, the average height of toddlers is 49.08 cm, and the average birth weight of toddlers is 3.12 kg.

3.2. Significant Test of Logistic Regression Model Parameters

Significance tests are divided into 2, namely simultaneous or simultaneous significant tests and partial or individual significant tests.

3.2.1.Simultaneous Test

Testing the regression coefficients simultaneously or simultaneously is a simultaneous test. The likelihood ratio test is used. The G test aims to test the role of predictor variables in the model simultaneously[7]. The test is carried out with decision criteria, namely reject H_0 if the value $G \geq x^2{}_{(\alpha,p)}$ or p-value <0.05, namely, $H_0 : \beta_0 = \beta_1 = \cdots = \beta_5 = 0$ (Predictor variables have no significant effect to the response variable simultaneously), H_1 : There is at least one $\beta_j \neq 0$, with $j = 1,2,\cdots,5$ (There is at least one predictor variable that has a significant effect on the response variable simultaneously).

Table 4. Simultaneous Logistic Regression Test Results

G	***Df***	$X^2_{(0.05;5)}$
21.031	5	11.0705

According to Table 4, information is obtained that the G value of the regression model formed is 21.031 with 5 degrees of freedom, which is greater than the value of $x^2_{(0.05;5)} = 11{,}0705$ so that H_0 is rejected. This means that there is a significant influence of at least one predictor variable on the response variable.

3.2.2.Partial Test

Partial test is a test of individual regression coefficients. Partial parameter testing uses the Wald test [8], testing is carried out with the decision criteria to reject H_0 if the value $W \geq X^2{}_{(\alpha,p)}$ or p-value <0.05, namely, H_0 : β_j=0,"with" j=1,2,⋯,5 (The predictor variable has no significant effect on the response variable partially), H_0 : $\beta_j = 0$, with $j = 1,2,\cdots,5$ (The predictor variable has a significant effect on the response variable partially).

Table 5. Partial Logistic Regression Test Results

Variable		**Estimate (*β*)**	***p-value***
Constant (0)	Y_0	24.499	0.000
Constant (1)	Y_1	16.610	0.006
Constant (2)	Y_2	14.809	0.014
Toddler Age	X_1	0.020	0.422
Toddler Gender			
-Female	X_2	0.007	0.989
-Male			
Birth Lenght	X_3	0.298	0.033
Birth Weight	X_4	0.919	0.191
Mother's Employment Status			
-Working	X_5	1.751	0.009
-Housewife			

Based on Table 5, it is known that 2 predictor variables have a significant effect on the nutritional status of toddlers, namely birth length (X_3), and mother's employment status (X_5). This is because the p-value is smaller than alpha 5% (0.05).

3.3. Odds Ratio

The interpretation of the logistic regression model used is by looking at the good ratio value. The odds ratio value is defined as follows:

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_0} \quad (1)$$

Odds ratio calculations are carried out only for significant variables. If the OR value = 1, the two variables have no relationship. If the OR value <1, then between the two variables, there is a negative relationship to changes in the category of the predictor variable. Likewise, if the OR value is> 1, then there is a positive relationship between the two variables with changes in the category of the predictor variable [9].

3.4. Ordinal Logistic Regression Model

Ordinal logistic regression can be applied to see the influence of a number of numerical or categorical predictor variables on categorical response variables in classification problems. [10]. The ordinal logistic regression model aims to determine the influence of predictor variables that influence indicators for assessing the nutritional status of toddlers based on the TB/U index. The ordinal logistic regression model obtained from partial test results is as follows.

$$\begin{aligned}\text{Logit}(\text{Y}_0) = {} & 24.499 + 0.020(\text{Toddler age}) + 0.007(\text{Gender}) + 0.298(\text{Birth Lenght})^* + \\ & 0.919(\text{Birth Weight}) + 1.751(\text{Mother's Employment Status})^*\end{aligned}$$

$$\begin{aligned}\text{Logit}(\text{Y}_1) = {} & 16.610 + 0.020(\text{Toddler age}) + 0.007(\text{Gender}) + 0.298(\text{Birth Lenght})^* + \\ & 0.919(\text{Birth Weight}) + 1.751(\text{Mother's Employment Status})^*\end{aligned}$$

$$\begin{aligned}\text{Logit}(\text{Y}_2) = {} & 14.809 + 0.020(\text{Toddler age}) + 0.007(\text{Gender}) + 0.298(\text{Birth Lenght})^* + \\ & 0.919(\text{Birth Weight}) + 1.751(\text{Mother's Employment Status})^*\end{aligned}$$

With (*) the predictor variable (significant) has an influence on the response variable.

Table 6. Odds Ratio Ordinal Logistic Regression Model

Variabel	Estimasi (β)	Odds Ratio
X_3	0.298	1.347
$\text{X}_{5(1)}$	1.751	5.760

The coefficient value of the toddler birth weight variable is 0.298, with an odds ratio value of 1.347. This shows a positive relationship with changes in the category of predictor variables; there is a tendency of 1.347 times for the nutritional status of toddlers to be better for each increase in birth length, assuming all other variables are considered constant. In other words, the taller the baby is at birth, the greater the chance of the baby having good nutrition.

The coefficient value for the mother's employment status variable is 1.751, with an odds ratio of 5.760. This shows a positive relationship with changes in the categories of predictor variables; there is an increase in the tendency of 5,760 times for a toddler whose mother works to have good nutrition compared to a toddler whose mother does not work.

3.5. Ordinal Logistic Regression Bagging Model

Bootstrap Aggregating (Bagging) is a method proposed by Breiman. Bagging is useful for improving the stability and accuracy of predictions, namely by reducing the variance of a predictor in classification and regression methods whose use is not limited to improving estimators. Multiple versions are formed by bootstrap replication of a dataset. Taking bootstrap samples by repeating $\{\mathcal{L}^B\}$ from data set L and forming $\{\varphi(x, \mathcal{L}^{(B)})\}$. When y is numeric data, we give $\varphi_B(x) = average_B \varphi(x, \mathcal{L}^{(B)})$. If y is a class label, then to determine the category by voting or averaging on $\{\varphi(x, \mathcal{L}^{(B)})\}$ from $\varphi_B(x)$. This bootstrap method is called bootstrap aggregating (bagging). [11].

Bagging is a method that can generally increase classification accuracy. So, bagging is used for this data in the hope of improving the model so that it can improve classification accuracy for the better. So the model formed is as follows.

$$\begin{aligned}\text{Logit}(\text{Y}_0) = {} & 27.264 + 0.015(\text{Toddler age}) + 0.329(\text{Gender}) + 0.325(\text{Birth Lenght})^* + \\ & 1.312(\text{Birth Weight}) + 2.208(\text{Mother's Employment Status})^*\end{aligned}$$

$$\begin{aligned}\text{Logit}(\text{Y}_1) = {} & 18.645 + 0.015(\text{Toddler age}) + 0.329(\text{Gender}) + 0.325(\text{Birth Lenght})^* + \\ & 1.312(\text{Birth Weight}) + 2.208(\text{Mother's Employment Status})^*\end{aligned}$$

$$\begin{aligned}\text{Logit}(\text{Y}_2) = {} & 16.927 + 0.015(\text{Toddler age}) + 0.329(\text{Gender}) + 0.325(\text{Birth Lenght})^* + \\ & 1.312(\text{Birth Weight}) + 2.208(\text{Mother's Employment Status})^*\end{aligned}$$

With (*) the predictor variable (significant) has an influence on the response variable.

Table 7. Odds Ratio Ordinal Logistic Regression Bagging Model

Variabel	Estimation (β)	Odds Ratio
X_3	0,325	1,384
$X_{5(1)}$	2,208	9,097

Based on this model, the coefficient value for the toddler birth length variable (X_3) significantly affects the toddler nutritional status indicator based on TB/U of 0.325. This shows with the odds ratio value that there is a tendency of 1.384 times for the nutritional status of toddlers to be better with each increase in body length at birth, assuming all other variables are considered constant; in other words, the longer the body is when the baby is born, the more likely the baby is to have good nutrition become bigger.

The coefficient value for the mother's employment status variable (X_5) is 2.208, with an odds ratio of 9.097. This shows an increased tendency of 9.097 times for a toddler from a working mother to have good nutrition compared to a toddler whose mother does not work. In other words, a toddler from a working mother needs to have good nutrition to become bigger.

3.6. A Measure of Classification Accuracy

Classification accuracy measures are used to determine whether the data is classified correctly or not. Table 8 is used to calculate classification accuracy as follows [12].

Table 8. Multiclass Confusion Matrix

Actual Y	**Predict Y**			
	Class 1	Class 2	Class 3	Class 4
Class 1	TP_1	TP_{12}	TP_{13}	TP_{14}
Class 2	TP_{21}	TP_2	TP_{23}	TP_{24}
Class 3	TP_{31}	TP_{32}	TP_3	TP_{34}
Class 4	TP_{41}	TP_{42}	TP_{43}	TP_4

The multiclass confusion matrix is a development of the binary confusion matrix where previously there were FP (False Positive), TP (True Positive), FN (False Negative), and TN (True Negative). By using table 1, the accuracy, specificity, sensitivity, and precision values can be calculated as follows [13]:

$$\textit{Accuracy} = \frac{(\mathrm{TP} + \mathrm{TN})}{\mathrm{N_{total}}} \tag{2}$$

$$\textit{Specificity} = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \tag{3}$$

$$\textit{Sensitivity} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{4}$$

$$\textit{Precision} = \frac{TP}{(TP + FP)} \tag{5}$$

The accuracy value is intended to measure the accuracy of the model classification when predicting data against all actual data, and sensitivity is used to see the likelihood that the prediction results are true positive when identifying true positive data. The higher the sensitivity value, the smaller the possibility of error in the prediction results. Specificity is used to measure the accuracy of negative predictions compared to all negative data, and precision is used to measure the ratio of true positive predicted results to all true positive actual data.

Table 9. Ordinal Logistic Regression Classification Results

Actual	***Predicted***			
	Tall	Normal	Short	Very Short
Tall	0	1	0	0
Normal	0	127	0	0

Actual	***Predicted***			
	Tall	Normal	Short	Very Short
Short	0	16	1	0
Very Short	0	4	1	0

It was found that the percentage value of accuracy or correctness of the Ordinal Logistic Regression model classification in predicting data against actual data was 85.33%. The sensitivity value is 26.47%; this illustrates that the level of accuracy of the toddler classification prediction results according to the actual correct data is very low, namely 26.47%. The specificity value is 76.98%, illustrating that the level of accuracy of specific predictions for toddlers from each category can be quite good, namely 76.98%. The precision value is 26.47%, which shows that 26.47% of the predictions for classified toddlers are based on the actual data.

This level of accuracy shows that the Ordinal Logistic Regression method is quite good in classifying the accuracy of nutritional status indicators for toddlers with the TB/U index. However, if the sensitivity and specificity values are low or unbalanced, then it can be said that the classification could be better [14]. An imbalance in the data used causes this, therefore causing a decrease in classification performance in the minority class. One of the ways to overcome the problem of class imbalance or imbalanced datasets is by combining (ensemble) methods. One of the most popular ensemble learning is bagging [15].

Table 10. Ordinal Logistic Regression Bagging classification results

Actual	***Predicted***			
	Tall	Normal	Short	Very Short
Tall	1	0	0	0
Normal	0	127	0	0
Short	0	1	16	0
Very Short	0	0	0	5

By combining the Ordinal Logistic Regression Bagging method, the percentage value of accuracy or accuracy of the classification of the Ordinal Logistic Regression Bagging model in predicting data against actual data is 99.33%. The sensitivity value of 98.91% illustrates that the level of accuracy of the toddler classification prediction results according to the actual correct data is good at 98.91%. The specificity value is 98.52%, illustrating that the level of accuracy of specific predictions for toddlers from each category can be said to be quite good, namely 98.52%. The precision value is 98.52%, which shows that 98.52% of the predictions for classified toddlers are based on the actual data.

This level of accuracy shows that the Ordinal Logistic Regression Bagging method is very good at classifying data. With the percentage values of sensitivity and specificity showing high and balanced values, this shows that the Ordinal Logistic Regression Bagging method is good in categorizing the accuracy of nutritional status of toddlers. This means that applying ordinal logistic regression bagging can increase classification accuracy by 14% and overcome dataset imbalances in classification.

4. CONCLUSIONS

The conclusions obtained based on the results of the analysis and discussion are as follows.

1) The Ordinal Logistic Regression Bagging Method is a method that can be used as a classification process. The classification value produced by the ordinal logistic regression method without Bagging is quite good, with an accuracy value of 85.33%. The Ordinal Logistic Regression Bagging Method increased classification accuracy from 85.33% in Ordinal Logistic Regression to 99.33% in Ordinal Logistic Regression Bagging, with an increase in classification accuracy of 14%.
2) From the results of the analysis carried out with five variables used, namely toddler age (X_1), toddler gender (X_2), toddler body length at birth (X_3), birth weight (X_4), and mother's employment status (X_5), variables that A significant influence on the nutritional status of toddlers based on the TB/U index is the variable Birth Body Length (X_3), and Maternal Employment Status (X_5) in the working mother category. It can be concluded that the body length of toddlers and the mother's employment status are factors that can influence

the assessment of children's nutritional status to be grouped into four categories according to the Height for Age index (TB/U).

3) The best model for toddler nutritional status based on the TB/U index uses the Ordinal Logistic Regression Bagging method.

$$\text{Logit}(Y_0) = 27.264 + 0.015(\text{Toddler age}) + 0.329(\text{Gender}) + 0.325(\text{Birth Lenght})^* + 1.312(\text{Birth Weight}) + 2.208(\text{Mother's Employment Status})^*$$

$$\text{Logit}(Y_1) = 18.645 + 0.015(\text{Toddler age}) + 0.329(\text{Gender}) + 0.325(\text{Birth Lenght})^* + 1.312(\text{Birth Weight}) + 2.208(\text{Mother's Employment Status})^*$$

$$\text{Logit}(Y_2) = 16.927 + 0.015(\text{Toddler age}) + 0.329(\text{Gender}) + 0.325(\text{Birth Lenght})^* + 1.312(\text{Birth Weight}) + 2.208(\text{Mother's Employment Status})^*$$

with (*) the predictor variable (significant) has an influence on the response variable.

REFERENCES

[1] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression: Hosmer/Applied Logistic Regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2000. doi: 10.1002/0471722146.

[2] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, z, doi: 10.1007/BF00058655.

[3] C. Sutton, “Adhesions and their prevention,” *Obstet. Gynaecol.*, vol. 7, no. 3, pp. 168–176, Jul. 2005, doi: 10.1576/toag.2005.7.3.168.

[4] I. K. P. Suniantara, I. G. E. W. Putra, and G. Suwardika, “Peningkatan Ketepatan Klasifikasi dengan Metode Bootstrap Aggregating pada Regresi Logistik Ordinal,” *INTENSIF J. Ilm. Penelit. Dan Penerapan Teknol. Sist. Inf.*, vol. 3, no. 1, p. 32, Feb. 2019, doi: 10.29407/intensif.v3i1.12587.

[5] R. D. Fitriani, H. Yasin, and T. Tarno, “Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal),” *J. Gaussian*, vol. 10, no. 1, pp. 11–20, Feb. 2021, doi: 10.14710/j.gauss.v10i1.30243.

[6] Kementrian Kesehatan Republik Indonesia, *Standar Antropometri Anak PMK Nomor 2*, Nomor 2. Jakarta, 2020.

[7] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, “Applied Logistic Regression, Third Edition,” *Hoboken N. J. John Wiley Sons 2013*, 2013.

[8] R. Ramandhani and D. Safitri, “Metode Bootstrap Aggregating Regresi Logistik Biner Untuk Ketepatan Klasifikasi Kesejahteraan Rumah Tangga Di Kota Pati,” vol. 6, no. 1, p. 10, 2017.

[9] A. F. Dewi and R. Pratiwi, “Analisis Regresi Logistik Biner pada Pengaruh Harga, Kualitas Pelayanan dan Promosi terhadap Kepuasan Pelanggan dalam Menggunakan Jasa Layanan Grab di Kabupaten Lamongan,” *Inferensi*, vol. 4, no. 2, p. 77, Sep. 2021, doi: 10.12962/j27213862.v4i2.8637.

[10] I. A. Adjei and R. Karim, “An Application of Bootstrapping in Logistic Regression Model,” *OALib*, vol. 03, no. 09, pp. 1–9, 2016, doi: 10.4236/oalib.1103049.

[11] P. P. Sari, M. Susilawati, and I. G. A. M. Srinadi, “Bootstrap Aggregating (Bagging) Regresi Logistik Ordinal Untuk Mengklasifikasikan Status Gizi Balita Di Kabupaten Klungkung,” *E-J. Mat.*, vol. 5, no. 3, p. 103, Aug. 2016, doi: 10.24843/MTK.2016.v05.i03.p128.

[12] K. Suryowati and M. T. Jatipaningrum, “Perbandingan Metode K-Nearest Neighbor Dan Random Forest Pada Klasifikasi Indeks Pembangunan Manusia Di Kabupaten/Kota Seluruh Indonesia”.

[13] N. Wang, N. N. Zeng, and W. Zhu, “Sensitivity, Specificity, Accuracy, Associated Confidence Interval And ROC Analysis With Practical SAS Implementations,” 2010.

[14] P. R. Sihombing and A. M. Arsani, "Comparison Of Machine Learning Methods In Classifying Poverty In Indonesia In 2018," *J. Tek. Inform. Jutif*, vol. 2, no. 1, pp. 51–56, Jan. 2021, doi: 10.20884/1.jutif.2021.2.1.52.

[15] B. W. Yap, K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin, and N. N. Abdullah, "An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets," in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, vol. 285, T. Herawan, M. M. Deris, and J. Abawajy, Eds., in Lecture Notes in Electrical Engineering, vol. 285. , Singapore: Springer Singapore, 2014, pp. 13–22. doi: 10.1007/978-981-4585-18-7_2.