

# ADE\_\_Cauchy\_Template (4)

*by* Turnitin Checker

---

**Submission date:** 08-Jun-2025 10:40PM (UTC-0400)

**Submission ID:** 2694955080

**File name:** ADE\_\_Cauchy\_Template\_4\_.pdf (651.15K)

**Word count:** 7654

**Character count:** 38761



# Modeling Risk Factors of Acute Respiratory Infections using Logistic Regression and Multivariate Adaptive Regression Splines

Ardi Kurniawan<sup>1,\*</sup>, Nathania Fauziah<sup>2</sup>, Arinda Mahadesyawardani<sup>3</sup>, Syifa' Azizah Putri Gunawan<sup>4</sup>, and Aurellia Calista Anggakusuma<sup>5</sup>

<sup>1,2,3,4,5</sup> *Statistics Study Program, Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*

*\*Corresponding author's email: [ardi-k@fst.unair.ac.id](mailto:ardi-k@fst.unair.ac.id)*

## Abstract

Acute Respiratory Infections (ARI) remain a leading cause of morbidity among toddlers, particularly in regions with limited healthcare access. This study aimed to model the risk factors of ARI in toddlers using Binary Logistic Regression and Multivariate Adaptive Regression Splines (MARS). Using secondary data from Southeast Aceh, seven predictor variables were analyzed, including maternal characteristics, breastfeeding status, and household conditions. Both models were statistically significant in identifying key predictors. Logistic regression showed superior performance with 86.96% accuracy, 85.00% precision, 91.89% recall, 81.25% specificity, and 88.30% F1-score. In contrast, MARS achieved a higher recall (97.30%) but lower specificity (62.50%), indicating higher sensitivity but a greater likelihood of false positives. Exclusive breastfeeding, home ventilation, and housing density were significant predictors in both models. Overall, logistic regression was found to be the more reliable and interpretable method, offering better balance in classification metrics. These findings support the use of logistic regression for identifying ARI risk factors in similar contexts and contribute to improved data-driven public health strategies aimed at reducing ARI incidence among vulnerable populations.

**Keywords:** Acute Respiratory Infections; Binary Logistic Regression; Multivariate Adaptive Regression Splines; Classification Model; Toddlers.

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

## 1 Introduction

Acute Respiratory Infections (ARI) are contagious respiratory diseases that represent a leading cause of death worldwide, accounting for 2.6 million deaths annually [1]. These infectious diseases are caused by viruses, fungi, bacteria, or a combination of all three, and are generally characterized by symptoms such as fever, sore throat, cough, and flu-like conditions [2]. According to reports from the World Health Organization (WHO), approximately 3.5% of the total global burden of diseases is attributed to ARI. Globally, ARI are responsible for about 15% of all deaths among toddlers, with the majority of these cases occurring in low- and middle-income countries. Respiratory infections contribute to 33% of total deaths among toddler, particularly in Southeast Asia [3].

As a developing country and part of Southeast Asia, Indonesia is not exempt from the burden of ARI. Data from the Ministry of Health show that in 2023, there was an increase in ARI cases, reaching 285,623 in July alone, with a prevalence rate of 3.5 per 100 toddler. The highest prevalence by age group

was recorded among toddlers aged 1 to 4 years [4]. Toddlers are especially vulnerable to ARI due to their still developing immune systems. This condition is exacerbated by various factors, including environmental conditions, socioeconomic disparities, and unequal access to healthcare services across regions. One area that deserves special attention is Aceh Tenggara Regency, which is geographically located in a hilly region with limited access in several sub-districts [5]. According to the 2024 Aceh Province Health Profile, pneumonia cases among toddler accounted for 10% of the total under-five population in the region, totaling 45,280 cases. Meanwhile, the number of ARI cases in Aceh Tenggara Regency is relatively high compared to several other regencies in Aceh Province. Although the Ministry of Health has introduced programs such as the Integrated Management of Childhood Illness (IMCI) and the revitalization of Posyandu (integrated health posts) to detect and manage ARI cases, operational barriers in remote areas have hindered the effectiveness of these initiatives in reaching vulnerable groups, particularly toddler in Southeast Aceh

The incidence of acute ARI in toddler is influenced by a variety of interconnected and complex risk factors. A study conducted by Titi Saparina L. and Rasni Intan [6] stated that the size of ventilation openings and housing density have a significant relationship with the incidence of ARI in toddler. Furthermore, [7] found that children who live in highly crowded rooms are at greater risk of developing ARI. Research by [8] also showed that exclusive breastfeeding is associated with a reduction in infectious diseases, including ARI among toddler. Exclusive breastfeeding provides natural antibodies that help strengthen the child's immune system. Socioeconomic factors such as low maternal education, unstable parental employment, and limited access to healthcare facilities, also contribute to increased vulnerability to this disease. [9] found that mothers with good knowledge and positive attitudes toward child health are more likely to recognize ARI symptoms early and take appropriate preventive actions. Therefore, maternal education also has an indirect impact by improving health literacy and access to medical services.

A data driven approach is becoming increasingly important for detecting and understanding risk factors in greater depth. Binary logistic regression as a parametric statistical method is considered more effective than standard linear approaches in identifying health risk factors, as it can accommodate various types of predictor variables, both categorical and numerical, making it more adaptable to the complex nature of health related data [10]. Meanwhile, Multivariate Adaptive Regression Splines (MARS) is a nonparametric method that builds regression models using combinations of basis functions, enabling it to capture nonlinear relationships and interactions between variables without rigid prior assumptions [11].

One study applied binary logistic regression to the case of ARI in toddler, conducted by [12], which analyzed the influence of knowledge, environmental, and behavioral factors on the incidence of ARI among toddler at a health center in Southeast Aceh. However, this study did not explicitly present the logistic regression model nor provide classification evaluation metrics to assess the model's performance. Moreover, the approach used was parametric, relying solely on logistic regression analysis to examine the relationships between variables through significance values and odds ratios, thus limiting its role as a predictive modeling tool. These limitations will be addressed in the present study by using the same dataset but constructing an explicit logistic regression model accompanied by classification evaluation metrics, which will then be compared to a nonparametric approach, namely Multivariate Adaptive Regression Splines (MARS). The comparison of parametric and nonparametric methods in this study is supported by [13], who compared binary logistic regression and binary MARS in modeling hypertension risk in Indonesia. Although the disease context differs, the study showed that MARS can be an effective alternative in disease risk modeling. Classification accuracy was used to compare model performance, providing a useful reference for assessing the effectiveness of both approaches in detecting ARI among toddler.

This study presents a novel contribution by not only reconstructing the binary logistic regression model using an existing dataset but also comparing it with MARS method, which has not been applied in this context before. The research aims to explore and model the risk factors associated with ARI in toddler using both logistic regression and MARS to identify which method offers better predictive performance and interpretability. Through this approach, the study enhances the understanding of health

data modeling for childhood diseases and provides insights that can support more effective prevention strategies. Additionally, the findings contribute to the achievement of Sustainable Development Goal (SDGs) 3, specifically Target 3.2 which aims to end preventable deaths of newborns and children under 5 years of age by 2030 [14].

## 2 Methods

### 2.1 Data Source and Variables

This study utilizes secondary data obtained from the thesis by [12]. The original data were collected directly by the researcher through interviews, questionnaires, and observations. This study employs variables suspected to be risk factors for the incidence of ARI in toddler at the Deleng Pokhkisen Health Center in Aceh Tenggara District. A detailed description of the response and predictor variables used in this study is presented in Table 1 below.

Table 1: Research Variables

| Variable Type | Variable Name             | Description                                                                                                      | Data Scale |
|---------------|---------------------------|------------------------------------------------------------------------------------------------------------------|------------|
| $Y$           | ARI Incidence in Toddlers | 1: Suffering from ARI<br>0: Not Suffering from ARI                                                               | Nominal    |
| $X_1$         | Mother's Age              | 0: 26–32 years<br>1: 18–25 years                                                                                 | Nominal    |
| $X_2$         | Mother's Education        | 1: No Schooling<br>2: Elementary School<br>3: Junior High School<br>4: Senior High School<br>5: Higher Education | Ordinal    |
| $X_3$         | Mother's Knowledge        | 0: Poor<br>1: Good                                                                                               | Nominal    |
| $X_4$         | Mother's Attitude         | 0: Negative<br>1: Positive                                                                                       | Nominal    |
| $X_5$         | Exclusive Breastfeeding   | 0: Not Given<br>1: Given                                                                                         | Nominal    |
| $X_6$         | House Ventilation         | 0: Does Not Meet the Standard<br>1: Meets the Standard                                                           | Nominal    |
| $X_7$         | Housing Density           | 0: Dense<br>1: Not dense                                                                                         | Nominal    |

### 2.2 Binary Logistic Regression

Logistic regression employs the logit function to estimate the probability of an event occurring [15]. In logistic regression, it is assumed that binary responses are independent, and thus follow a binomial distribution. From Equation 1, the probability of success for the  $i$  observation, given the predictor variables  $x_i$ , is modeled using a binomial distribution as expressed in Equation 1 below.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)} \quad (1)$$

The interpretation of  $\beta$  in logistic regression refers to the log odds, which is obtained by exponentiating the coefficient  $e^{\beta_i}$  [16]. The odds in a binary logistic model are defined as the ratio between the probability of success and the probability of failure. To assess the simultaneous influence of the predictors, the Likelihood Ratio Test (G-test) is employed at a 5% significance level. The hypotheses are formulated as follows :

113  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$   
 114  $H_1: \text{At least one } \beta_i \neq 0 \text{ for } i = 1, 2, 3, 4, 5, 6, 7$

115 The G-test statistic is defined as:

$$G = -2 \log \left( \frac{L_0}{L_1} \right) \quad (2)$$

116 The null hypothesis  $H_0$  is rejected if  $G > \chi^2_{(\alpha, p-1)}$  or the p-value  $< \alpha$ , indicating that the predictor  
 117 variables jointly have a significant effect on the incidence of ARI in toddler. To test the partial effect of  
 118 each predictor variable, the Wald test is applied. The hypotheses are:

119  $H_0: \beta_i = 0 \text{ for } i = 1, 2, 3, 4, 5, 6, 7$   
 120  $H_1: \beta_i \neq 0 \text{ for } i = 1, 2, 3, 4, 5, 6, 7$

121 The Wald test statistic is given by Equation 3:

$$W = \left( \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right)^2 \quad (3)$$

122 here,  $\hat{\beta}_i$  represents the estimated parameter for the  $i$ th variable and  $SE(\hat{\beta}_i) = \sqrt{(\sigma^2(\hat{\beta}_i))}$  denotes the  
 123 standard error associated with that estimate. The null hypothesis is rejected if  $W \geq Z_{\alpha/2}$  or the  $p$ -  
 124 value  $< \alpha$ , indicating that the  $i$  predictor has a statistically significant effect on the probability of ARI  
 125 occurrence in toddler.

126 The model's fit to the observed data in logistic regression is evaluated using the Hosmer-Lemeshow  
 127 test. This test examines how closely the predicted outcomes from the model match the actual outcomes  
 128 recorded. A good model should satisfy the Goodness of Fit (GoF) criterion. The test statistic for the  
 129 Hosmer and Lemeshow method is determined using Equation 4:

$$\hat{C} = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i(1 - E_i)} \quad (4)$$

130 where  $O_i$  denotes the observed count in the  $i$ th group,  $E_i$  is the expected count in the same group, and  $g$   
 131 refers to the total number of groups. The model is considered to have a good fit with the data when the  
 132 test statistic satisfies  $\hat{C} < \chi^2_{(\alpha, g-2)}$  or  $p\text{-value} > \alpha$ .

### 133 2.3 Multivariate Adaptive Regression Splines (MARS)

134 MARS is a nonparametric regression method developed by Friedman in 1991, as expressed in Equa-  
 135 tion 5 [17].

$$y_i = a_0 + \sum_{m=1}^M a_m B_{mi}(x, t) + \varepsilon_i \quad (5)$$

136 In this study, the dependent variable is binary, meaning it consists of two categories. Therefore, the  
 137 outcome can be interpreted as the probability of a particular event occurring. To identify the predictor  
 138 variables that significantly influence the binary outcome, one may analyze the odds ratio [13]. The  
 139 probability that  $Y = 1$  is given in Equation 6.

$$P(Y = 1 | X_1, X_2, \dots, X_p) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}} \quad (6)$$

140 Accordingly, this probability can also be expressed using the basis functions of MARS, as shown in  
 141 Equation 7 below:

$$P = \frac{\exp(BF_0 + BF_i(X_i))}{1 + \exp(BF_0 + BF_i(X_i))} \quad (7)$$

142 where  $i$  denotes the basis function components obtained from the optimal MARS model.

The modeling process in MARS involves forward selection and backward elimination, which are iteratively performed until the Generalized Cross Validation (GCV) score is minimized, indicating the most optimal model. The GCV criterion is defined in Equation 8 [18].

$$GCV(M) = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2}{\left[1 - \frac{C(M)}{n}\right]^2} \quad (8)$$

As explained by [19], significance testing is also conducted in MARS to examine parameter significance and assess model adequacy. A simultaneous test is performed using the F-statistic shown in Equation 9:

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (N - M - 1)} \quad (9)$$

In this study, the hypotheses for the simultaneous test are as follows:

$H_0: a_1 = a_2 = \dots = a_M = 0$  (The model is not significant)

$H_1: \text{At least one } a_m \neq 0$  (the model is significant)

The null hypothesis is rejected if  $F_{\text{stat}} > F_{\alpha}(M, N - M - 1)$  or  $p\text{-value} < \alpha$ , indicating that at least one predictor variable significantly affects the response variable. For partial significance testing of individual parameters,  $t$ -test shown in Equation 10:

$$t = \frac{\hat{a}_m}{\sqrt{\text{var}(\hat{a}_m)}} \quad (10)$$

The hypotheses for the partial test are:

$H_0: a_m = 0$  (basis function  $B_m(x)$  has no effect on the model)

$H_1: a_m \neq 0$  (basis function  $B_m(x)$  has a significant effect on the model)

The null hypothesis is rejected if  $t > t_{\alpha/2(N-M-1)}$  or  $p\text{-value} < \alpha$ .

## 2.4 Evaluation Matrix

The evaluation matrix serves as a tool to measure how well a classification model performs in identifying specific factors. In health-related studies, typical performance indicators include accuracy, precision, recall, and the F1-score. These indicators are typically derived from the confusion matrix, as illustrated in Table 2 below:

**Table 2:** Classification evaluation metrics and their formulas

| Classification Evaluation | Calculation                                                                               |
|---------------------------|-------------------------------------------------------------------------------------------|
| Accuracy                  | $\frac{TP + TN}{TP + TN + FP + FN}$                                                       |
| Precision                 | $\frac{TP}{TP + FP}$                                                                      |
| Recall                    | $\frac{TP}{TP + FN}$                                                                      |
| F1-Score                  | $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ |

where:

**TP** (True Positive): Positive cases that were accurately identified as positive.

166 **TN** (True Negative): Negative cases that were accurately identified as negative.  
 167 **FP** (False Positive): Negative cases that were mistakenly classified as positive.  
 168 **FN** (False Negative): Positive cases that were mistakenly classified as negative.

## 169 2.5 Research Stage

- 170 1. The factors influencing ARI were modeled using binary logistic regression with *SPSS* software,  
 171 following these steps:
  - 172 (a) Testing the independence among predictor variables.
  - 173 (b) Estimating logistic regression parameters and evaluating their significance both simultane-  
 174 ously using the Likelihood Ratio test and individually using the Wald test.
  - 175 (c) Evaluating the overall model fit using the Hosmer–Lemeshow goodness-of-fit test.
  - 176 (d) Interpreting the model outcomes and calculating epidemiological association measures.
- 177 2. The factors influencing ARI were modeled using MARS with the *MARS* software, through the  
 178 following procedure:
  - 179 (a) Seven predictor variables were used, with the number of basis functions limited to 14, 21,  
 180 and 28. The maximum number of interactions was set to 1 and 2, while the minimum number  
 181 of observations between knots was set at 0, 1, and 2.
  - 182 (b) The optimal MARS model was selected based on the lowest Generalized Cross Validation  
 183 (GCV), the minimum Mean Squared Error (MSE), and the highest R-squared ( $R^2$ ) values  
 184 from various combinations of basis functions, interaction levels, and knot intervals.
  - 185 (c) A binary MARS model was built using the parameter estimates from the previously selected  
 186 best basis function model.
  - 187 (d) The significance of the MARS model parameters was evaluated to assess model fit using both  
 188 simultaneous and partial regression coefficient tests.
- 189 3. Comparing the classification performance of the binary logistic and MARS models based on eval-  
 190 uation metrics.

## 191 3 Results and Discussion

### 192 3.1 Descriptive Analysis of Data

193 The proportion of toddlers suffering and not suffering from ARI is presented in Figure 1 below.

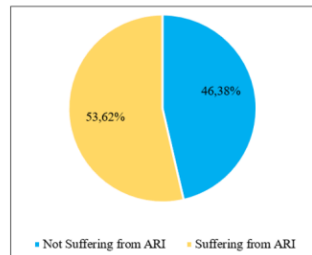


Figure 1: Distribution of Toddlers Based on ARI Status

194 Based on the distribution in Figure 1, indicates that more than half of the observed toddlers experi-  
 195 enced ARI, which may be influenced by various environmental and maternal factors. These proportions  
 196 provide an overview of the prevalence of ARI in the studied population and justify further investigation  
 197 into its risk factors.

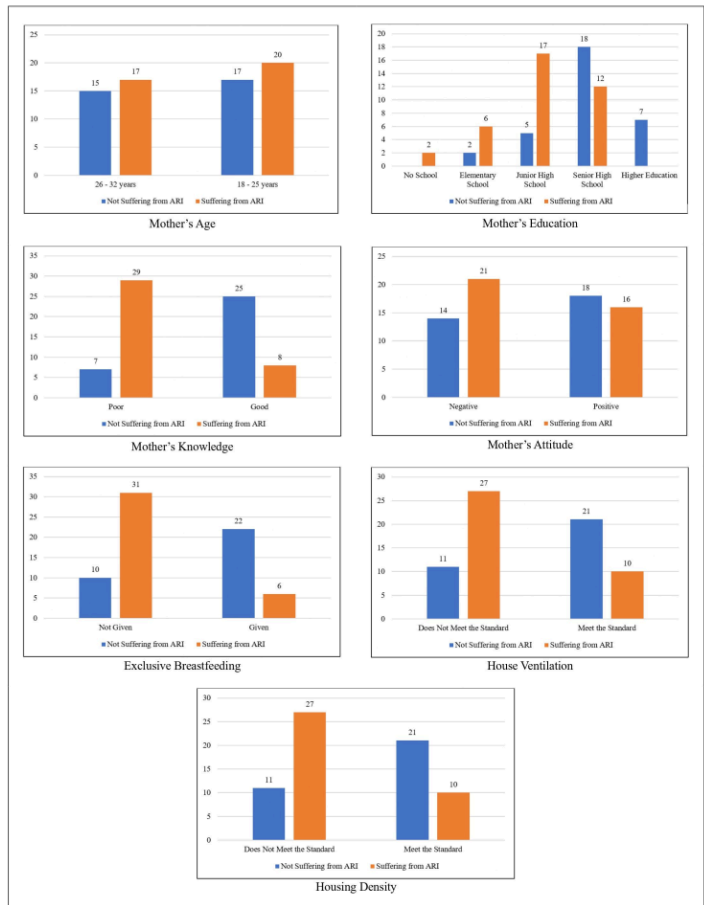


Figure 2: Distribution of ARI Status by Predictor Variables

Figure 2 illustrates the distribution of ARI cases in toddlers based on seven predictor variables. The figure reveals that higher incidences of ARI are observed among children of younger mothers ( $X_1$ ), mothers with lower education levels ( $X_2$ ), poor maternal knowledge ( $X_3$ ), and negative maternal attitudes ( $X_4$ ). Furthermore, toddlers who were not exclusively breastfed ( $X_5$ ), lived in homes with inadequate ventilation ( $X_6$ ), and resided in areas with high housing density ( $X_7$ ) also exhibited a noticeably higher proportion of ARI. These patterns suggest that both maternal factors and living conditions play a critical role in the risk of developing acute respiratory infections in toddler.

## 3.2 Binary Logistic Regression Modelling

### 3.2.1 Multicollinearity Test

The multicollinearity test is used to detect whether there is a strong correlation between two or more predictor variables in a binary logistic regression analysis. In an ideal regression model, the predictor variables should not exhibit high intercorrelation. This test is conducted by examining the *Variance Inflation Factor* (VIF). The hypotheses used are as follows:

$H_0$ : There is no multicollinearity among the predictor variables.

$H_1$ : There is multicollinearity among the predictor variables.

The critical region is defined such that  $H_0$  is rejected if  $VIF > 10$ . The results of the multicollinearity test are presented in Table 3 below:

**Table 3:** Multicollinearity Test Results for Factors Affecting ARI Cases in Toddlers

| Predictor Variable      | Tolerance | VIF   |
|-------------------------|-----------|-------|
| Mother's Age            | 0.813     | 1.230 |
| Mother's Education      | 0.459     | 2.179 |
| Mother's Knowledge      | 0.473     | 2.114 |
| Mother's Attitude       | 0.514     | 1.944 |
| Exclusive Breastfeeding | 0.782     | 1.278 |
| Home Ventilation        | 0.795     | 1.258 |
| Household Density       | 0.800     | 1.249 |

Based on Table 3, it can be concluded that all VIF values are below 10, and therefore we fail to reject  $H_0$ . This indicates that there is no multicollinearity among the predictor variables used in the model.

### 3.2.2 Significance Testing of Binary Logistic Regression Model

The significance test of the model parameters is conducted to evaluate how much the predictor variables contribute in explaining the variation of the response variable, namely the incidence of ARI in toddler. In testing the model parameters will be done simultaneously with the G test and partially with the Wald test. The outcomes of the simultaneous parameter significance test are shown in Table 4.

**Table 4:** Simultaneous Parameter Test

|       | Chi-Square | df | P-Value |
|-------|------------|----|---------|
| Step  | 54.492     | 10 | 0.000   |
| Block | 54.492     | 10 | 0.000   |
| Model | 54.492     | 10 | 0.000   |

Based on Table 4, the p-value obtained is 0.000, which is less than  $\alpha = 5\%$ . Therefore, the decision is to reject  $H_0$ , and it can be concluded that there is a significant effect of the predictor variables on the incidence of ARI in toddler. Another side, the results of the partial parameter significance test are presented in Table 5.

**Table 5:** Partial Parameter Test

|                         | B       | S.E.      | Wald  | df | P-Value |
|-------------------------|---------|-----------|-------|----|---------|
| Constant                | 23.502  | 28296.171 | 0.000 | 1  | 0.999   |
| Mother's Age            | 1.234   | 0.919     | 1.801 | 1  | 0.180   |
| Mother's Education (1)  | -21.599 | 28296.171 | 0.000 | 1  | 0.999   |
| Mother's Education (2)  | -20.136 | 28296.171 | 0.000 | 1  | 0.999   |
| Mother's Education (3)  | -20.895 | 28296.171 | 0.000 | 1  | 0.999   |
| Mother's Education (4)  | -41.058 | 31289.017 | 0.000 | 1  | 0.999   |
| Mother's Knowledge      | -1.265  | 1.109     | 1.302 | 1  | 0.340   |
| Mother's Attitude       | 0.852   | 0.894     | 0.909 | 1  | 0.340   |
| Exclusive Breastfeeding | -2.497  | 1.027     | 5.908 | 1  | 0.015   |
| Home Ventilation        | -2.073  | 0.966     | 4.610 | 1  | 0.032   |
| Housing Density         | -2.208  | 0.955     | 5.351 | 1  | 0.021   |

Based on Table 5, it was found that the variables *exclusive breastfeeding*, *home ventilation*, and *housing density* have p-values less than  $\alpha$  (5%), leading to the decision to reject  $H_0$ . Therefore, it can be concluded that these three variables have a partial effect on the incidence of ARI in toddler. Meanwhile, the other variables do not have a partial effect, as their p-values are greater than  $\alpha$  (5%).

### 3.2.3 Goodness of Fit Test

A model fit test was conducted to assess the suitability of the binary logistic regression model, specifically to determine whether there is a significant difference between the observed and predicted values. The Hosmer–Lemeshow test was used for this evaluation. The results of the model fit test are presented in Table 6 below.

**Table 6:** Model Goodness-of-Fit Test Results

| Chi-Square | df | P - Value |
|------------|----|-----------|
| 12.899     | 8  | 0.115     |

Based on Table 6, the p-value is 0.115, which is greater than  $\alpha$  (5%). Therefore, we fail to reject  $H_0$  and conclude that the binary logistic regression model fits the data adequately.

### 3.2.4 Coefficient of Determination Test

One approach to assess how well a logistic regression model fits the data is through the Nagelkerke R Square test. A higher value of this statistic indicates that a larger portion of the variance in the dependent variable is accounted for by the independent variables in the model. A value near 1 suggests strong explanatory capability, whereas a value closer to 0 implies weak explanatory power. The outcomes of the model's goodness-of-fit assessment using the Nagelkerke R Square are shown in Table 7.

**Table 7:** Determination Coefficient Test Results

| -2 Log Likelihood | Cox & Snell R Square | Nagelkerke R Square |
|-------------------|----------------------|---------------------|
| 12.899            | 8                    | 0.115               |

Based on Table 7, the Nagelkerke R Square value is 0.115. This indicates that the predictive variables in the model are able to explain 11.5% of the variation in the response variable. In other words, 88.5% of the variation in the response variable is influenced by other factors that are not included in this logistic regression model.

### 3.2.5 Parameter Estimation of Binary Logistic Regression Model

The model estimation was performed using analysis results that included only predictor variables with a statistically significant effect on the incidence of ARI among toddler. The coefficient estimates of these significant variables are shown Table 8 below.

**Table 8:** Binary Logistic Regression Model Estimation

| Variable in the Equation | B      | S.E   | Wald   | df | Sig.  |
|--------------------------|--------|-------|--------|----|-------|
| Exclusive Breastfeeding  | -2.598 | 0.781 | 11.062 | 1  | 0.001 |
| Home Ventilation         | -2.409 | 0.792 | 9.247  | 1  | 0.002 |
| Housing Density          | -1.985 | 0.715 | 7.705  | 1  | 0.006 |
| Constant                 | 3.292  | 0.810 | 16.525 | 1  | 0.000 |

Based on Table 8, the parameter values  $\beta$  used to construct the initial binary logistic regression model were obtained by considering the factors that influence the incidence of ARI in toddler. The logistic regression model obtained is:

$$\pi(x_i) = \frac{\exp(3.292 - 2.598x_5 - 2.409x_6 - 1.985x_7)}{1 + \exp(3.292 - 2.598x_5 - 2.409x_6 - 1.985x_7)} \quad (11)$$

Since Equation 11 is nonlinear, a logit transformation is required to form the logistic regression model:

$$g(x) = 3.292 - 2.598x_5 - 2.409x_6 - 1.985x_7 \quad (12)$$

Based on the resulting model, it can be concluded that an increase of one unit in exclusive breastfeeding reduces the risk of ARI by 2.598 units. Likewise, an improvement of one unit in home ventilation reduces the risk of ARI by 2.409 units, and a one-unit decrease in housing density results in a reduction of 1.985 units in ARI risk.

### 3.2.6 Classification Results of the Binary Logistic Regression Model

The evaluation of a binary logistic regression model involves several performance metrics that indicate how well the model can distinguish between categories in the dependent variable. Table 9 presents the classification results of the binary logistic regression model.

**Table 9:** Classification Results of the Binary Logistic Regression Model

| Observed               | Predicted              |                    |
|------------------------|------------------------|--------------------|
|                        | Not Suffering from ARI | Suffering from ARI |
| Not Suffering from ARI | 26                     | 6                  |
| Suffering from ARI     | 3                      | 34                 |

## 3.3 Multivariate Adaptive Regression Splines (MARS) Modeling

### 3.3.1 Best MARS Model

Before estimating the MARS model, the first step is to determine the optimal combination of parameters. This study used 7 predictor variables, with the number of basis functions (BF) set to 14, 21, and 28; the maximum interaction (MI) set to 1 and 2; and the minimum number of observations (MO) set to 0, 1, and 2. All possible combinations of BF, MI, and MO were evaluated to identify the best model configuration based on the lowest GCV and MSE values, and the highest  $R^2$  value. The calculation results are presented in Table 10 below.

**Table 10:** GCV Calculation Results for Each Model Combination

| Model    | BF        | MI       | MO       | GCV          | MSE          | $R^2$        |
|----------|-----------|----------|----------|--------------|--------------|--------------|
| 1        | 14        | 1        | 0        | 0.173        | 0.142        | 0.323        |
| 2        | 14        | 1        | 1        | 0.180        | 0.137        | 0.292        |
| 3        | 14        | 1        | 2        | 0.180        | 0.137        | 0.292        |
| <b>4</b> | <b>14</b> | <b>2</b> | <b>0</b> | <b>0.155</b> | <b>0.123</b> | <b>0.393</b> |
| 5        | 14        | 2        | 1        | 0.155        | 0.123        | 0.393        |
| 6        | 14        | 2        | 2        | 0.155        | 0.123        | 0.393        |
| 7        | 21        | 1        | 0        | 0.181        | 0.142        | 0.292        |
| 8        | 21        | 1        | 1        | 0.192        | 0.137        | 0.273        |
| 9        | 21        | 1        | 2        | 0.181        | 0.142        | 0.292        |
| 10       | 21        | 2        | 0        | 0.155        | 0.123        | 0.393        |
| 11       | 21        | 2        | 1        | 0.155        | 0.123        | 0.393        |
| 12       | 21        | 2        | 2        | 0.155        | 0.123        | 0.393        |
| 13       | 28        | 1        | 0        | 0.192        | 0.142        | 0.248        |
| 14       | 28        | 1        | 1        | 0.205        | 0.155        | 0.201        |
| 15       | 28        | 1        | 2        | 0.192        | 0.142        | 0.248        |
| 16       | 28        | 2        | 0        | 0.155        | 0.123        | 0.393        |
| 17       | 28        | 2        | 1        | 0.155        | 0.123        | 0.393        |
| 18       | 28        | 2        | 2        | 0.155        | 0.123        | 0.393        |

Based on Table 10, several model parameter combinations produce identical evaluation results, notably models 4–6, 13–15, and 22–24. All of these models yield the minimum GCV value of 0.155, MSE of 0.123, and the highest  $R^2$  value of 0.393, meeting the criteria for the best MARS model. According to the parsimony principle—preferring the simplest model among those with equivalent predictive performance. Model 4 is selected as the optimal model, with BF = 14, MI = 2, and MO = 0. This model not only demonstrates strong performance but is also structurally simpler, making it easier to interpret and apply.

### 3.3.2 Parameter Estimation of MARS Model

Based on the best MARS model previously obtained, estimation can be conducted to model ARI cases in toddler as shown in Table 11 below.

**Table 11:** Parameter Estimation Results of the MARS Model

| Basis Function (BF)                                      | Parameter Estimation |
|----------------------------------------------------------|----------------------|
| $BF_0$                                                   | 0.840                |
| $BF_3 = (X_7 = 1)BF_1$ ; where $BF_1 = (X_3 = 1)$        | −0.553               |
| $BF_8 = \max(0, X_2 - 1)BF_2$ ; where $BF_2 = (X_5 = 1)$ | −0.117               |

Based on Table 11, the MARS model for estimating ARI cases in toddler at the Deleng Pookhisen Health Center, Aceh Tenggara Regency is as follows:

$$Y = 0.840 - 0.553BF_3 - 0.117BF_8 \quad (13)$$

Based on Equation 13, the MARS model indicates two main interaction terms occurring in  $BF_3$  and  $BF_8$ , with the interpretation of each basis function described as follows:

- Basis function  $BF_3 = (X_7 = 1)(X_3 = 1)$  captures the interaction between housing density ( $X_7$ ) and maternal knowledge about ARI ( $X_3$ ).  $BF_3$  equals 1 only if the child lives in a low-density household and the mother has good knowledge of ARI. With a negative coefficient of −0.553, this suggests that such a combination significantly reduces the risk of ARI in toddler at Deleng Pookhisen Health Center.

- b. Basis function  $BF_8 = \max(0, X_2 - 1)(X_5 = 1)$  represents the interaction between maternal education level ( $X_2$ ) and exclusive breastfeeding practice ( $X_5$ ).  $BF_8$  equals 1 when the mother has education above primary school and provides exclusive breastfeeding. Its coefficient of  $-0.117$  implies a moderate reduction in ARI risk, although less impactful than  $BF_3$ .

Since the response variable is binary, the probability model for ARI occurrence is expressed as:

$$P(Y = 1) = \pi(x) = \frac{\exp(0.840 - 0.553BF_3 - 0.117BF_8)}{1 + \exp(0.840 - 0.553BF_3 - 0.117BF_8)} \quad (14)$$

Based on Equation 14, the negative coefficients of  $BF_3$  and  $BF_8$  indicate that fulfillment of these conditions decreases the log-odds of ARI. Specifically, when  $BF_3 = 1$ , the logit decreases by 0.553, showing a substantial protective effect. When  $BF_8 = 1$ , the logit decreases by 0.117, indicating a milder, yet still beneficial effect.

### 3.3.3 Significance Testing of MARS Model

The following are the hypotheses formulated to test whether the set of basis functions significantly influences the response variable in the MARS model:

$H_0: a_3 = a_8 = 0$  (The model is not significant)

$H_1: \text{At least one } a_m \neq 0, m = 3 \text{ and } m = 8$  (the model is significant)

At a significance level of 5%, the critical region for the above hypotheses is: reject  $H_0$  if the test statistic  $F > F_{\text{critical}}$  or if the p-value  $< \alpha$ .

**Table 12:** Simultaneous Significance Test Results for the MARS Model

|                                 |                            |
|---------------------------------|----------------------------|
| F-Statistic                     | 36.826                     |
| Standard Error of Regression    | 0.351                      |
| P-value                         | $0.181332 \times 10^{-10}$ |
| Residual Sum of Squares (RSS)   | 8.110                      |
| (MDF, NDF)                      | (2, 66)                    |
| Regression Sum of Squares (SSR) | 9.050                      |

Based on Table 12, it can be seen that the F-statistic is 36.826, which is greater than  $F_{(2,66)} = 3.13592$ . In addition, the p-value of  $0.181332 \times 10^{-10}$  is less than 0.05. Therefore, we reject  $H_0$  and conclude that the model is statistically significant, indicating that at least one basis function with  $\beta \neq 0$  significantly influences the incidence of ARI among toddler at the Deleng Pokhkisen Public Health Center in Aceh Tenggara Regency.

The next step is to perform a partial significance test to examine whether each basis function has a significant effect on the response variable in the MARS model. The hypotheses used are as follows:

$H_0: a_m = 0$  (basis function  $B_m(x)$  has no effect on the model)

$H_1: a_m \neq 0, m = 3 \text{ and } 8$  (basis function  $B_m(x)$  has a significant effect on the model)

At a 5% significance level,  $H_0$  is rejected if the p-value  $< \alpha$ . The results of the partial significance tests for each basis function are as follows:

**Table 13:** Partial Significance Test Results for the MARS Model

| Parameter        | Estimate | Standard Error | t-Ratio | P-value                   |
|------------------|----------|----------------|---------|---------------------------|
| Constant         | 0.840    | 0.056          | 15.088  | $0.99201 \times 10^{-3}$  |
| Basis Function 3 | -0.553   | 0.099          | -5.565  | $0.514515 \times 10^{-6}$ |
| Basis Function 8 | -0.117   | 0.030          | -3.873  | $0.249783 \times 10^{-3}$ |

Based on Table 13, the results of the partial significance test for each basis function are as follows:

- a. For basis function 3, the test statistic is  $t = -5.565$  and the p-value is  $0.514515 \times 10^{-6}$ . Since the p-value is less than 0.05, we reject  $H_0$  and conclude that basis function 3 significantly affects the incidence of ARI among toddler at the Deleng Pookhisen Public Health Center in Aceh Tenggara Regency.
- b. For basis function 8, the test statistic is  $t = -3.873$  and the p-value is  $0.249783 \times 10^{-3}$ . Again, the p-value is below 0.05, leading us to reject  $H_0$  and conclude that basis function 8 also significantly affects the incidence of ARI in the same setting.

### 3.3.4 Predictor Variable Importance Level

To explain the extent of contribution of each predictor variable to the probability model of ARI (Acute Respiratory Infection) incidence in toddlers, the output of relative variable importance is viewed in terms of the importance level and cost of omission, as shown in Table 14 below.

**Table 14:** Relative Variable Importance Output

| Variable                 | Importance Level | Cost of Omission |
|--------------------------|------------------|------------------|
| Mother's Knowledge       | 100.000          | 0.201            |
| Household Density        | 100.000          | 0.201            |
| Mother's Education Level | 51.947           | 0.168            |
| Exclusive Breastfeeding  | 51.947           | 0.168            |
| Mother's Age             | 0.000            | 0.155            |
| Mother's Attitude        | 0.000            | 0.155            |
| House Ventilation        | 0.000            | 0.155            |

Based on Table 14, the following interpretation is obtained:

- a. The variables *mother's knowledge* and *household density* are the two most significant predictors in the model, each with an importance level of 100%, indicating that both provide the strongest influence on ARI incidence in toddlers. In addition, these two variables have a cost of omission of 0.201, meaning that their contribution is substantial—removing either of them from the model would significantly reduce its predictive power. Thus, these variables are crucial to retain in the model.
- b. The variables *mother's education level* and *exclusive breastfeeding* also contribute meaningfully, each with an importance level of 51.947%. Although not as influential as the previous two, both variables have a cost of omission of 0.168, indicating that omitting either would moderately increase model error. Even though their impact is less than that of *mother's knowledge* and *household density*, these variables remain important and should not be disregarded, as they help reduce the model's generalization error.
- c. The variables *mother's age*, *mother's attitude*, and *house ventilation* show no significant contribution to the model, as they have an importance level of 0%. However, they still have a cost of omission of 0.155. Therefore, removing these three variables from the model may be considered, as it would not significantly reduce the model's effectiveness.

### 3.3.5 Classification Results of MARS Model

The evaluation of a MARS model involves several performance metrics that indicate how well the model can distinguish between categories in the dependent variable. Table 15 presents the classification results of MARS model.

37  
Table 15: Classification Results of MARS Model

| Observed               | Predicted              |                    |
|------------------------|------------------------|--------------------|
|                        | Not Suffering from ARI | Suffering from ARI |
| Not Suffering from ARI | 20                     | 12                 |
| Suffering from ARI     | 1                      | 36                 |

351 3.4 Selection of the Best Regression Method

352 Based on Table 9 and Table 15, the accuracy, precision, recall, specificity, and F1-score are calculated  
353 to determine the best regression model between binary logistic regression and MARS. The comparison  
354 results are presented in Table 15 as follows:

Table 16: Comparison of Classification Accuracy between Logistic Regression and MARS

| Method                     | Accuracy | Precision | Recall | Specificity | F1-Score |
|----------------------------|----------|-----------|--------|-------------|----------|
| Binary Logistic Regression | 86.96    | 85.00     | 91.89  | 81.25       | 88.30    |
| MARS                       | 81.16    | 75.00     | 97.30  | 62.50       | 84.70    |

355 Based on Table 16, the Binary Logistic Regression model outperforms the MARS model in terms of  
356 accuracy, precision, specificity, and F1-score. Although MARS achieves a higher recall value (97.30%),  
357 it comes at the cost of lower specificity (62.50%), which suggests that the model is more prone to false  
358 positives. In contrast, the Binary Logistic Regression model maintains a better balance between sen-  
359 sitivity and specificity, making it a more reliable model for classifying ARI cases in toddlers overall.  
360 Therefore, Binary Logistic Regression is considered the better model in this case.

361 4 Conclusion

362 This study aimed to model the risk factors of Acute Respiratory Infections (ARI) in toddlers by com-  
363 paring Binary Logistic Regression and Multivariate Adaptive Regression Splines (MARS). The findings  
364 showed that both models were statistically significant and capable of identifying key predictors of ARI.  
365 However, logistic regression demonstrated superior classification performance, with higher accuracy  
366 (86.96%), precision (85.00%), specificity (81.25%), and F1-score (88.30%). In contrast, the MARS  
367 model achieved a higher recall (97.30%), indicating better sensitivity to actual positive cases but at the  
368 cost of lower specificity (62.50%), making it more prone to false positives. Overall, binary logistic  
369 regression emerged as the more reliable and interpretable method, offering a better balance between sen-  
370 sitivity and specificity that essential attribute for effective public health interventions and policymaking.  
371 Importantly, exclusive breastfeeding, home ventilation, and housing density emerged as significant  
372 predictors in the logistic regression model, while MARS highlighted the interaction effects between  
373 maternal knowledge and housing conditions, as well as between maternal education and breastfeeding.  
374 These findings offer new insights into how individual and environmental factors interact in influencing  
375 ARI incidence in toddlers.

376 While the study provides a robust comparison of parametric and non-parametric modeling approaches,  
377 it also has limitations. The dataset was relatively small and specific to one health center, which may affect  
378 the generalizability of the findings. Future research could extend this work using larger, more diverse  
379 datasets and explore hybrid or ensemble modeling techniques to further improve prediction accuracy.

380 CRediT Authorship Contribution Statement

381 Ardi Kurniawan: Conceptualization. Nathania Fauziah: Methodology, Formal Analysis. Arinda  
382 Mahadesyawardani: Software, Validation. Syifa' Azizah Putri Gunawan: Visualization, Data Cura-  
383 tion. Aurellia Calista Anggakusuma: Writing-Review & Editing.

## Declaration of Generative AI and AI-assisted technologies

AI-assisted technologies were used to assist in language refinement and initial draft structuring. All scientific content and interpretations were developed and validated by the authors.

## Declaration of Competing Interest

The authors declare no competing interests.

## Funding and Acknowledgments

This research received no external funding. The authors would like to express their sincere gratitude to Department of Mathematics, Faculty of Science Technology, Universitas Airlangga for the academic and institutional support provided throughout the research process.

## Data Availability

The dataset used in this study was obtained from the Master's thesis by Rahman Sabri (2019), titled "Faktor yang Memengaruhi Balita terhadap Penyakit ISPA di Puskesmas Deleng Pookhiksen Kabupaten Aceh Tenggara", submitted to the Master of Public Health Program, Faculty of Public Health, Helvetia Institute of Health, Medan, Indonesia<sup>1</sup>.

## References

- [1] X. Yu, H. Wang, S. Ma, W. Chen, L. Sun, and Z. Zou, "Estimating the global and regional burden of lower respiratory infections attributable to leading pathogens and the protective effectiveness of immunization programs," *International Journal of Infectious Diseases*, vol. 149, p. 107 268, 2024. DOI: [10.1016/j.ijid.2024.107268](https://doi.org/10.1016/j.ijid.2024.107268).
- [2] R. Mirino, D. Dary, and R. Tampubolon, "Identification of factors causing acute respiratory infection (ari) of under-fives in community health center work area in north jayapura sub-district," *Journal of Tropical Pharmacy and Chemistry*, vol. 6, no. 1, pp. 15–20, 2022. DOI: [10.30872/j.trop.pharm.chem.v6i1.214](https://doi.org/10.30872/j.trop.pharm.chem.v6i1.214).
- [3] H. Zhu, K. Huang, X. Han, *et al.*, "The burden of acute respiratory infection in children under 5 attributable to economic inequality in low- and middle-income countries," *BMJ Global Health*, vol. 10, no. 3, pp. 1–9, 2025. DOI: [10.1136/bmjgh-2024-017409](https://doi.org/10.1136/bmjgh-2024-017409).
- [4] S. Y. Nurjamillah and C. M. Dwiriani, "Status gizi dengan kejadian ispa balita di wilayah kerja puskesmas unyur kota serang sebelum dan selama pandemi covid-19," *Jurnal Ilmu Gizi dan Dietetik*, vol. 1, no. 2, pp. 95–102, 2022. DOI: [10.25182/jigd.2022.1.2.95-102](https://doi.org/10.25182/jigd.2022.1.2.95-102).
- [5] M. Irwansyah, S. Zuliansyah, and I. Hasan, "Sustainable landscape for high urban temperature mitigation in the disaster-prone coastal city of banda aceh, indonesia," in *IOP Conference Series: Earth and Environmental Science*, vol. 630, IOP Publishing, 2021, p. 012010. DOI: [10.1088/1755-1315/630/1/012010](https://doi.org/10.1088/1755-1315/630/1/012010).
- [6] R. Intan, "Relationship of the physical environment with the incidence of ari in toddlers," *Miracle Journal of Public Health*, vol. 4, no. 2, pp. 176–186, 2021. DOI: [10.36566/mjph/Vol4.Iss2/268](https://doi.org/10.36566/mjph/Vol4.Iss2/268).

<sup>1</sup><https://helvetia.ac.id/>

- 420 [7] E. Saputri, P. Eka Sudiarti, and Z. R. Z., "Hubungan kepadatan hunian kamar dan jenis bahan bakar  
421 memasak dengan kejadian ispa pada balita di desa pulau rambai wilayah kerja upt puskesmas  
422 kampa tahun 2023," *Jurnal Ners*, vol. 7, no. 2, pp. 20234–21841, 2023.
- 423 [8] A. Intiyati, R. D. Y. Putri, I. S. Edi, *et al.*, "Correlation between exclusive breastfeeding, com-  
424 plementary feeding, infectious disease with wasting among toddlers: A cross-sectional study,"  
425 *Amerta Nutrition*, vol. 8, no. 2SP, pp. 1–8, 2024. DOI: [10.20473/amnt.v8i2SP.2024.1-](https://doi.org/10.20473/amnt.v8i2SP.2024.1-8)  
426 [8](https://doi.org/10.20473/amnt.v8i2SP.2024.1-8).
- 427 [9] I. Yasmin, I. A. Pramesty, T. T. Affandi, and Y. Naldi, "Hubungan antara tingkat pengetahuan,  
428 tingkat pendidikan ibu, serta status gizi balita terhadap kejadian infeksi saluran pernapasan akut  
429 (ispa) pada balita di puskesmas kesunean kota Cirebon Jawa Barat," *Jurnal Kedokteran & Kese-*  
430 *hatan*, vol. 5, no. 1, pp. 1–8, 2024.
- 431 [10] E. T. Yuniarsih, M. Salam, M. H. Jamil, and A. N. Tenriawaru, "Determinants determining the  
432 adoption of technological innovation of urban farming: Employing binary logistic regression  
433 model in examining rogers' framework," *Journal of Open Innovation: Technology, Market, and*  
434 *Complexity*, vol. 10, no. 2, p. 100307, 2024. DOI: [10.1016/j.joitmc.2024.100307](https://doi.org/10.1016/j.joitmc.2024.100307).
- 435 [11] J. Sharma and S. K. Mitra, "Developing a used car pricing model applying multivariate adaptive  
436 regression splines approach," *Expert Systems with Applications*, vol. 236, no. 1, p. 121277, 2024.  
437 DOI: [10.1016/j.eswa.2023.121277](https://doi.org/10.1016/j.eswa.2023.121277).
- 438 [12] R. Sabri, I. Effendi, and N. Aini, "Factors affecting the level of ispa disease in children in deleng  
439 pokhkisen health center Aceh Tenggara district," *Contagion :Scientific Periodical of Public Health*  
440 *and Coastal Health*, vol. 1, no. 2, pp. 69–82, 2019. DOI: [10.30829/contagion.v1i2.](https://doi.org/10.30829/contagion.v1i2.6883)  
441 [6883](https://doi.org/10.30829/contagion.v1i2.6883).
- 442 [13] N. Chamidah, A. T. Hendrawan, F. S. Ardiyanto, M. S. Hammami, N. Izzah, and S. N. Hariadi,  
443 "Modeling hypertension disease risk in Indonesia using multivariate adaptive regression spline  
444 and binary logistic regression approaches," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*,  
445 vol. 18, no. 4, pp. 2217–2230, 2024. DOI: [10.30598/barekengvol18iss4pp2217-2230](https://doi.org/10.30598/barekengvol18iss4pp2217-2230).
- 446 [14] J. E. Lawn, Z. A. Bhutta, C. Ezeaka, and O. Saugstad, "Ending preventable neonatal deaths: Mul-  
447 ticountry evidence to inform accelerated progress to the sustainable development goal by 2030,"  
448 *Neonatology*, vol. 120, no. 4, pp. 491–499, 2023. DOI: [10.1159/000530496](https://doi.org/10.1159/000530496).
- 449 [15] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd. New York:  
450 John Wiley & Sons, 2013. DOI: [10.1002/9781118548387](https://doi.org/10.1002/9781118548387).
- 451 [16] J. Wilson, K. Lorenz, and L. Selby, "Introduction to binary logistic regression," in *Modeling*  
452 *Binary Correlated Responses*, ser. ICSA Book Series in Statistics, Cham: Springer, 2024. DOI:  
453 [10.1007/978-3-031-62427-8\\_1](https://doi.org/10.1007/978-3-031-62427-8_1).
- 454 [17] H. Dukalang and B. Otok, "Modified partial least square structural equation model with multi-  
455 variate adaptive regression spline: Parameter estimation technique and applications," *MethodsX*,  
456 vol. 14, p. 103381, 2025. DOI: [10.1016/j.mex.2025.103381](https://doi.org/10.1016/j.mex.2025.103381).
- 457 [18] M. B. Adiguzel and M. Cengiz, "Model selection in multivariate adaptive regressions splines  
458 (mars) using alternative information criteria," *Heliyon*, vol. 9, no. 9, e19964, 2023. DOI: [10.](https://doi.org/10.1016/j.heliyon.2023.e19964)  
459 [1016/j.heliyon.2023.e19964](https://doi.org/10.1016/j.heliyon.2023.e19964).
- 460 [19] S. Risambessy, S. Aulele, and F. Lembang, "Misclassification analysis of elementary school ac-  
461 creditation data in ambon city using multivariate adaptive regression spline," *Jurnal Matematika,*  
462 *Statistika Dan Komputasi*, vol. 18, no. 3, pp. 394–406, 2022. DOI: [10.20956/j.v18i3.](https://doi.org/10.20956/j.v18i3.19451)  
463 [19451](https://doi.org/10.20956/j.v18i3.19451).

## ORIGINALITY REPORT

18%

SIMILARITY INDEX

14%

INTERNET SOURCES

11%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

|   |                                                                                                                                                                                                                                    |      |
|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1 | <a href="https://ojs3.unpatti.ac.id">ojs3.unpatti.ac.id</a><br>Internet Source                                                                                                                                                     | 1 %  |
| 2 | <a href="https://ijeais.org">ijeais.org</a><br>Internet Source                                                                                                                                                                     | 1 %  |
| 3 | <a href="https://journal.universitasbumigora.ac.id">journal.universitasbumigora.ac.id</a><br>Internet Source                                                                                                                       | 1 %  |
| 4 | <a href="https://jurnal.uinsu.ac.id">jurnal.uinsu.ac.id</a><br>Internet Source                                                                                                                                                     | 1 %  |
| 5 | <a href="https://repository.poltekkes-manado.ac.id">repository.poltekkes-manado.ac.id</a><br>Internet Source                                                                                                                       | 1 %  |
| 6 | Ardi Kurniawan, Mochammad Firmansyah, Toha Saifudin. "MODELING CHRONIC FILARIASIS CASES IN WEST JAVA USING A MULTIVARIATE ADAPTIVE APPROACH REGRESSION SPLINES", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2024<br>Publication | <1 % |
| 7 | <a href="https://www.coursehero.com">www.coursehero.com</a><br>Internet Source                                                                                                                                                     | <1 % |
| 8 | Mohammed H., Abdel-Rahman Hedar, Taysir H., Yousef B.. "SS-SVM (3SVM): A New Classification Method for Hepatitis Disease Diagnosis", International Journal of Advanced Computer Science and Applications, 2013<br>Publication      | <1 % |
| 9 | <a href="https://garuda.kemdikbud.go.id">garuda.kemdikbud.go.id</a><br>Internet Source                                                                                                                                             |      |

<1 %

10 [www.mdpi.com](http://www.mdpi.com)  
Internet Source

<1 %

11 [medium.com](https://medium.com)  
Internet Source

<1 %

12 [ebin.pub](http://ebin.pub)  
Internet Source

<1 %

13 [tropmedhealth.biomedcentral.com](http://tropmedhealth.biomedcentral.com)  
Internet Source

<1 %

14 Pratyush Banerjee, Jatin Pandey, Manish Gupta. "People Analytics - Theory, Tools and Techniques", Routledge, 2025  
Publication

<1 %

15 [www.eea-esem.com](http://www.eea-esem.com)  
Internet Source

<1 %

16 [jurnal.univpgri-palembang.ac.id](http://jurnal.univpgri-palembang.ac.id)  
Internet Source

<1 %

17 [docobook.com](http://docobook.com)  
Internet Source

<1 %

18 [scholarworks.montana.edu](http://scholarworks.montana.edu)  
Internet Source

<1 %

19 [sciencedocbox.com](http://sciencedocbox.com)  
Internet Source

<1 %

20 [journals.bilpubgroup.com](http://journals.bilpubgroup.com)  
Internet Source

<1 %

21 [pure.ncue.edu.tw](http://pure.ncue.edu.tw)  
Internet Source

<1 %

22 [www.ukessays.com](http://www.ukessays.com)  
Internet Source

<1 %

|    |                                                                                                                                                                                                                                                                                                                  |      |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 23 | Submitted to University of Reading<br>Student Paper                                                                                                                                                                                                                                                              | <1 % |
| 24 | <a href="https://repository.helvetia.ac.id">repository.helvetia.ac.id</a><br>Internet Source                                                                                                                                                                                                                     | <1 % |
| 25 | Submitted to University College Dublin (UCD)<br>Student Paper                                                                                                                                                                                                                                                    | <1 % |
| 26 | Submitted to University of Glasgow<br>Student Paper                                                                                                                                                                                                                                                              | <1 % |
| 27 | Vieira, Margarida Jorge Cerqueira. "Classifying and Explaining Pornography Detection:An End-to-End Framework.", Universidade do Porto (Portugal)<br>Publication                                                                                                                                                  | <1 % |
| 28 | <a href="https://journals.ums.ac.id">journals.ums.ac.id</a><br>Internet Source                                                                                                                                                                                                                                   | <1 % |
| 29 | Abdelmonem Afifi, Susanne May, Virginia A. Clark. "Practical Multivariate Analysis", Chapman and Hall/CRC, 2019<br>Publication                                                                                                                                                                                   | <1 % |
| 30 | Mohanad S. AL-Musaylh, Ravinesh C. Deo, Yan Li, Jan F. Adamowski. "Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting", Applied Energy, 2018<br>Publication | <1 % |
| 31 | Muthmainah, Hanik Badriyah Hidayati, Budi Yanti. "Improving Health for Better Future Life: Strengthening from Basic Science to Clinical Research", CRC Press, 2023<br>Publication                                                                                                                                | <1 % |

32 Sarbani Biswas, In Kim. "A thorough review of phytogenic feed additives in non-ruminant nutrition: production, gut health, and environmental concerns", Journal of Animal Science and Technology, 2025

Publication

<1 %

33 Satria Bangsawan, MS Mahrinasari, Ernie Hendrawaty, Rindu Rika Gamayuni et al. "The Future Opportunities and Challenges of Business in Digital Era 4.0", Routledge, 2020

Publication

<1 %

34 Submitted to Trine University

Student Paper

<1 %

35 repository.mocu.ac.tz

Internet Source

<1 %

36 Cramer, Duncan. "EBOOK: ADVANCED QUANTITATIVE DATA ANALYSIS", EBOOK: ADVANCED QUANTITATIVE DATA ANALYSIS, 2003

Publication

<1 %

37 Yunjong Kim, Kawon Kang, Juneyoung Park, Cheol Oh. "A methodology for prioritizing safety indicators using individual vehicle trajectory data", Journal of Transportation Safety & Security, 2023

Publication

<1 %

38 amrsjournals.com

Internet Source

<1 %

39 journal.ipm2kpe.or.id

Internet Source

<1 %

40 Ruwayd Hussain Charfare, Aditya Uttam Desai, Nishad Nitin Keni, Aditya Suresh Nambiar, Mimi Mariam Cherian. "Smart

<1 %

Healthcare Framework: Real-Time Vital Monitoring and Personalized Diet and Fitness Recommendations Using IoT and Machine Learning", International Journal of Robotics and Control Systems, 2025

Publication

|    |                                                                                                                                                                                                                                   |      |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 41 | <a href="https://bioone.org">bioone.org</a><br>Internet Source                                                                                                                                                                    | <1 % |
| 42 | <a href="https://etd.aau.edu.et">etd.aau.edu.et</a><br>Internet Source                                                                                                                                                            | <1 % |
| 43 | H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in Healthcare Innovation", CRC Press, 2025<br>Publication                                                                                                | <1 % |
| 44 | <a href="https://ijrpr.com">ijrpr.com</a><br>Internet Source                                                                                                                                                                      | <1 % |
| 45 | <a href="https://rjoas.com">rjoas.com</a><br>Internet Source                                                                                                                                                                      | <1 % |
| 46 | <a href="https://www.gssrr.org">www.gssrr.org</a><br>Internet Source                                                                                                                                                              | <1 % |
| 47 | <a href="https://www.researchsquare.com">www.researchsquare.com</a><br>Internet Source                                                                                                                                            | <1 % |
| 48 | <a href="https://www.sciencegate.app">www.sciencegate.app</a><br>Internet Source                                                                                                                                                  | <1 % |
| 49 | Claudia F. Parvanta, Katherine K. Thomas, Khandaker S. Zaman. "Changing Nutrition Behavior in Bangladesh: Successful Adaptation of New Theories and Anthropological Methods ", Ecology of Food and Nutrition, 2007<br>Publication | <1 % |
| 50 | <a href="https://digital.library.unt.edu">digital.library.unt.edu</a><br>Internet Source                                                                                                                                          | <1 % |

|    |                                                                                                                                                                                                                                                          |      |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 51 | <a href="http://journal.multitechpublisher.com">journal.multitechpublisher.com</a><br>Internet Source                                                                                                                                                    | <1 % |
| 52 | <a href="http://journal.unhas.ac.id">journal.unhas.ac.id</a><br>Internet Source                                                                                                                                                                          | <1 % |
| 53 | <a href="http://journal.ypidathu.or.id">journal.ypidathu.or.id</a><br>Internet Source                                                                                                                                                                    | <1 % |
| 54 | Andrew S. Fullerton, Jun Xu. "Ordered Regression Models - Parallel, Partial, and Non-Parallel Alternatives", Chapman and Hall/CRC, 2019<br>Publication                                                                                                   | <1 % |
| 55 | Istianah Surury, Melikhatun Azizah, Nastiti Dyah Prastiwi. "Spatial Analysis of Acute Respiratory Infection (ARI) Based On the Air Pollution Standard Index (PSI) at DKI Jakarta Region in 2018-2019", JURNAL KESEHATAN LINGKUNGAN, 2022<br>Publication  | <1 % |
| 56 | Sandeep Gawdiya, Dinesh Kumar, Bulbul Ahmed, Ramandeep Kumar Sharma, Pankaj Das, Manoj Choudhary, Mohamed A. Mattar. "Field scale wheat yield prediction using ensemble machine learning techniques", Smart Agricultural Technology, 2024<br>Publication | <1 % |
| 57 | Submitted to Syntax Corporation<br>Student Paper                                                                                                                                                                                                         | <1 % |
| 58 | <a href="http://ajbasweb.com">ajbasweb.com</a><br>Internet Source                                                                                                                                                                                        | <1 % |
| 59 | <a href="http://aridbiom.yazd.ac.ir">aridbiom.yazd.ac.ir</a><br>Internet Source                                                                                                                                                                          | <1 % |
| 60 | <a href="http://civilejournal.org">civilejournal.org</a><br>Internet Source                                                                                                                                                                              |      |

<1 %

61

[dghs.portal.gov.bd](http://dghs.portal.gov.bd)

Internet Source

<1 %

62

[doaj.org](http://doaj.org)

Internet Source

<1 %

63

[dspace.daffodilvarsity.edu.bd:8080](http://dspace.daffodilvarsity.edu.bd:8080)

Internet Source

<1 %

64

[ir.ucc.edu.gh](http://ir.ucc.edu.gh)

Internet Source

<1 %

65

[jmi.rivierapublishing.id](http://jmi.rivierapublishing.id)

Internet Source

<1 %

66

[journals.plos.org](http://journals.plos.org)

Internet Source

<1 %

67

[jurnal.unprimdn.ac.id](http://jurnal.unprimdn.ac.id)

Internet Source

<1 %

68

[maxwellsci.com](http://maxwellsci.com)

Internet Source

<1 %

69

[newinera.com](http://newinera.com)

Internet Source

<1 %

70

[repository.ipb.ac.id](http://repository.ipb.ac.id)

Internet Source

<1 %

71

[www.atlantis-press.com](http://www.atlantis-press.com)

Internet Source

<1 %

72

Ade Gafar Abdullah, Isma Widiaty, Cep Ubad Abdullah. "Global Competitiveness: Business Transformation in the Digital Era", Routledge, 2019

Publication

<1 %

73

Ahmed A. Abd El-Latif, Mohammed A ElAffendi, Mohamed Ali AlShara, Yassine

<1 %

Maleh. "Cybersecurity, Cybercrimes, and Smart Emerging Technologies", CRC Press, 2025

Publication

74

Brady T. West, Steven G. Heeringa, Patricia A. Berglund. "Applied Survey Data Analysis", CRC Press, 2025

Publication

<1 %

75

E.T. Yuniarsih, Muslim Salam, Muhammad Hatta Jamil, A. Nixia Tenriawaru.  
"Determinants determining the adoption of technological innovation of urban farming: Employing binary logistic regression model in examining Rogers' framework", Journal of Open Innovation: Technology, Market, and Complexity, 2024

Publication

<1 %

76

Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On