

Determinants of Stunting Among Under-Five Years Children Using the Ordinal Logistic Regression Model

Marisa Rifada^{1,3*} Nur Chamidah^{1,3}, Putri Nuraini¹, Faris Dwiki Gunawan¹, Lailatul Muniroh^{2,3}

¹ Department of Mathematics, Universitas Airlangga, Surabaya, Indonesia

² Department of Health Nutrition, Universitas Airlangga, Surabaya, Indonesia

³ Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia

*Corresponding author: marisa.rifada@fst.unair.ac.id

ABSTRACT

Stunting is a kind of chronic nutritional problem among under-five years children that characterized by a shorter height compared to other children of their age. The ordinal logistic regression is one statistical method that used to analyze the ordinal response variable with three or more categories and predictor variables that are categorical or continuous. This method has been used in various fields, one of them is in the health field. In this study, ordinal logistic regression is used to describe the correlation between stunting status of children and birth weight, birth length, mother's age, mother's height, mother's knowledge about nutrition, parenting, health services, hygiene and sanitation, number of children and number of family members. The stunting status of children in this study is divided into three kind which are severely stunted, moderately stunted, and normal stunted. The aim of this study is to describe the characteristics of children and to determine the factors that influence the stunting status of children in priority villages for stunting prevention i.e. Bucurwetan and Alaspandan village in the Probolinggo district, East Java, Indonesia. The prevalence of stunting was 30% (22.5% moderately, 7.5% severely). The mother's height and health services were the important determinants of stunting.

Keywords: Children, Ordinal Logistic Regression, Stunting.

1. INTRODUCTION

Phenomenon of stunting is a major nutritional problem that occurs in Indonesia. Stunting will be one of the strategic issues that are priorities for health development in Indonesia for the next five years (2020-2024). Stunting is a kind of chronic nutritional problem in toddlers which is characterized by a shorter height than other children of their age.

In 2017, 22.2%, or around 150.8 million toddlers in the world were stunted. According to stunting prevalence data that collected by the World Health Organization (WHO), among the third countries, Indonesia is the highest prevalence in the South-East Asia Regional (SEAR). The average prevalence of stunting among under-five years children in Indonesia in 2005-2017 is 36.4% [1]. The results of Basic Health Research in 2018 showed that as many as 30.8% or around 7 million toddlers were stunted. This

number is still far from the minimum stunting rate set by WHO, which is 20%. In East Java Province, the prevalence of stunting among under-five years children is 32.81%. One of the areas in East Java that has a fairly high prevalence rate of stunting is Probolinggo Regency, which is 39.9% [2].

It is important to reduce the incidence of stunting in toddlers as early as possible to avoid long-term adverse effects such as stunted child development. Stunting can affect brain development so that the child's intelligence level is not optimal. This is at risk of reducing productivity as an adult. Stunting also makes children more susceptible to disease. Children who are stunted have a higher risk of suffering from chronic diseases in adulthood [1].

Several studies examining the factors that are at risk of influencing the incidence of stunting in toddlers have been carried out. Ni'mah and Nadhiroh

[3] using the Chi-Square and Fisher Exact tests and concluded that there is a correlation between the birth length, a history of exclusive breastfeeding, family income, mother’s education, and mother's knowledge about nutrition against the incidence of stunting among toddlers. Another factor related to the incidence of stunting is birth weight. Research in Nepal shows that babies with low birth weight have a higher risk of becoming stunted [4]. The birth length of the baby is also associated with the incidence of stunting. Research in Kendal shows that babies with short birth lengths have a high risk of stunting in toddlers [5]. This study aims to develop a risk model for the incidence of stunting among under-five years children using an ordinal logistic regression approach. If the form of the model is known, it is expected that the incidence of stunting among under-five years children can be anticipated from the start and preventive steps can be taken against the factors that have a significant effect.

2. LITERATURE REVIEW

2.1. Multicollinearity Detection

Multicollinearity is one of the problems in regression that occurs because of the high correlation between the predictor variables. To detect cases of multicollinearity, it can be done by calculating the VIF (Variance Inflation Factor) value which is defined as the following equation:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{1}$$

where R_j^2 is the coefficient of determination between one predictor variable X_j and another predictor variables. The predictor variable is said to have multicollinearity if the VIF value > 10.

2.2. Ordinal Logistic Regression

The ordinal logistic regression is one of statistical method that used for analyzing response variables which have an ordinal scales, it should consist of three or more categories. Predictor variables that could be included in the model are categorical or continuous data that consist of two or more variables [6]. Models that can be used for ordinal logistic regression are cumulative logit models. For example, Y_i is an ordinal scale response variable that has q categories and $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^T$ is a vector of the predictor variable in the i -th observation where $i = 1, 2, \dots, n$, then the cumulative logit model can be defined as follows:

$$\text{logit}[P(Y_i \leq j | \mathbf{x}_i)] = \theta_j + \mathbf{x}_i^T \boldsymbol{\beta} \tag{2}$$

where $j = 1, 2, \dots, q-1$, $P(Y_i \leq j | \mathbf{x}_i)$ is the cumulative probability of the j -th category response variable with terms \mathbf{x}_i , θ_j is the intercept parameter and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ as usual denotes the vector of the regression coefficients. The logit link function can be defined as follows:

$$\text{logit}[P(Y_i \leq j | \mathbf{x}_i)] = \ln \left[\frac{P(Y_i \leq j | \mathbf{x}_i)}{1 - P(Y_i \leq j | \mathbf{x}_i)} \right] \tag{3}$$

based on Equation (2) and Equation (3), the ordinal logistic regression model can be stated as follows:

$$\ln \left[\frac{P(Y_i \leq j | \mathbf{x}_i)}{1 - P(Y_i \leq j | \mathbf{x}_i)} \right] = \theta_j + \mathbf{x}_i^T \boldsymbol{\beta} \tag{4}$$

Therefore, from Equation (4) we can get

$$P(Y_i \leq j | \mathbf{x}_i) = \frac{\exp(\theta_j + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_j + \mathbf{x}_i^T \boldsymbol{\beta})} \tag{5}$$

If there are three response categories, then the cumulative probability of the j -th category response variable is as follows:

$$P(Y_i \leq 1 | \mathbf{x}_i) = \frac{\exp(\theta_1 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_1 + \mathbf{x}_i^T \boldsymbol{\beta})} \tag{6}$$

$$P(Y_i \leq 2 | \mathbf{x}_i) = \frac{\exp(\theta_2 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_2 + \mathbf{x}_i^T \boldsymbol{\beta})} \tag{7}$$

Suppose $\pi_j(\mathbf{x}_i) = P(Y_i = j | \mathbf{x}_i)$; $j = 1, 2, \dots, q$ is the probability that the response variable in the i -th observation has j -th category, then $P(Y_i \leq j | \mathbf{x}_i) = \pi_1(\mathbf{x}_i) + \pi_2(\mathbf{x}_i) + \dots + \pi_j(\mathbf{x}_i) = 1$, so that

$$\pi_1(\mathbf{x}_i) = \frac{\exp(\theta_1 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_1 + \mathbf{x}_i^T \boldsymbol{\beta})} \tag{8}$$

$$\pi_2(\mathbf{x}_i) = \frac{\exp(\theta_2 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_2 + \mathbf{x}_i^T \boldsymbol{\beta})} - \frac{\exp(\theta_1 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_1 + \mathbf{x}_i^T \boldsymbol{\beta})} \tag{9}$$

$$\pi_3(\mathbf{x}_i) = 1 - \frac{\exp(\theta_2 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_2 + \mathbf{x}_i^T \boldsymbol{\beta})} \tag{10}$$

The probability value for each response category is used as a guide for classification. An observation will be classified as the response variable of the j -th category based on the greatest probability value.

2.3. Estimating Parameters

Estimating parameters of ordinal logistic regression models can be obtained using the Maximum Likelihood Estimation (MLE) method. The first step in the MLE method is to make the likelihood function of the sample with n independent observations (Y_i, x_i) as follows:

$$L(\boldsymbol{\varphi}) = \prod_{i=1}^n \pi_1(\mathbf{x}_i)^{n_{1i}} \pi_2(\mathbf{x}_i)^{n_{2i}} \dots \pi_q(\mathbf{x}_i)^{n_{qi}}$$

$$= \prod_{i=1}^n \gamma_1(\mathbf{x}_i)^{n_{1i}} [\gamma_2(\mathbf{x}_i) - \gamma_1(\mathbf{x}_i)]^{n_{2i}} \dots [1 - \gamma_{q-1}(\mathbf{x}_i)]^{n_{qi}}$$

where n_{ji} will be 1 if $Y_i = j$ and will be 0 if $Y_i \neq j$, and $\boldsymbol{\varphi} = (\theta_1, \theta_2, \dots, \theta_{q-1}, \beta_1, \beta_2, \dots, \beta_p)^T$ is a parameter which is estimated by maximizing the likelihood function and $\gamma_j(\mathbf{x}_i) = P(Y_i \leq j | \mathbf{x}_i)$. Next, the \ln -transformation is performed as follows:

$$\ell = \ln(L(\boldsymbol{\varphi}))$$

$$= \sum_{i=1}^n \sum_{j=1}^q \left\{ n_{ji} \ln \left[\frac{\exp(\theta_j + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_j + \mathbf{x}_i^T \boldsymbol{\beta})} - \frac{\exp(\theta_{j-1} + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_{j-1} + \mathbf{x}_i^T \boldsymbol{\beta})} \right] \right\} \quad (11)$$

From the results of the derivation in Equation (11), a nonlinear function is obtained so that the Newton Raphson iteration method is used in solving as follows:

$$\boldsymbol{\varphi}^{(i+1)} = \boldsymbol{\varphi}^{(i)} - (\mathbf{H}^{(i)})^{-1} \mathbf{r}^{(i)} \quad (12)$$

where $\mathbf{r}^{(i)} = \frac{d\ell}{d\boldsymbol{\varphi}^{(i)}}; i = 0, 1, \dots$ and Hessian matrix

$\mathbf{H}^{(i)} = \frac{d^2\ell}{d\boldsymbol{\varphi} d\boldsymbol{\varphi}^T}$. The parameter vector estimate will be obtained when the process has converged, expressed as $\max |\boldsymbol{\varphi}^{(i+1)} - \boldsymbol{\varphi}^{(i)}| < \delta$, with δ is a small positive number.

2.4. Simultaneous Test

Simultaneous testing is carried out to determine the effect of predictor variables simultaneously on the response variable with the following hypothesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k; k = 1, 2, \dots, p$$

$$H_1 : \text{there is at least one } \beta_k \neq 0; k = 1, 2, \dots, p$$

The test statistic used to test the hypothesis is the likelihood ratio test which is defined as follows [7]:

$$G^2 = -2 \ln \left[\frac{\left(\frac{n_1}{n} \right)^{n_1} \left(\frac{n_2}{n} \right)^{n_2} \dots \left(\frac{n_q}{n} \right)^{n_q}}{\prod_{i=1}^n \pi_1(\mathbf{x}_i)^{n_{1i}} \pi_2(\mathbf{x}_i)^{n_{2i}} \dots \pi_q(\mathbf{x}_i)^{n_{qi}}} \right] \quad (13)$$

where $n_1 = \sum_{i=1}^n n_{1i}; n_2 = \sum_{i=1}^n n_{2i}; n_q = \sum_{i=1}^n n_{qi}$ and $n = n_1 + n_2 + \dots + n_q$. The area of rejection for H_0 is $G^2 > \chi^2_{(\alpha, p)}$ or p-value less than α .

2.5. Partial Test

This partial test is used to determine the effect of each predictor variable on the response variable with the following hypothesis:

$$H_0 : \beta_k = 0; k = 1, 2, \dots, p$$

$$H_1 : \beta_k \neq 0; k = 1, 2, \dots, p$$

The test statistic used to test this hypothesis is the Wald test statistic which is defined as follows [7]:

$$W_k = \frac{\beta_k}{s(\beta_k)}; k = 1, 2, \dots, p \quad (14)$$

The area of rejection for H_0 is $|W_k| > Z_{(\alpha/2)}$ or p-value less than α .

2.6. Goodness-of-Fit Test

The goodness-of-fit test is used to determine whether the model is appropriate or not, this test is done by comparing the actual model with the fit model with the following hypothesis:

$$H_0 : \text{Ordinal logistic regression models are appropriate}$$

$$H_1 : \text{Ordinal logistic regression model are not appropriate}$$

The test statistic used is the deviance test which is defined as follows:

$$D = 2 \sum_{i=1}^n \left\{ n_{1i} \ln \left(\frac{\pi_1(\mathbf{x}_i)}{\pi_1(\mathbf{x}_i)} \right) + n_{2i} \ln \left(\frac{\pi_2(\mathbf{x}_i)}{\pi_2(\mathbf{x}_i)} \right) + \dots + n_{qi} \ln \left(\frac{\pi_q(\mathbf{x}_i)}{\pi_q(\mathbf{x}_i)} \right) \right\} \quad (15)$$

The area of rejection for H_0 if $D > \chi_{\alpha, (q-1)(J-p-1)}^2$, where J is the number of level combinations of different predictor variables [7].

2.6. Classification Accuracy

Apparent Error Rate (APER) is a value used to see the probability of error in classifying objects. Suppose the response variable Y has q categories, then the APER value is

$$APER = \frac{\sum_{i \neq j=1}^q n_{ij}}{\sum_{i,j=1}^q n_{ij}} \times 100\% \tag{16}$$

where n_{ij} is the number of Y events in category i from the observation results classified into Y events in category j from the prediction results [8]. Based on the APER value in Equation (16), classification accuracy is obtained $100\% - APER$.

3. RESEARCH METHODOLOGY

3.1. Data Source

The data used in this study are sample data of toddlers recorded in the One Posyandu Application and survey sampling in priority villages for stunting prevention i.e. Bucorwetan and Alaspandan village in Probolinggo district, East Java, Indonesia. The sample used was 80 toddlers.

3.2. Research Variable

The response variable (Y) used in this study is the nutritional status of toddlers based on the height/age index which is categorized into 3, i.e. severely stunted ($Y=1$), moderately stunted ($Y=2$) and normal ($Y=3$). Meanwhile, the predictor variable (X) consists of 11 continuous variables, namely birth weight (X_1), birth length (X_2), parity (X_3), mother's age (X_4), mother's height (X_5), mother's knowledge of nutrition (X_6), parenting (X_7), health services (X_8), hygiene and sanitation (X_9), number of children (X_{10}) and number of family members (X_{11}).

3.3. Methods

The steps of data analysis in this research are as follows:

- Descriptive analysis to determine the characteristics of each variable

- Detection of multicollinearity
- Ordinal logistic regression analysis
 - Estimated parameters using the MLE method
 - Test the significance of parameters simultaneously and partially
 - Ordinal logistic regression model building
 - Interpretation of factors affecting the nutritional status of toddlers based on Height/Age index
- Goodness-of-Fit Test using the deviance test statistic
- Calculating the classification accuracy value of the model formed

4. RESULTS AND DISCUSSION

4.1. Characteristics of Research Variables

Analysis of the characteristics of the variables in this study describes the condition of toddlers in the sample. The first characteristic analysis was carried out on the response variable, namely the nutritional status of toddlers based on height/age index. The results of the analysis can be shown in the following Figure 1:

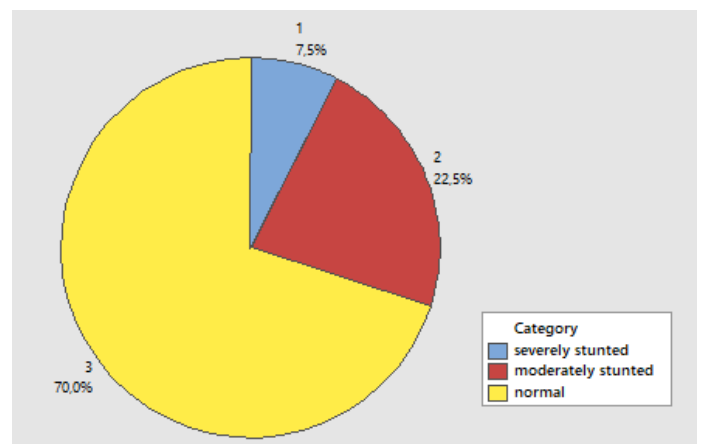


Figure 1. Percentage of nutritional status of toddlers based on height/age index

Based on Figure 1, it can be seen that the prevalence of stunting cases in the sample is 30% with moderately stunted is 22.5% and severely stunted is 7.5%. Furthermore, the characteristic analysis of the predictor variables is as follows:

Table 1. Descriptive statistics on predictor variables

Variable	Mean	Median	Minimum	Maximum	Standard Deviation
X_1	2.978	3	1.820	4.8	0.468
X_2	49.7	49	37	57	2.523
X_3	1.789	2	1	4	0.741
X_4	29.913	30	20	43	5.408
X_5	149.959	150	139	165	4.846
X_6	63.25	70	30	100	15.326
X_7	74.5	80	20	100	21.103
X_8	84.688	100	25	100	18.007
X_9	62	60	0	100	20.277
X_{10}	1.825	2	1	4	0.759
X_{11}	4.813	5	2	9	1.274

Based on Table 1, it can be seen that the variable X_1 (birth weight) has a mean of 2.978 kg but in the minimum value there is still a birth weight that is less than 2 kg which is 1.82 kg. This shows that there are still toddlers who are born prematurely. Furthermore, the variable X_4 (mother’s age) has a the median value of 30 years but in the minimum and maximum ages of 20 years and 43 years it is a period beyond the age of the mother to produce healthy babies. So there is a possibility that toddlers are born less healthy. In other variables, the variable X_7 (parenting) has a median value of 80 and a minimum value of 20. This proves that the implementation of good parenting is not optimal for toddlers. Meanwhile, the variable X_9 (hygiene and sanitation) has a median value of 60 but it has the minimum value of 0. This indicates that there is a lack of hygiene and sanitation in toddlers.

4.2. Multicollinearity Detection

Multicollinearity detection is performed to determine whether there is a linear relationship between predictor variables in the model. Multicollinearity detection correlates between predictor variables, where the estimation of multicollinearity cases occurs if there is a high correlation between the predictor variables. The existence of multicollinearity cases can also be seen based on the VIF value with a limit of 10. If the VIF value of a variable is more than 10 then multicollinearity occurs. The following table shows the multicollinearity detection of the variables in this study:

Table 2. Multicollinearity test of predictor variables

Variable	VIF
Birth weight	2.516
Birth length	2.421
Parity	18.773
Mother’s age	1.869
Mother’s height	1.190
Mother's knowledge of nutrition	1.204
Parenting	1.183
Health services	1.109
Hygiene and sanitation	1.303
Number of children	17.241
Number of family members	1.120

Based on Table 2, it was found that the variable X_3 (parity) and X_{10} (number of children) had a VIF value more than 10 so that they had multicollinearity cases and were excluded from the model.

4.3. Ordinal Logistic Regression

Ordinal logistic regression modeling was done by including all predictor variables that did not have cases of multicollinearity. To perform ordinal logistic regression testing, a simultaneous test was conducted first. The results of the simultaneous test for these variables can be seen in the following table:

Table 3. Concurrent Test with Likelihood Ratio

Model	G^2	Chi-Square	df	Sig.
Intercept Only	124.730			
Final	101.282	23.448	9	0.005

Based on Table 3, shows that the value of G^2 is 23.448, which means that $G^2 > \chi^2_{(0.05;9)} = 16.919$ or the significance value of 0.005 is less than the alpha value of 5% so that the decision taken is to reject H_0

which means that at least one parameter coefficient β is significant to the ordinal logistic regression model. The next analysis is a partial coefficient test. The partial test results can be seen in the following Table 4:

Table 4. Partial Parameter Estimation

Variable	Estimate	Std. Error	Wald	Sig.	95% Confidence Interval	
					Lower	Upper
[Y = 1]	33.038	12.069	7.494	0.006	9.384	56.693
[Y = 2]	35.198	12.167	8.369	0.004	11.351	59.044
β_1	0.131	0.977	0.018	0.893	-1.783	2.045
β_2	0.221	0.166	1.772	0.183	-0.104	0.545
β_4	-0.037	0.055	0.455	0.500	-0.144	0.070
β_5	0.174	0.062	7.788	0.005*	0.052	0.297
β_6	-0.033	0.020	2.614	0.106	-0.072	0.007
β_7	0.015	0.014	1.184	0.277	-0.012	0.043
β_8	0.032	0.015	4.505	0.034*	0.002	0.062
β_9	-0.018	0.015	1.407	0.236	-0.046	0.011
β_{11}	-0.149	0.217	0.473	0.492	-0.575	0.276

* significant at alpha = 0.05

Based on Table 4, it can be seen that two variables have a significant value less than the alpha value of 5%, i.e 0.005 and 0.034. They are the variables X_5 (mother's height) and X_8 (health services). The height of the parents is related to the physical growth of the child. Genetic factors in mothers, namely height affect the

incidence of stunting in toddlers. Likewise regarding health services for toddlers will affect the fulfillment of toddler nutrition.

Thus an ordinal logistic regression model can be built based on variables of mother's height and health services. The results can be seen in the following table:

Table 5. Parameter Estimates for Logit Model

Variable	Estimate	Std. Error	Wald	Sig.	95% Confidence Interval	
					Lower	Upper
[Y = 1]	25.797	8.613	8.971	0.003	8.916	42.678
[Y = 2]	27.722	8.697	10.160	0.001	10.675	44.768
β_5	0.174	0.057	9.357	0.002	0.062	0.285
β_8	0.032	0.014	5.204	0.023	0.005	0.060

* significant at alpha = 0.05

The logit model is structured based on significant parameters in Table 5 with the following results:

$$\text{logit} \left[\hat{P}(Y_i \leq 1 | x_i) \right] = 25.797 + 0.174X_5 + 0.032X_8$$

$$\text{logit} \left[\hat{P}(Y_i \leq 2 | x_i) \right] = 27.722 + 0.174X_5 + 0.032X_8$$

4.4. Model Fit Test

This test aims to determine the suitability of the regression model. The model is considered suitable if the p-value at Deviance is more than 5%. The results of

the model suitability test can be seen in the following table:

Table 6. Model Suitability Test

	Chi-Square	df	Sig.
Pearson	140.625	82	0.000
Deviance	83.029	82	0.447

Based on Table 6, it can be seen that the p-value at Deviance of 0.447 indicates more than 5% so that a decision to accept H_0 can be taken and it can be concluded that the model obtained is appropriate.

4.5. Accuracy Of Model Classification

The accuracy of model classification is done by calculating the value of the classification accuracy between the actual value and the predicted value obtained from the model that has been formed.

Table 7. The accuracy of model classification

Observation	Predicted Response Category		
	Severely stunted	Moderately stunted	Normal
Severely stunted	0	2	4
Moderately stunted	0	3	15
Normal	0	3	53

Based on Table 7, it is found that the model classification accuracy value is 70%.

5. CONCLUSION

Analysis of the characteristics of the variables in this study describes that the prevalence of stunting cases is 30% with moderately stunted is 22.5% and severely stunted is 7.5%, there are still toddlers who are born prematurely, mothers have period beyond the age of the mother to produce healthy babies and the implementation of good parenting is not optimal for toddlers. Based on the ordinal logistic regression model, two variables are significant determinants of stunting among under-five years children, i.e mother's height and health services. The model classification accuracy value is 70%.

ACKNOWLEDGMENTS

The authors specially appreciate to Universitas Airlangga for funding this research with a contract number 824/UN3.14/PT/2020. We also thank to editors and peer reviewers who have given useful suggestions and criticisms for the improvement of this paper.

REFERENCES

- [1] Pusat Data dan Informasi Kementerian Kesehatan RI, Situasi Balita Pendek (Stunting) di Indonesia, Buletin Jendela Data dan Informasi Kesehatan, 2018.
- [2] Kementerian Kesehatan RI, Hasil Utama Riskesdas 2018 Provinsi Jawa Timur, 2018.
- [3] Ni'mah, K. and Nadhiroh, S. R., Faktor yang Berhubungan dengan Kejadian Stunting pada Balita, *Media Gizi Indonesia*, vol. 10, 2015, pp. 13-19.
- [4] Paudel, R., Pradhan, B., Wagle, R. R., Pahari, D. P. and Onta, S. R., Risk Factors for Stunting Among Children: A Community based Case Control Study in Nepal, *Kathmandu University Medical Journal*, vol. 10, 2012, pp. 18-24. DOI: 10.3126/kumj.v10i3.8012
- [5] Meilyasari, F. and Isnawati, M., Faktor Risiko Kejadian Stunting pada Balita Usia 12 Bulan di Desa Purwokerto Kecamatan Patebon Kabupaten Kendal, *Journal of Nutrition College*, vol. 3, 2014, pp.16-25. DOI: <https://doi.org/10.14710/jnc.v3i2.5437>
- [6] Agresti, A., *Categorical Data Analysis*, John Wiley and Sons Inc, New Jersey, 2002.
- [7] Hosmer, D. W., and Lemeshow, S., *Applied Logistic Regression* 2nd ed, John Wiley and Sons Inc, New York, 2000.
- [8] Johnson, R. A., and Wichern, D. W., *Applied Multivariate Statistical Analysis* 6th ed, Pearson Education Inc, United States of America, 2007.