BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1541–1556
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1541-1556

COMPARISON OF MARS AND BINARY LOGISTIC REGRESSION MODELS FOR IDENTIFYING STUNTING RISK FACTORS IN TODDLERS IN TELUK WARU, EAST SERAM REGENCY

Johan Bruiyf Bension[1], Ferry Kondo Lembang[2*], Nurul Fadillah Idris[3], Norisca Lewaherilla[4]

[1]*Medical School Study Program, Faculty of Medicine, Universitas Pattimura*
[2,3,4]*Statistics Study Program, Faculty of Science and Technology, Universitas Pattimura*
Jln. Ir. M. Putuhena, Kampus Unpatti Poka, Ambon, 97233, Indonesia

*Corresponding author's e-mail: * ferrykondolembang@gmail.com*

Article Info

Article History:
Received: 18th June 2025
Revised: 13th August 2025
Accepted: 24th September 2025
Available Online: 26th January 2026

Keywords:
Binary Logistic Regression;
MARS;
Prevalence;
Stunting.

ABSTRACT

In 2022, the prevalence of chronic stunting in Indonesia reached 21.6%, surpassing the World Health Organization (WHO) threshold of 20%. East Seram Regency reported an even higher prevalence of 24.1%, with Teluk Waru District identified as one of the areas most affected due to low compliance with healthy lifestyle practices. This study aimed to compare the performance of Multivariate Adaptive Regression Splines (MARS) and Binary Logistic Regression in analyzing risk factors for toddler stunting in Teluk Waru District, East Seram Regency. Data were collected through direct anthropometric measurements at the Integrated Health Post (Posyandu) of Teluk Waru Health Center with 60 respondents. The findings revealed that Binary Logistic Regression outperformed MARS, achieving $R^2 = 72.7\%$ accuracy in predicting stunting. Significant determinants of toddler stunting included a history of illness, provision of supplementary food for pregnant women, and iron tablet consumption during pregnancy. The novelty of this study lies in the application of a comparative modeling approach—MARS versus Binary Logistic Regression—in identifying stunting risk factors at a district level with high prevalence. Practically, the results can assist local health authorities in prioritizing maternal nutrition and disease prevention programs to reduce stunting.

How to cite this article:

J. B. Bension, F. K. Lembang, N. F. Idris andN. Lewaherilla".,, "COMPARISON OF MARS AND BINARY LOGISTIC REGRESSION MODELS FOR IDENTIFYING STUNTING RISK FACTORS IN TODDLERS IN TELUK WARU, EAST SERAM REGENCY *BAREKENG: J. Math. & App.,* vol. 20, no. 2, pp. 1541-1556, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Toddlers are children aged 0-59 months who experience rapid growth and development, requiring many high-quality nutrients [1]. Toddlers who experience nutritional disorders that inhibit the achievement of maximum genetic potential in terms of growth and development are called stunted toddlers. According to the Ministry of Health of the Republic of Indonesia, toddlers are considered to have severe stunting if they have a z-score of less than minus three standard deviations [2]. Stunting is a chronic problem related to nutrition, which, upon closer examination, is often found to be more frequently caused by health and social issues. Stunting is a serious problem affecting millions of toddlers worldwide, particularly in developing countries. Stunting has adverse consequences, including impaired immune function and an elevated risk of chronic diseases in early childhood.

East Seram Regency is one of several regencies in Maluku Province where cases of stunted growth in toddlers are still prevalent. In early 2023, two villages in Teluk Waru District, East Seram Regency, were designated as national pilot project villages for preventing stunting and malnutrition, namely Madak Village and Kampung Baru Village. According to the results of the Indonesian Nutritional Status Survey (SSGI) at the end of 2022, the prevalence of stunting in the East Seram Regency reached 24.1%, which is still above the global target set by the WHO, which is below 20%. Therefore, because the impact of stunting significantly affects the health of toddlers in the future, as they are susceptible to disease, early intervention or prevention efforts are needed from the beginning of pregnancy until birth [3]. By applying regression methods, one form of stunting intervention involves identifying the primary determinants of stunting.

The regression method is a statistical analysis tool that finds the relationship between predictor and response variables [4]. Stunting incidents can be classified into two groups: those that did not occur and those that did. The appropriate regression model used for cases of stunting incidents is the regression method, with the dichotomous response variable [5]. Two regression method approaches that can be used are parametric and non-parametric regression [6]. The regression methods generally used to determine the effect of predictor variables on dichotomous response variables are the Multivariate Adaptive Regression Spline (MARS) method and Binary Logistic Regression [7]. The MARS method and Binary Logistic Regression have been widely applied in various fields, including health science. The second method can be utilized in developing predictive models for disease diagnosis, clinical outcome prediction, or estimating individual health risks based on complex factors such as medical history or genetic predisposition, with stunting serving as one example.

Several previous studies related to stunting, such as research conducted by Elisa [8], have shown that factors influencing the incidence of stunted toddlers include birth weight, infectious disease status, and access to healthy toilets. Another study by Utami showed that the factors influencing stunted toddlers in East Java were gender, exclusive breastfeeding, complete basic immunization, deworming, birth weight, birth height, pregnant women receiving additional food, and pregnant women receiving iron tablets, with a classification accuracy of 66.7% [9]. Multivariate Adaptive Regression Splines (MARS) have been applied in several international health and biomedical studies — for example, to predict improvements in HbA1c in diabetic patients [10], in QSAR modelling for antitumor research [11], and to estimate biological age in postmenopausal women [12].

This research aims to address this gap by comparing the performance of Multivariate Adaptive Regression Splines (MARS) and Binary Logistic Regression in identifying risk factors for stunting among toddlers in the Teluk Waru Health Center area. The evaluation criteria include the Apparent Error Rate (APER), Total Accuracy Rate (TAR), and Coefficient of Determination (R2). The novelty of this study lies in applying and comparing advanced inferential statistical models to stunting analysis at the district level, an approach rarely used in Maluku Province, where research has predominantly employed descriptive methods. By doing so, this study not only provides methodological contributions but also practical insights to support government efforts in accelerating stunting reduction in East Seram Regency, Maluku Province, and Indonesia more broadly.

2. RESEARCH METHODS

2.1 Data Sources

The data sources used in this study are primary data collected through field observations, utilizing direct measurement and interview methods. The study included 60 toddlers whose body length and height were measured using anthropometric tools during the implementation of integrated health posts in August 2023 at the Teluk Waru Health Center, East Seram Regency. Logistic Regression remains widely used in health research with small datasets, as it provides robust and interpretable estimates for binary outcomes when the number of predictors is limited. MARS, on the other hand, is a flexible nonparametric technique capable of capturing nonlinearities and interactions without strict distributional assumptions. While larger samples generally maximize the performance of MARS, its application in smaller datasets remains meaningful, particularly when supported by cross-validation to mitigate overfitting risks. In this study, the goal is not to produce a nationwide generalization but to evaluate the relative performance of parametric and nonparametric approaches in a high-prevalence district with real-world field constraints. Thus, using both Logistic Regression and MARS provides methodological insights as well as practical evidence for local health interventions.

2.2 Research Variables

Research variables are characteristics or attributes of individuals or organizations that can be measured or observed, and they have certain variations determined by the researcher [13]. This study's research variables consist of response and predictor variables. The response variable (Y) used in this study is the stunting status of toddlers, and as many as nine predictor variables (X) are presented in Table 1.

Table 1. Research Variable

Variable	Information	Operational Definition	Category	Scale
Y	Stunting Status	Growth conditions in toddlers	0 = No stunting 1 = Stunting	Nominal
X_1	Having illness history	Medical history of the toodler	0 = No 1 = Yes	Nominal
X_2	Age of complementary feeding introduction for baby	Toddler's age at the introduvtion of complementary feeding	0 = < 6 months 1 = 6 months 2 = > 6 months	Ordinal
X_3	Formula drink consumption	Mother's acknowledgment of providing formula milk as a supplement to breastfeeding	0 = No 1 = Yes	Nominal
X_4	Exclusive breastfeeding	Exclusive breastfeeding for babies up to 6 months of age	0 = No 1 = Yes	Nominal
X_5	Regularly measure the child's height	Activities conducted to monitor toddler growth	0 = No 1 = Yes	Nominal
X_6	Regularly measure the child's weight	Activities conducted to monitor toddler weight growth	0 = No 1 = Yes	Nominal
X_7	Pregnant women receive supplementary food provision	Maternal acknowledgment during pregnancy of receiving supplementary feeding	0 = No 1 = Yes	Nominal
X_8	Mother receives iron supplementation during pregnancy	Maternal acknowledgment during pregnancy of receiving iron supplement tablet	0 = No 1 = Yes	Nominal

2.3 MARS Model

The MARS model is a development of the Recursive Partitioning Regression (RPR) approach, combining of spline methods. RPR is a computational program for processing high-dimensional data, and the spline method is a polynomial cut with continuous derivatives at the knot. MARS aims to overcome the problem of high-dimensional data with predictor variable sizes between $3 \leq v \leq 20$ and sample sizes of $50 \leq n \leq 1000$ and to improve the weaknesses of RPR, namely, the model produced is continuous at the knot [14]. In the MARS method, there are several things to consider: knots and basis functions [15]. A knot is a point that separates the end of a data area from the beginning of another location. Knots in MARS are determined using forward stepwise and backward stepwise. The placement of knots depends on the determination of the number of observations for each knot. The number of observations at each knot is known as the Minimum Observation (MO). The MO used is 0,1,2 and 3. Basis Function (BF) is a function separated by knot points explaining the relationship between the predictor and response variables. The maximum BF allowed is 2 to 4 times the number of predictor variables [16], while the maximum number of interactions (MI) is 1, 2, and 3. If there are more than three interactions, it will cause a very complex model interpretation. The MARS model estimator is as follows:

$$\hat{f}(x) = \hat{a}_0 + \sum_{m=1}^{M} \hat{a}_m \prod_{k=1}^{K_m} \left[S_{km} \cdot \left(x_{j(k,m)} - t_{km}\right)\right]_+, \quad (1)$$

where

$\hat{a}_0$	:	constant coefficient of the basis function;
$\hat{a}_m$	:	coefficient of the m −th basis function;
M	:	maximum of the basis function;
K_m	:	degree of interaction in the m-th basis function;
S_{km}	:	sign (+ or –) for the k −th interaction in the m −th basis function;
$x_{j(k,m)}$	:	j-th predictor variable, k-th interaction, and m −th basis function;
t_{km}	:	knot value of predictor variable $x_{j(k,m)}$.

According to [10], the model estimation uses the Maximum Likelihood Estimation (MLE) method for the MARS model with binary response variables.

2.4 MARS Model Significance Testing

If the residuals in the MARS model meet the assumptions in parametric regression, then parameter significance testing is carried out, and the model's suitability is evaluated. Testing is carried out by simultaneously or partially testing the regression coefficients.

2.4.1 Simultaneous Regression Coefficient Testing

Simultaneous regression coefficient testing is done by simultaneously testing the parameters contained in the MARS model. This test aims to determine whether the MARS model is generally appropriate. The hypothesis used is as follows:

$H_0: a_1 = a_2 = \cdots = a_m = 0$ (model not significant);

H_1: there is at least one $a_m \neq 0$; $m = 1,2, \cdots M$ (model significant).

Test statistic:

$$F_{count} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / M}{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / N - M - 1}. \quad (2)$$

The critical region rejects H_0 if the $F_{count} > F_{\alpha(M;N-M-1)}$ or $p - value < \alpha$.

2.4.2 Partial Regression Coefficient Testing

Suppose in the simultaneous regression coefficient test, it is concluded that there is at least one significant parameter. In that case, it is necessary to know the essential and insignificant parameters. We perform a partial regression coefficient test using the following hypotheses:

$H_0: a_m = 0$ (coefficient a_m does not affect the model).

$H_1: a_m \neq 0$ for each m, where $m = 1,2, \cdots M$ (coefficient a_m influence on the model).

Test statistics:

$$t_{count} = \frac{\hat{a}_m}{Se(\hat{a}_m)}, \quad (3)$$

with $Se(\hat{a}_m) = \sqrt{var(\hat{a}_m)}$.

Critical area rejects H_0 if $t_{count} > t_{(\frac{\alpha}{2};N-M)}$ or $p - value < \alpha$.

2.5 Criteria for Selecting the Best MARS Model

The criteria for selecting the best MARS model are generally based on the minimum Generalized Cross-Validation (GCV) method [17]. In general, GCV is defined as follows:

$$GCV(M) = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{f}_M(x_i)\right]^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]}. \quad (4)$$

2.6 Independence Test

The Chi-Square test is used to determine whether there is a relationship between two categorical variables. The hypothesis used is as follows:

H_0: There is no relationship between the predictor variable and the response variable;

H_1: There is a relationship between the predictor variable and the response variable.

Test statistics:

$$X^2_{count} = \sum_{i,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}, \quad (5)$$

where:

O_{ij} : observation value/observation row i, column j;

E_{ij} : expected value of row I, column j.

Critical area rejects H_0 if $X^2_{count} > X_{(\alpha)(a-1)(b-1)}$.

2.7 Regresi Logistik Biner

Binary logistic regression is a data analysis method used to find the relationship between response variables (y), which are binary or dichotomous, and predictor variables (x), which are polychotomous [18]. We can express the logistic regression function by the following formula:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}, \quad (6)$$

where p: number of predictor variables x_i.

2.8 Model Fit Test

The model suitability test functions to determine whether or not there is a difference in the results of observations and model predictions with the following hypothesis:

H_0: Model is appropriate, or is there no difference in the results of observations and predictions;

H_1: The model is not appropriate, or there is a difference in the results of observations and predictions.

Test statistics:

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n_k\bar{\pi}_k)^2}{n_k\bar{\pi}_k(1 - \bar{\pi}_k)}. \quad (7)$$

Failed to reject H_0 if $\hat{C} < X^2_{(\alpha,g-2)}$ or $p - value > \alpha$ and it can be concluded that there is no difference between observations and model predictions [19].

2.9 Significance Testing of Binary Logistic Regression Model

Simultaneous significance testing of binary logistic regression model parameters aims to determine whether predictor variables have an overall effect or not on the model with the following hypothesis [20]:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_j = 0$;

H_1: at least one $\beta_j \neq 0$, where $j = 1,2, \cdots p$.

Test statistics:

$$G = -2\ ln\left[\frac{\left(\frac{n_0}{n}\right)^{n_0}\left(\frac{n_1}{n}\right)^{n_1}}{\prod_{i=1}^{n}(\hat{\pi}_i)^{y_i}(1-\hat{\pi}_i)^{1-y_i}}\right]. \tag{8}$$

Critical area rejects H_0 if $G > X^2_{(df,\alpha)}$ or $p - value < \alpha$.

Next, a partial significance test is carried out to find out which predictor variables have a significant influence on the response variable with the following hypothesis:

$H_0: \beta_j = 0$;

$H_1: \beta_j \neq 0$, where $j = 1,2, \cdots p$.

Test statistics:

$$W^2 = \left[\frac{\hat{\beta}_J}{SE(\hat{\beta}_J)}\right]^2 \tag{9}$$

Critical area rejects H_0 if $W^2 > X^2_{(df,\alpha)}$ or $p - value < \alpha$.

2.10 Classification Accuracy

Classification accuracy is needed to determine the grouping of data that is classified correctly in its group. Apparent Error Rate (APER) is defined as the proportion of incorrectly classified samples. The proportion of correctly classified samples can be calculated from the Total Accuracy Rate (TAR) value [21].

2.11 Coefficient of Determination (R^2)

The coefficient of determination (R^2) is a fundamental measure used to assess the explanatory power of a regression model. It quantifies the proportion of variance in the dependent variable that is explained by the set of independent variables. An R^2 value approaching 1 indicates that the model accounts for a large proportion of the observed variability, while a value near 0 suggests limited explanatory capacity. Although widely applied as an indicator of model performance, R^2 should be interpreted with caution, as a high value does not imply causality and may be affected by overfitting or multicollinearity.

3. RESULT AND DISCUSSION

3.1 Descriptive Analysis

Descriptive statistical analysis generally explains the characteristics of the data, specifically determining the stunting status of toddlers in Teluk Waru District, SBT Regency, which is measured using anthropometric tools, including body length and height, during the integrated health post in August 2023. The results of the measurements will determine whether toddlers tend to be categorized as stunted or not. Stunting can be visually shown as follows:

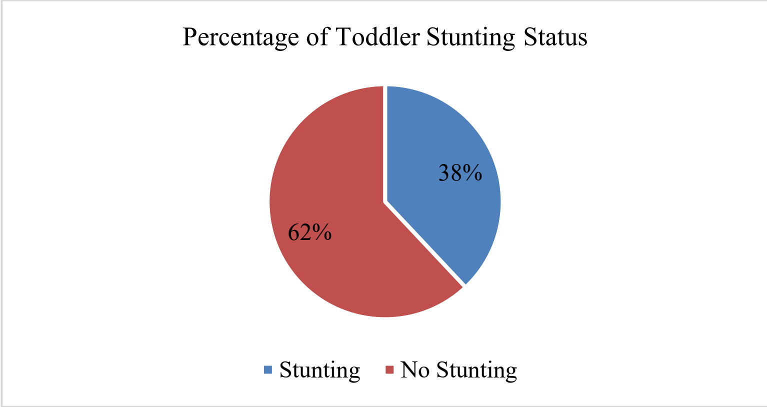

Figure 1. Toddler Stunting Status Bar Chart

Based on the information in Figure 1, it can be observed that, among the 60 respondents measured, the results indicate that 47 toddlers, or 62%, were in the non-stunting category, while 13 toddlers, or 38%, were in the stunting category.

3.2 Independence Testing

Before the modeling stage is carried out, it is necessary to analyze the relationship between the response variable and the predictor variable through independence testing using the Chi-Square Test. The results of the test are shown as follows:

Table 2. Chi-Square Test

Variable	X^2	df	X^2_{table}	Decision	Information
Medical History (X_1)	7.183	1	2.706	Reject H_0	There is a Relationship
Age of complementary feeding introduction for baby (X_2)	1.061	2	3.065	Failed to reject H_0	There is no Relationship
Consuming Formula Milk (X_3)	1.910	1	2.706	Failed to reject H_0	There is no Relationship
Given Exclusive Breastfeeding (X_4)	0.060	1	2.706	Failed to reject H_0	There is no Relationship
Regularly Measuring Child's Height (X_5)	17.897	1	2.706	Reject H_0	There is a Relationship
Regularly Measuring Child's Weight (X_6)	15.545	1	2.706	Reject H_0	There is a Relationship
Pregnant women receive supplementary food provision (X_7)	10.523	1	2.706	Reject H_0	There is a Relationship
Mother receives iron supplementation during pregnancy (X_8)	26.643	1	2.706	Reject H_0	There is a Relationship

Table 2 shows that the age variable of giving complementary feeding introduction for the baby, consuming formula milk, and being given exclusive breastfeeding is insignificant because the value of $X^2 < X^2_{table}$ fails to reject H_0. While for the variables of having a history of illness, routinely measuring height, routinely measuring weight, pregnant women receiving supplementary food provision, and mothers receiving iron tablets during pregnancy are significant because the value of $X^2 > X^2_{table}$, so that H_0 is rejected. Therefore, it can be concluded that all independent variables have a relationship with the dependent variable except for the age variable of giving MPASI, consuming formula milk, and being given exclusive breastfeeding. Since three variables are not significant in forming the model, the insignificant variables are excluded from the model.

3.3 MARS Model

The MARS model is explained in section 2.3, where the formation process can be done by trial and error for all BF, MI, and MO combinations. The number of BF used is 2 to 4 times the number of predictor variables. At the modeling stage, the number of predictor variables included is five, so the number of BF used is 10, 15, and 20. The MI used is 1, 2, and 3. While the MO used is 0, 1, 2, and 3. The best model selection can be seen from the minimum Generalized Cross Validation (GCV) value as shown below.

Table 3. Trial and Error Formation of the MARS Model

BF	MI	MO	GCV
10	1	0	0.106
		1	0.096
		2	0.106
		3	0.106
10	**2**	0	0.105
		1	**0.087**
		2	0.102
		3	0.105
10	3	0	0.105
		1	0.088
		2	0.102
		3	0.105
15	1	0	0.116
		1	0.105
		2	0.116
		3	0.116
15	2	0	0.120
		1	0.088
		2	0.111
		3	0.111
15	3	0	0.120
		1	0.088
		2	0.105
		3	0105
20	1	0	0.121
		1	0.112
		2	0.121
		3	0.121
20	2	0	0.138
		1	0.088
		2	0.125
		3	0.125
20	3	0	0.138
		1	0.088
		2	0.116
		3	0.116

Based on Table 3, it is evident that 36 models are formed for each combination. Of the 36 models, the model with a minimum GCV value of 0.087 combines 10 BFs, 2 MIs, and 1 MO.

3.3.1 The Best MARS Model

The best MARS model is the model generated from the minimum GCV value. Therefore, the model generated by BF, MI, and MO is the most effective. The model from the combination of BF = 10, MI = 2, and MO = 1 by using (1) is as follows:

$$\hat{f}(x) = 0.344 + 0.468BF_3 + 0.474BF_5 - 0.353BF_9,$$

with $BF_1 = (X_8 = 0)$, $BF_2 = (X_8 = 1)$, $BF_3 = (X_5 = 0) * BF_2$, $BF_5 = (X_7 = 0) * BF_1$, and $BF_9 = (X_1 = 0) * BF_2$.

From the best model produced, it can be concluded that the variables of Medical History (X_1), Routine Measurement of Child's Height (X_5), Pregnant Women Receiving Supplementary Food Provision (X_7), and Mothers Receiving TTD During Pregnancy (X_8) affect the stunting status of toddlers.

3.3.2 The Best MARS Significant Test

This test is conducted to see the significance of the parameters and evaluate the model's suitability by checking the regression coefficients simultaneously and partially. The tests consist of simultaneous and partial.

1. Simultaneous Test

   A simultaneous test is conducted using the F-test to determine the effect of predictor variables on the response variable simultaneously, based on the basis function coefficient. Decision: Based on the results of data processing using the MARS 2.0 software, the $F_{count} = 40.557$, while $F_{(0.05;3;56)} = 2.769$. Because the $F_{count} > F_{(0.05;3;56)}$ then reject H_0. It can be concluded that the predictor variables collectively have a significant effect on the response variable, making them suitable for use in modeling stunting status.

2. Partial Test

   After conducting a simultaneous test that shows the predictor variables collectively affect the response variable, this partial test will examine which predictor variables individually affect the response variable. The test statistic used is the t-test.

   The decision: t_{count} states the calculated value of the t-test and $t_{(0.025;56)} = 2.003$ states the table values of the t-test statistics at the significance level used, with degrees of freedom of 56. Based on the test statistics used, it is known that all BF parameters have a value of $|t_{count}| > t_{(0.025;56)}$, and then reject H_0. This result indicates that each BF parameter representing the predictor variable has a partial and significant effect on the response variable.

3.3.3 Classification

This section presents the classification accuracy of the MARS method for both the observation and prediction phases. The results summarize the performance of MARS in accurately classifying the data under each status.

Table 4. MARS Classification Accuracy

Observation		Predicting Status	
		No Stunting	**Stunting**
Status	**No Stunting**	43	4
	Stunting	1	12

According to Table 4, we calculate the accuracy as follow:

$$APER(\%) = \frac{n_{10}+n_{01}}{n} \times 100\% = \frac{1+4}{60} \times 100\% = 8.33\%,$$

$$TAR(\%) = 1 - APER = 1 - 8.33\% = 91.67\%.$$

The TAR value of 91.67% indicates that the MARS model obtained was able to predict accurately with a 91.67% accuracy.

3.4 Binary Logistic Regression Model

The binary logistic regression model is explained in Section 2.7, where it is formed from predictor variables of medical history, routine height measurement, routine weight measurement, pregnant women receiving supplementary food provision, and mothers receiving iron tablets (TTD) during pregnancy, with the response variable of toddler stunting status. The initial step in this modeling is to estimate the parameters.

3.4.1 Parameter Estimation

Parameter estimation calculations are required to obtain the initial model. The resulting parameter estimation values are as follows.

Table 5. Initial Model Parameter Estimation

Predictor Variable	$\widehat{\beta_j}$
Constants β_0	-0.989
Medical History $X_{1(1)}$	-5.642
Routinely Measure Children's Height $X_{5(1)}$	19.841
Routinely Measure Children's Weight $X_{6(1)}$	-17.791
Pregnant women receive supplementary food provision $X_{7(1)}$	3.219
Mother receives iron supplementation during pregnancy $X_{8(1)}$	4.301

By looking at Table 5 above, the initial model formed according to Eq. (6) is as follows:

$$\pi(x) = \frac{e^{-0.989-5.642+19.841-17.791+3.219+4.301}}{1+e^{-0.989-5.642+19.841-17.791+3.219+4.301}}.$$

3.4.2 Significant Parameter Test

After obtaining parameter estimates, the next step is to conduct parameter significance tests simultaneously and partially.

1. Likelihood Ratio Test

   Hypothesis:

   $H_0: \beta_1 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0;$

   H_1: there is at least one $\beta_j \neq 0$, where $j = 1,5,6,7,8$.

   Significant Level: $\alpha = 5\%$.

   Test statistic:

   $$G = -2\ ln\left[\frac{\left(\frac{n_0}{n}\right)^{n_0}\left(\frac{n_1}{n}\right)^{n_1}}{\prod_{i=1}^{n}(\hat{\pi}_i)^{y_i}(1-\hat{\pi}_i)^{1-y_i}}\right] = 41.386.$$

   Critical area: Reject H_0 if $G > X^2_{(df,\alpha)}$.

   Decision: Because $G = 41.386 > X^2_{(5;0,05)} = 11.070$, therefore, there is at least one predictor variable that influences the response variable, or simultaneously, the predictor variables influence the stunting status of toddlers.

2. Wald Test

   Hypothesis:

   $H_0: \beta_j = 0;$

   $H_1: \beta_j \neq 0$, where $j = 1,5,6,7,8$.

   Significance level: $\alpha = 5\%$.

   Test statistic: $W^2 = \left[\frac{\widehat{\beta_j}}{SE(\widehat{\beta_j})}\right]^2$.

   Critics area, reject H_0 if $W^2 > X^2_{(df,\alpha)}$.

   The Wald values for each variable are as follows:

Table 6. Initial Wald Model Values

Predictor Variable	Wald	$X^2_{(1;0.05)}$	Decision
Constants β_0	0.322	3.481	
Medical History $X_{1(1)}$	5.221	3.481	Reject H_0
Routinely Measure Children's Height $X_{5(1)}$	0.000	3.481	Failed to reject H_0
Routinely Measure Children's Weight $X_{6(1)}$	0.000	3.481	Failed to reject H_0

Predictor Variable	Wald	$X^2_{(1;0.05)}$	Decision
Pregnant Women Get Supplementary Food Provision $X_{7(1)}$	3.754	3.481	Reject H_0
Mother receives iron supplementation during pregnancy $X_{8(1)}$	9.219	3.481	Reject H_0

Based on the information in Table 6, the Wald test results indicate that medical history (Wald = 5.221), supplementary food provision for pregnant women (Wald = 3.754), and maternal iron tablet consumption during pregnancy (Wald = 9.219) are significant predictors of stunting, underscoring the dual importance of infection control in children and adequate maternal nutrition. In contrast, other variables do not affect the model. These results are consistent with previous studies in Indonesia, which found that frequent childhood illness and inadequate maternal nutrition during pregnancy were strongly associated with impaired growth outcomes. For instance, [9] reported that maternal supplementation programs, including iron tablets and additional food, significantly reduced the risk of stunting, while [8] emphasized the role of infectious diseases as a critical determinant. In contrast, routine measurement of children's height and weight showed no statistical significance, echoing findings in Maluku Province, where monitoring activities were not sufficient to influence outcomes unless coupled with effective follow-up interventions. Taken together, these findings reinforce the need for integrated strategies that combine early detection with maternal and child health programs to accelerate stunting prevention in high-burden areas such as East Seram.

Furthermore, it is necessary to form a new model using variables that only affect the response variable. The following are the likelihood ratio and a Wald test for the three predictor variables.

1. Likelihood Ratio Test

   Hypothesis:

   $H_0: \beta_1 = \beta_7 = \beta_8 = 0;$

   H_1: there is at least one $\beta_j \neq 0$, where $j = 1,7,8$.

   Significance level: $\alpha = 5\%$.

   Test Ststistic:

   $$G = -2\ ln\left[\frac{\left(\frac{n_0}{n}\right)^{n_0}\left(\frac{n_1}{n}\right)^{n_1}}{\prod_{i=1}^{n}(\hat{\pi}_i)^{y_i}(1-\hat{\pi}_i)^{1-y_i}}\right] = 38.240.$$

   Critical area: reject H_0 if $G > X^2_{(df,\alpha)}$.

   Decision: Since $G = 38.240 > X^2_{(3;0.05)} = 7.815$, therefore, there is at least one predictor variable that influences the response variable, or simultaneously, the predictor variables influence the stunting status of toddlers.

2. Wald Test

   Hypothesis:

   $H_0: \beta_j = 0$

   $H_1: \beta_j \neq 0$, where $j = 1,7,8$

   Significance level: α =5%

   Test Statitics: $W^2 = \left[\frac{\widehat{\beta_j}}{SE(\widehat{\beta_j})}\right]^2$

   Critical area: reject H_0 if $W^2 > X^2_{(df,\alpha)}$

   The Wald values for each variable are as follows:

Table 7. Initial Wald Model Values

Predictor Variable	Wald	$X^2_{(1;0.05)}$	Decision
Constant	0.010	3.481	
Medical History $X_{1(1)}$	7.885	3.481	Reject H_0
Pregnant women receive supplementary food provision $X_{7(1)}$	3.783	3.481	Reject H_0
Mother receives iron supplementation during pregnancy $X_{8(1)}$	11.843	3.481	Reject H_0

At a significance level of 5%, it can be concluded that the variables that influence the stunting status of toddlers are the variables of having a Medical History (X_1), Pregnant women receive supplementary food provision (X_7), and mothers receiving iron supplements (TTD) during pregnancy (X_8).

3.4.3 Model Fit Test

The model fit test is conducted to evaluate whether the proposed model adequately represents the observed data. The hypotheses for the model fit test are formulated as follows:

H_0: The model is appropriate or there is no difference in observation and prediction results;

H_1: The model is not appropriate or there is a difference in observation and prediction results.

Significant level: $\alpha = 5\%$.

Statistic Test: $\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n_k\bar{\pi}_k)^2}{n_k\bar{\pi}_k(1-\bar{\pi}_k)} = 0.186$.

Critical area: failed to reject H_0 if $\hat{C} < X^2_{(0.05;2)}$.

Decision: Since $\hat{C} = 0.18 < X^2_{(0.05;2)} = 5.591$, at a significance level of 5%, we can conclude that the variables that influence the stunting status of toddlers are the variables of having a medical history.

$$\hat{\pi}(x) = \frac{exp\left(0.132 - 5.559X_{1(1)} + 2.532X_{7(1)} + 4.313X_{8(1)}\right)}{1 + exp\left(0.132 - 5.559X_{1(1)} + 2.532X_{7(1)} + 4.313X_{8(1)}\right)}.$$

3.4.4 Classification

This subsection presents the classification results obtained using the binary logistic regression method. The classification accuracy reflects the model's ability to correctly predict the observed outcomes.

Table 8. Accuracy of Binary Logistic Regression Classification

Observation		Prediction Status	
		No Stunting	Stunting
Status	No Stunting	44	3
	Stunting	2	11

Then, we found:

$$APER(\%) = \frac{n_{10}+n_{01}}{n} \times 100\% = \frac{3+2}{60} \times 100\% = 8.33\%.$$

$$TAR(\%) = 1 - APER = 1 - 8{,}33\% = 91.67\%.$$

The TAR value of 91.67% indicates that the obtained binary logistic model could predict accurately with 91.67% accuracy.

3.5 Comparison of the performance of the MARS model and Binary Logistic Regression

This subsection presents a performance comparison between the MARS model and Binary Logistic Regression. The comparison is conducted in modeling the incidence of stunting among toddlers in Teluk Waru District, SBT Regency, as presented in Table 9.

Table 9. Comparison of MARS Model Performance and Binary Logistic Regression

Accuracy	MARS	Binary Logistic Regression
APER	8.33%	8.33%
TAR	91.67%	91.67%
R^2	68.5%	72.7%

Based on the description of Table 9, it can be seen that both methods have the same performance in classifying stunting status, whereas the APER and TAR values each have the same value, namely 8.33% and 91.67%. However, when viewed from the performance of the model's goodness-of-fit measure, the binary logistic regression model has better performance than the MARS model because it has a larger R^2 value of 72.7%. The results of this study are in line with the conclusions of the study conducted by [22] regarding the evaluation of the performance of the binary logistic regression method with the MARS method for mapping groundwater potential, which states that for small sample sizes, the performance of the binary logistic regression method is better than the MARS.

4. CONCLUSION

Based on the results of the analysis and discussion, several conclusions can be drawn. These conclusions are presented as follows:

1. The number of respondents observed is that as many as 47 or 62% of toddlers are in the non-stunting category, while toddlers in the stunting category are 13 or 38%.
2. The best MARS model for the incidence of toddler stunting in Teluk Waru District, SBT Regency, is the MARS model with a combination of 10 BFs, 2 MIs, and 1 MO, which produces a minimum GCV value of 0.087. The best MARS model can be written as

   $$\hat{f}(x) = 0.344 + 0.468BF_3 + 0.474BF_5 - 0.353BF_9,$$

   with predictor variables that have a significant effect on the incidence of toddler stunting, including the variables of medical history (X_1), routine measuring of children's height (X_5), pregnant women receive supplementary food provision (X_7), and mother receives iron supplementation during pregnancy (X_8) have an effect on the status of toddler stunting.
3. The binary logistic regression model for the incidence of stunting in toddlers in Teluk Waru District, SBT Regency, can be mathematically written as follows:

   $$\hat{\pi}(x) = \frac{exp\left(0.132 - 5.559X_{1(1)} + 2.532X_{7(1)} + 4.313X_{8(1)}\right)}{1 + exp\left(0.132 - 5.559X_{1(1)} + 2.532X_{7(1)} + 4.313X_{8(1)}\right)},$$

   with predictor variables that significantly influence the incidence of toddler stunting, including variables of having a history of illness, pregnant women receiving a supplementary food provision, and mothers receiving iron supplementation during pregnancy.
4. The performance of both models in predicting classification accuracy produced the same APER and TAR values, namely 8.33% and 91.67%, respectively; however, when viewed from the performance of the model's goodness of fit measure, the binary logistic regression model is more appropriate for analyzing the factors influencing the incidence of toddler stunting in Teluk Waru District, SBT Regency because it produces an R^2 value of 72.7% greater than the MARS model.
5. The study is limited by its small sample size and restricted set of variables, suggesting the need for future research with larger datasets and broader determinants such as socioeconomic and environmental factors. Nevertheless, the findings provide practical insights for public health, emphasizing the importance of maternal nutrition programs and infection control as priority strategies to accelerate stunting prevention efforts in high-burden regions of Indonesia.

Author Contributions

Johan Bruiyf Bension: Conceptualization, Data curation, and Formal Analysis. Ferry Kondo Lembang: Data Curation, Formal Analysis, Writing - Review and Editing, and Validation. Nurul Fadillah Idris: Data

Curation, Formal Analysis, Writing - Review and Editing, and Validation. Norisca Lewaherilla: Validation and Investigation. All authors reviewed and approved the final version of the manuscript.

Funding Statement

The author received research funding assistance from the 2024 DIPA PNBP Budget of the Faculty of Medicine, Pattimura University, Ambon.

Acknowledgement

The researcher extends their gratitude to the leadership of Pattimura University and the leadership of the Faculty of Medicine at Pattimura University for their financial support provided through the 2024 Budget Implementation List (DIPA) of Pattimura University, Ambon.

Declarations

The authors confirm that there are no conflicts of interest related to the publication of this manuscript. Since the research did not involve human participants or animal testing, ethical approval and participation consent are not required. All authors have reviewed and approved the final manuscript and have agreed to its publication. Research data can be accessed by contacting the corresponding author, subject to reasonable request. The specific roles and responsibilities of each author are described in the Author Contributions section.

Declaration of Generative AI and AI-assisted technologies

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript, including for writing, editing, data analysis, or the creation of tables and figures.

REFERENCES

[1] Nur Latifah Ariyani, "HUBUNGAN POLA ASUH MAKAN DAN KEBIASAAN MAKAN KELUARGA TERHADAP STATUS GIZI BALITA DI DESA TUMIYANG KECAMATAN PEKUNCEN," Universitas Muhammadiyah Purwokerto, Purwokerto, 2017.

[2] Kementerian Kesehatan Republik Indonesia, "LAPORAN NASIONAL RISET KESEHATAN DASAR TAHUN 2018," 2018.

[3] Y. Karyati, "PENGARUH JUMLAH PENDUDUK MISKIN, LAJU PERTUMBUHAN EKONOMI, DAN TINGKAT PENDIDIKAN TERHADAP JUMLAH STUNTING DI 10 WILAYAH TERTINGGI INDONESIA TAHUN 2010-2019," *Jurnal Riset Ilmu Ekonomi Dan Bisnis*, pp. 101–108, 2021, doi: https://doi.org/10.29313/jrieb.v1i2.401.

[4] R. Kurniawan and B. Yoniarto, *ANALISIS REGRESI DASAR DAN PENERAPANNYA DENGAN R*. Jakarta: Kencana, 2016.

[5] Achmad Faqih, "ANALISIS FAKTOR RESIKO STUNTING MENGGUNAKAN REGRESI LOGISTIK BINER," Universitas Islam Negeri Sunan Ampel, Surabaya, 2020.

[6] D. Ariesta, N. Gusriani, and K. Parmikanti, "ESTIMASI PARAMETER MODEL REGRESI NONPARAMETRIK B-SPLIN E PADA ANGKA KEMATIAN MATERNAL," *Jurnal Matematika UNAND*, vol. 10, no. 3, pp. 342–354, 2021, doi: https://doi.org/10.25077/jmu.10.3.342-354.2021.

[7] R. Binadari, Y. Wilandari, and S. Suparti, "PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) PADA PEMINATAN JURUSAN SMA (STUDI KASUS SMA NEGERI 2 SEMARANG)," *Jurnal Gaussian*, vol. 4, no. 4, pp. 987–996, 2015, doi: https://doi.org/10.14710/j.gauss.4.4.987-996.

[8] T. R. Elisa, "PENERAPAN METODE MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) UNTUK KLASIFIKASI BALITA STUNTING DI KECAMATAN PADANG TIMUR," Universitas Andalas, Padang, 2022.

[9] A. Utami, "ANALISIS FAKTOR-FAKTOR YANG DIDUGA BERPENGARUH TERHADAP KASUS BALITA STUNTING DI JAWA TIMUR MENGGUNAKAN REGRESI LOGISTIK BINER," Institut Teknologi Sepuluh Nopember, Surabaya, 2024.

[10] R. Lu, T. Duan, M. Wang, H. Liu, S. Feng, X. Gong, H. Wang, J. Wang, Z. Cui, Y. Liu, C. Li and Jun Ma., "THE APPLICATION OF MULTIVARIATE ADAPTIVE REGRESSION SPLINES IN EXPLORING THE INFLUENCING FACTORS AND PREDICTING THE PREVALENCE OF HBA1C IMPROVEMENT," *Ann Palliat Med*, vol. 10, no. 2, pp. 1291303–1296303, 2021, doi: https://doi.org/10.21037/apm-19-406.

[11] M. Gackowski, K. Szewczyk-Golec, R. Pluskota, M. Koba, K. Mądra-Gackowska, and A. Woźniak, "APPLICATION OF MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARSPLINES) FOR PREDICTING ANTITUMOR ACTIVITY OF ANTHRAPYRAZOLE DERIVATIVES," *Int J Mol Sci*, vol. 23, no. 9, p. 5132, 2022, doi: https://doi.org/10.3390/ijms23095132.

[12] C.-F. Chang, T.-W. Chu, C.-H. Liu, S.-T. Wu, and C.-C. Yang, "EQUATION BUILT BY MULTIPLE ADAPTIVE REGRESSION SPLINE TO ESTIMATE BIOLOGICAL AGE IN HEALTHY POSTMENOPAUSAL WOMEN IN TAIWAN," *Diagnostics*, vol. 15, no. 9, p. 1147, 2025, doi: https://doi.org/10.3390/diagnostics15091147.

[13] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alphabet, 2019.

[14] E. Y. Sari, "ANALISIS SURVIVAL DENGAN PENDEKATAN MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) UNTUK DATA RESAMPLING," Universitas Lampung, Lampung, 2016.

[15] A. Adityaningrum, "ESTIMASI PROSPENSITY SCORE MATCHING BERDASARKAN PENDEKATAN MULTIVARIATE ADAPTIVE REGRESSION SPLINES," Institut Teknologi Sepuluh Nopember, Surabaya, 2017.

[16] S. Anam, S. Sugiman, and S. Sunarmi, "KETEPATAN KLASIFIKASI DENGAN MENGGUNAKAN METODE MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) PADA DATA KELOMPOK RUMAH TANGGA KABUPATEN CILACAP," *Unnes Journal of Mathematics*, vol. 6, no. 1, pp. 92–101, 2017, doi: https://doi.org/10.15294/ujm.v6i1.13638.

[17] M. B. Adiguzel and M. A. Cengiz, "MODEL SELECTION IN MULTIVARIATE ADAPTIVE REGRESSIONS SPLINES (MARS) USING ALTERNATIVE INFORMATION CRITERIA," *Heliyon*, vol. 9, no. 9, 2023, doi: https://doi.org/10.1016/j.heliyon.2023.e19964.

[18] C. Conoscenti, M. Ciaccio, N. A. Caraballo-Arias, Á. Gómez-Gutiérrez, E. Rotigliano, and V. Agnesi, "ASSESSMENT OF SUSCEPTIBILITY TO EARTH-FLOW LANDSLIDE USING LOGISTIC REGRESSION AND MULTIVARIATE ADAPTIVE REGRESSION SPLINES: A CASE OF THE BELICE RIVER BASIN (WESTERN SICILY, ITALY)," *Geomorphology*, vol. 242, pp. 49–64, 2015, doi: https://doi.org/10.1016/j.geomorph.2014.09.020.

[19] D. W. Hosmer, S. Lemeshow, and S. May, "APPLIED SURVIVAL ANALYSIS," *Wiley Series in Probability and Statistics*, p. 60, 2008.

[20] A. Strzelecka, A. Kurdyś-Kujawska, and D. Zawadzka, "APPLICATION OF LOGISTIC REGRESSION MODELS TO ASSESS HOUSEHOLD FINANCIAL DECISIONS REGARDING DEBT," *Procedia Comput Sci*, vol. 176, pp. 3418–3427, 2020, doi: https://doi.org/10.1016/j.procs.2020.09.055.

[21] S. Risambessy, S. N. Aulele, and F. K. Lembang, "MISCLASSIFICATION ANALYSIS OF ELEMENTARY SCHOOL ACCREDITATION DATA IN AMBON CITY USING MULTIVARIATE ADAPTIVE REGRESSION SPLINE," *Jurnal Matematika, Statistika Dan Komputasi*, vol. 18, no. 3, pp. 394–406, 2022, doi: https://doi.org/10.20956/j.v18i3.19451.

[22] S. Park, S.-Y. Hamm, H.-T. Jeon, and J. Kim, "EVALUATION OF LOGISTIC REGRESSION AND MULTIVARIATE ADAPTIVE REGRESSION SPLINE MODELS FOR GROUNDWATER POTENTIAL MAPPING USING R AND GIS," *Sustainability*, vol. 9, no. 7, p. 1157, 2017, doi: https://doi.org/10.3390/su9071157.