Modeling the Percentage of Tuberculosis Cure in Indonesia Using a Multivariate Adaptive Regression Spline Approach

Toha Saifudin[1*], Dita Aris Novianti[2], Nadia Dwi Marwanda[3], dan Suliyanto[4]
[1,2,3,4]Statistics Program Study, Department of Mathematics, Universitas Airlangga, Surabaya, Indonesia
*Corresponding author: tohasaifudin@fst.unair.ac.id

Received: 3 May 2024
Revised: 19 June 2024
Accepted: 24 June 2024

ABSTRACT –Tuberculosis (TB) is an infectious disease caused by the bacterium Mycobacterium Tuberculosis. After India, Indonesia is the country with the second highest number of TB sufferers in the world. TB prevention efforts in Indonesia have been carried out, even since 1995. However, in general, 2006-2022 the TB cure in Indonesia tends to experience a downward trend. Therefore, it is important to know what variables have a significant effect and how the pattern relates to the percentage of TB cures. We urgently need this information to optimize TB handling efforts and achieve Sustainable Development Goals (SDGs) point 3, which focuses on good health and well-being. For that purpose, this study used the Multivariate Adaptive Regression Spline (MARS) approach. MARS is considered more flexible in overcoming cases of predictor variables that do not form a certain pattern to their response variables and can accommodate possible interactions between predictor variables. The best model was obtained at $BF = 18$, $MI = 2$, and $MO = 0$ with minimum GCV value is 37.053 and R^2 is 91.6%, with significant predictor variables are food management sites meet the requirements according to standards, complete treatment, smoking population over 15 years, families with healthy latrines, and districts/municipalities implement healthy living *germas* policy. The significance of the nine predictors should prioritize enhancing the quality of health services for example ensuring a fair distribution of complete treatment for TB patients.

Keywords–**Tuberculosis, Multivariate Adaptive Regression Spline (MARS), Nonparametric Regression**

I. INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by the bacterium *Mycobacterium tuberculosis*. Sources of infectious diseases can usually occur through the air (*airborne disease)* [1]. *Mycobacterium tuberculosis* dies quickly when exposed to direct sunlight, but can live several hours in dark, moist places. TB transmission often occurs when patients with BTA pulmonary TB are talking, sneezing or coughing and indirectly the patient releases sputum splashes in the air and there are ±3000 sputum splashes containing bacteria [2]. Home environmental factors, such as room ventilation, room humidity, sunlight, and frequency of room use, statistically correlate with the incidence of tuberculosis [3].

This disease became an epidemic the world. After India, Indonesia is around the second-highest number of pulmonary tuberculosis sufferers in the world in 2021. TB control efforts in Indonesia have been carried out, even since 1995. However, in general, in 2006-2022, the TB cure rate in Indonesia tends to experience a downward trend [4]. This information is urgently needed so that TB handling efforts can be carried out optimally to achieve the Sustainable Development Goals (SDGs) point 3, namely good health and well-being, which aims to ensure good health and well-being for all people at all ages. According to its strategic plan, the Indonesian government has committed to reducing the incidence of tuberculosis to 65 cases per 100,000 population by 2030. Tuberculosis control efforts in Indonesia between 2020-2024 aim to accelerate Indonesia's efforts to achieve the goal of tuberculosis elimination by 2030 as well as end the tuberculosis epidemic by 2050 [5]. Adherence to taking medication is a key factor determining the success of treatment to achieve recovery. Some patients in many countries stop treatment before treatment is complete for various reasons. Nonadherence rates are difficult to assess, but it is estimated that more than a quarter of TB patients do not complete treatment for 6 months. Non-adherence to treatment increases the risk of treatment failure and recurrence and is considered one of the most important causes of the emergence of drug-resistant tuberculosis. In particular, multidrug resistant TB (MDR-TB) and extensively resistant TB pose a serious threat to public health [6].

To find out which variables have a significant effect on tuberculosis cure concentration, one method that can be used is regression analysis, namely a non-parametric regression model. Because the relationship of response variables with predictors is not known with certainty, so it uses nonparametric concepts. One regression method with a nonparametric approach is the Multivariate Adaptive Regression Spline (MARS) method. MARS is considered more flexible in overcoming cases of predictor variables that do not form a certain pattern for their response variables and can accommodate possible interactions between predictor variables. Because it combines *truncated spline* and recursive separation spline regression [7].

In addition, research [8] on modeling factors that affect tuberculosis cure in East Java province with truncated spline nonparametric regression with variable results that affect tuberculosis cure rates are management sites meet the requirements according to standards, complete treatment, smoking population over 15 years, families with healthy latrines, and districts/municipalities implement a healthy living *germas* policy. The study had a weakness in the lack of detail about truncated spline nonparametric regression analysis methods, which could result in complexity of interpretation of the results. Furthermore, the presence of unidentified variables and variable measurement problems may impact the reliability of the findings. Spline methods are also prone to overfitting and require careful parameter settings to produce a well-interpretable model.

From the reference above and the purpose of this research analysis, researchers want to know what factors have a significant influence on the percentage of tuberculosis cure in 2022 in Indonesia with a MARS approach. The results of this study are expected to be a reference for the central government to make policies related to the factors that influence the number of tuberculosis cures in the hope that it can improve tuberculosis cures in Indonesia and it is also expected that this study can provide learning and understanding for readers.

II. LITERATURE REVIEW

A. Nonparametric Regression

Regression analysis is a method used to find patterns of relationships between the response variable (y) and predictor variables (x) [9]. This relationship pattern is usually estimated through data visualization using scatter plots. If the scatter plot does not show a clear pattern of relationships, such as linear, quadratic, or other patterns then a nonparametric regression approach can be used. In this approach, no particular form is assumed for regression functions, so the method is more flexible in capturing complex patterns of relationships between these variables [10]. Variable y and variable x with n observations can be expressed generally as a nonparametric regression model which can be seen in the general model in equation (1) :

$$y = f(x_i) + \varepsilon_i, \quad i = 1,2,\dots,n \tag{1}$$

With

$f(x_i)$ = Assumed smooth

ε_i = Random error that follows a normal distribution with zero mean and variance σ^2 or $(\varepsilon_i \sim IIDN(0, \sigma^2))$

B. Multivariate Adaptive Regression Splines (MARS)

MARS were introduced by Friedman in 1991 [11]. MARS is a type of nonparametric regression model that assumes no specific form for the relationship between bound variables and independent variables. MARS has great flexibility in the form of functions produced. The main focus of MARS is to deal with problems in high dimensions, large samples, and number of variables, which often require complex calculations based on the smallest values of Generalized Cross Validation (GCV) [12]. The MARS method can produce more accurate prediction models for continuous and categorical binary response variables, by considering several predictor variables in the form of piecewise regression.

Piecewise regression is a type of segmented regression, where if a regression line is unable to explain the entire data, several regression lines will be used to adjust all data derived from independent variables. The points at which the pattern of change occurs are referred to as knots [13]. The MARS model is the result of a complex combination of spline and recursive partitioning techniques [14]. The main goal is to produce continuous estimation of regression functions along knot points in the data. MARS uses the spline shape as smoothing. The placement and number of knot points are automatically determined by the data through modification of the recursive partitioning algorithm [15]. MARS modeling is shown in the equation (2).

$$f(x_i) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [S_{km}(X_v(k,m) - t_{km})]_+ + \varepsilon_i \tag{2}$$
$$y_i = a_0 + \sum_{m=1}^{M} a_m B_m(x) + \varepsilon_i$$

where

a_0 = The constant (intercept)

a_m = The coefficient of BF_m

S_{km} = +1 (knots to the right of the subset) and −1 (knots to the left of the subset)

K_m = Degrees of interaction

$X_v(k,m)$ = Predictor variables to-v, sorting to-k subsets to-m

t_{km} = Knot value of the variable $X_v(k,m)$

M =The number of basis function

$B_m(x)$ = $\prod_{k=1}^{K_m} [S_{km}(X_v(k,m) - t_{km})]_+$

Equation (2) can also be written in matrix form, which is contained in equation (3) :

$$\boldsymbol{y} = \boldsymbol{Ba} + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$, $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} S_{1m}(X_1(1,m) - t_{1m}) & \cdots & \prod_{k=1}^{K_M} S_{Mm}(X_1(M,m) - t_{Mm}) \\ 1 & \prod_{k=1}^{K_1} S_{1m}(X_2(1,m) - t_{1m}) & \cdots & \prod_{k=1}^{K_M} S_{Mm}(X_2(M,m) - t_{Mm}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} S_{1m}(X_n(1,m) - t_{1m}) & \cdots & \prod_{k=1}^{K_M} S_{Mm}(X_n(M,m) - t_{Mm}) \end{bmatrix}$

C. Determination of the Best Model

The determination of the best model in the MARS method can be seen in the minimum GCV value. The determination uses forward and backward stepwise algorithms. The formula of GCV is in equation (5) [16, 17].

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{y}_i]^2}{[1 - Trace(\boldsymbol{B}(\boldsymbol{B^T B})^{-1}\boldsymbol{B^T})/N]^2} \tag{4}$$

where
N = Number of observation
y_i = Actual value of observation i
$\hat{y}_i$ = Fitted value of observation i

III. METHODOLOGY

A. Data Source

This study used secondary data taken from "Profil Kementerian Kesehatan Indonesia year 2022" and "Badan Pusat Statistik Indonesia". The data used are the percentage of cure for TB and the factors that are thought to affect it. There are 34 provinces in Indonesia as observation units.

B. Research Variables

The variables in this study are found in **Table 1** [18].

Table 1 Research Variables

Variable	Variable Name	Unit
Y	TB Cure	Percentage
X_1	Food Management Sites Meet the Requirements According to Standards	Percentage
X_2	Households Have Proper Sanitation	Percentage
X_3	Complete Treatment	Percentage
X_4	Health Insurance Ownership	Percentage
X_5	Smoking Population Over 15 Years	Percentage
X_6	*Puskesmas* According to Health Type Standards	Percentage
X_7	Households Have Access to Adequate Drinking Water	Percentage
X_8	Families with Healthy Latrines	Percentage
X_9	Districts/Municipalities Implement Healthy Living *Germas* Policy	Percentage

C. Research Steps

Measures to analyze factors thought to affect tuberculosis cure:

1. Describe the percentage of TB cure in Indonesia and factors thought to affect it.
2. Modelling the percentage of TB cure in Indonesia with the MARS method using MARS software, as for the following steps.
   a. Prepare response and predictor variables data with sav format.
   b. Import data into MARS software.
   c. Specifies the Basis Function (BF) value, which is between 18 and 45.
   d. Specifies the Maximum Interaction (MI), which is one, two, and three.
   e. Specifies the Minimum Observation (MO), which is zero, one, two, and three.
   f. Combine predefined BF, MI, and MO values. Then compare the GCV value, the minimum GCV is the best model.
   g. Perform simultaneous tests on all BF coefficients with F test statistics, to find out whether together the BF affects the response variables [17].
   h. Perform a partial test on each BF coefficient with t test statistics, to find out whether each BF has a significant influence on the response variables [17].
3. Interpretation of the best model that has been tested significantly on response variables by inferring the results obtained and explaining the meaning contained in the values of BF.

IV. RESULTS AND DISCUSSIONS

A. Descriptive Statistics

To find out an overview of the percentage of TB cure in Indonesia and the factors that allegedly affect it is to descriptive the data. The results of the data description can be seen at **Table 2**.

Table 1 Descriptive Research Data

Variable	Mean	Standard Deviation	Minimum	Maximum
Y	43.04	9.77	24.80	67.80
X_1	58.14	12.25	34.80	80.60
X_2	81.00	9.78	40.34	96.21
X_3	61.85	7.26	42.30	75.60
X_4	72.49	10.35	55.91	97.50
X_5	26.76	3.74	17.91	33.81
X_6	55.34	22.92	7.60	100.00
X_7	87.47	7.82	65.39	98.42
X_8	86.59	11.17	54.70	100.00
X_9	70.06	32.81	0.00	100.00

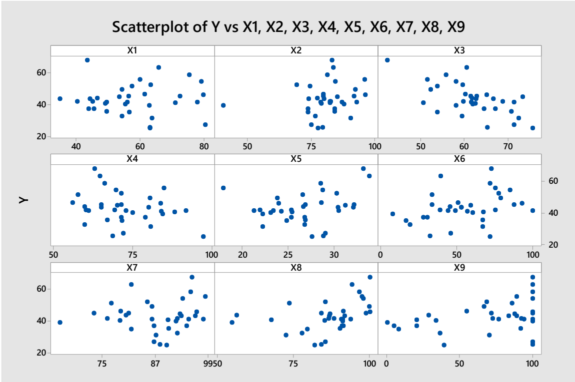

Figure 1 *Scatter Plot* of Variables $\boldsymbol{Y}$ Against Each Variables $\boldsymbol{X}$

At **Figure 1** indicates that scatter plots between variables Y and each variable X does not have a clear pattern of relationships. Therefore, a nonparametric regression method is used to analyze factors that are thought to affect TB cure in Indonesia.

B. MARS Model Estimation

In MARS software, testing is carried out by determining the BF, MI, and MO values. The BF value entered are $18, 27, 36$, and 45 then the MI value used are $1, 2, 3$, and for MO are $0, 1, 2$, and 3. The best combination was obtained on $BF = 18, MI = 2$, and $MO = 0$. This is based on the minimum GCV Value, namely 37.053, MSE value is 9.754, and R^2 is 91.6%. So that estimates are obtained as in **Table 3.**

Table 2 Estimation of t-Test Parameters and Values from the Basis Function

Basis Functions	Model	Estimation	$t_{statistics}$
Constant	−	$\mathbf{61.915}$	−
BF_1	$\max(0, X_3 - 42.3)$	-0.87	-10.263
BF_2	$\max(0, X_5 - 26.86)$	−	−
BF_6	$\max(0, X_9 - 88.9)$	-4.84	-7.771
BF_7	$\max(0, 88.9 - X_9)$	-0.176	-6.665
BF_8	$\max(0, X_8 - 54.7) * BF_6$	0.113	7.054
BF_9	$\max(0, X_1 - 34.8) * BF_2$	0.096	7.489
BF_{11}	$\max(0, 56.0 - X_1) * BF_7$	0.012	4.007

Based on **Table 3** the MARS model was obtained to estimate the percentage of TB cure in Indonesia which can be seen in equation (5):

$$\begin{aligned}\hat{y} &= 61.915 - 0.870BF_1 - 4.840BF_6 - 0.176BF_7 + 0.113BF_8 + 0.096BF_9 + 0.012BF_{11} \\ &= 61.915 - 0.870(\max(0, X_3 - 42.3)) - 4.840(\max(0, X_9 - 88.9)) - 0.176(\max(0, 88.9 - X_9)) \\ &\quad + 0.113\left(\max(0, X_8 - 54.7) * (\max(0, X_9 - 88.9))\right) \\ &\quad + 0.096\left(\max(0, X_1 - 34.8) * (\max(0, X_5 - 26.86))\right) \\ &\quad + 0.012\left(\max(0, 56.0 - X_1) * (\max(0, 88.9 - X_9))\right)\end{aligned} \quad (5)$$

Based on the MARS model obtained in equation (5), there is an interaction between predictor variables. These interactions are found in BF_8, BF_9, and BF_{11}. In BF_8 there is an interaction between variables X_8 and X_9, in BF_9there is an interaction between variables X_1 and X_5, while in BF_{11} there is an interaction between variables X_1 and X_9.

Furthermore, the basis function coefficient test is carried out which includes simultaneous tests and individual tests.

1. Simultaneous Tests of MARS Model Basis Function Coefficients

   Simultaneous or concurrent testing aims to determine whether in general the selected MARS model is a suitable model. The hypothesis used is as follows:

   $H_0: a_1 = a_6 = a_7 = a_8 = a_9 = a_{11} = 0$

   H_1: There is at least one $q_m \neq 0$; $m = 1,6,7,8,9,11$

   Based on the results of the MARS test, it can be seen that the value of $F_{statistic}$ is 49.271 by using $\alpha = 0.05$ so obtained $F_{(0,05;6.27)} = 2.459$. Because $F_{statistic} > F_{(0,05;6.27)}$, So the decision on taken is to reject H_0 which means that there is at least one that is not equal to zero so that it can be said that the model obtained is appropriate and shows the exact relationship between the predictor and response variables.

2. Individual Test of MARS Model Basis Function Coefficients

   Individual tests aim to find out whether the model containing these parameters has been able to describe the actual state of the data. The hypothesis used is as follows:

   $H_0: a_m = 0, m = 1,6,7,8,9,11$

   $H_1: a_m \neq 0, m = 1,6,7,8,9,11$

Based on **Table 3** obtained the value of $|t_{statistic}|$ by using $\alpha = 0.05$ then obtained value of $t(\frac{\alpha}{2}, v) = t_{(0.025;28)} = 2{,}048$. Because of all values of $|t_{statistic}| > t_{(0.025;28)}$, So the decision taken is reject H_0, which means that there is a relationship between all base functions and response variables

Obtained from the selection of the best model with a minimum GCV value, it can be seen that the predictor variables included in the model are five variables, namely where food management meets the requirements according to the standard (X_1), complete treatment (X_3), smoking population over 15 years (X_5), Families with healthy latrines (X_8), and districts/cities carry out healthy living *germas* policies (X_9) with the importance of the predictor variables contained in the model described as in **Table 4**.

Table 3 Importance of Predictor Variables

Variables	Importance	GCV
X_3	100.0%	128.862
X_5	69.113%	80.906
X_8	64.074%	74.745
X_1	57.933%	67.866
X_9	48.877%	58.985
X_2	0.00%	37.053
X_4	0.00%	37.053
X_6	0.00%	37.053
X_7	0.00%	37.053

Based on **Table 4** then it can be seen that the complete treatment variable (X_3)is the most important variable in the model with an importance level of 100%, if the variable X_3 is included in the GCV model it will decrease by 128.862. Then followed by the variable population smoking over 15 years (X_5) with an importance of 69.113%, if variable X_5 included in the GCV model will be reduced by 80.906, and so on. Other variables have no importance (0%) because it is already represented by the previous five variables.

C. Analysis and Interpretation of the MARS Model

After obtaining the estimation of the following equation (5) model, a plot between Y and $\hat{Y}$ is presented in **Figure 2:**

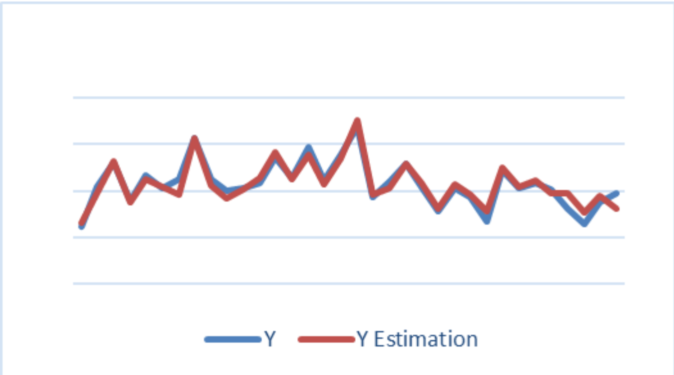

Figure 2 Plot of Actual Value VS Fitted Value

The following is given the interpretation of the MARS model contained in equation (5) as follows:

1. $BF_1 = \max(0, X_3 - 42.3)$ with a coefficient of -0.87. It means BF_1 will be meaningful for provincies with a percentage of Complete Treatment (X_3) more than 42.3%, hence every increase of BF_1 one unit will reduce the TB cure percentage by 0.87% assuming the other BF is constant.
2. $BF_6 = \max(0, X_9 - 88.9)$ with a coefficient of -4.84. It means BF_6 will be meaningful for provincies with a percentage of districts/municipalities implement healthy living *germas* policy (X_9) more than 88.9%, hence every increase of BF_6 one unit will reduce the TB cure percentage by 4.84% assuming the other BF is constant.
3. $BF_7 = \max(0, 88.9 - X_9)$ with a coefficient of -0.176. It means BF_7 will be meaningful for provincies with a percentage of districts/municipalities implement healthy living *germas* policy (X_9) less than 88.9%, hence every drop BF_7 one unit will reduce the TB cure percentage by 0.176% assuming the other BF is constant.
4. $BF_8 = \max(0, X_8 - 54.7) * BF_6 = \max(0, X_8 - 54.7) * (0, X_9 - 88.9)$ with a coefficient of 0.113. It means BF_8 will be meaningful for provinces with a percentage of families with healthy latrines (X_8) more than 54.7% and percentage of districts/municipalities implement healthy living *germas* policy (X_9) more than 88.9%, hence every increase of BF_8 one unit will increase the TB cure percentage by 0.113% assuming the other BF is constant. In BF_8, there is an interaction between the variables X_8 and X_9.
5. $BF_9 = \max(0, X_1 - 34.8) * BF_2 = \max(0, X_1 - 34.8) * (X_5 - 26.8)$ with a coefficient of 0.0096%. It means BF_9 will be meaningful for percentage provinces with a food management sites meet the requirements according to standards (X_1) more than 34.8% and percentage of smoking population over 15 years (X_5) more than 26.8%,

hence every increase of BF_9 one unit will increase the TB cure percentage by 0.0096% assuming the other BF is constant. In BF_9, there is an interaction between the variables X_1 and X_5.

6. $BF_{11} = \max(56.0 - X_1) * BF_7 = \max(56.0 - X_1) * (0{,}88.9 - X_9)$ with a coefficient of 0.012%. It means BF_9 will be meaningful for percentage of provinces with a food management sites meet the requirements according to standards (X_1) less than 56% and percentage of districts/municipalities implement healthy living *germas* policy (X_9) less than 88.9%. hence every increase of BF_{11} one unit will increase the TB cure percentage by 0.012% assuming the other BF is constant. In BF_{11}, there is an interaction between the variables X_1 and X_9.

V. CONCLUSIONS AND SUGGESTIONS

A. Conclusion

Based on the results of the analysis and discussion, the conclusions obtained from this study are as follows:

1. In 2022, Indonesia has an average tuberculosis cure percentage rate of 46.5%, where the highest percentage of cure in West Nusa Tenggara Province, which is 67.80% and the lowest cure percentage rate is in Aceh Province, which is 24.80%.
2. The best model obtained using the MARS method is $BF = 18, MI = 2$, and $MO = 0$ with a minimum GCV value is 37.053, the Mean Square Error (MSE) value is 9.754, and R^2 value is 91.6%.
3. The results of modeling using the MARS method produced five variables that had a significant effect on the percentage of tuberculosis cure in Indonesia, namely: food management sites meet the requirements according to standards (X_1), complete treatment (X_3), smoking population over 15 years (X_5), families with healthy latrines (X_8), and districts/municipalities implement healthy living *germas* policy (X_9).
4. The variable of complete treatment variable was the most important variable in the model with an importance of 100%, Then followed by the variable of smoking population over 15 years with an importance level of 69.113%, next is the variable of families with a healthy latrine with an importance level of 64.074%, for the variable of food management sites meet the requirements according to standard with an importance level of 57.933%, and the last variable is districts/municipalities implement healthy living *germas* policy with an important level of 48.877%. Other variables have no importance (0%) because it is already represented by the previous five variables.

B. Suggestions

Based on the conclusions that have been obtained, the suggestions that can be given are as follows:

1. Based on the level of importance of the nine predictor variables, it is expected to be a recommendation for the Health Institutions of each province in Indonesia that prioritize enhancing the quality of health services for example ensuring a fair distribution of complete treatment for TB patients.
2. Furthermore, researchers are expected to be able to compare other methods such as nonparametric methods spatially in order to obtain models according to the circumstances carried out or socio-geographical sufferers.

ACKNOWLEDGMENT

We are very grateful to Universitas Airlangga, Faculty of Science and Technology, especially S1 Statistics Study Program for the opportunity given to implement the knowledge we have gained and create this journal. Not to forget, thanks to the accompanying lecturers, namely Dr. Toha Saifudin, M. Si. and Drs. Suliyanto, M. Si. who have provided valuable guidance. All their support and assistance is greatly appreciated in this process.

REFERENCES

[1] R. Miggiano, M. Rizzi and D. M. Ferraris, "Myctobacterium Tuberculosis Pathogenesis, Infection Preventation an Treatment," *Pathogens*, vol. 9, no. 5, p. 385, 2020.

[2] Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan, Pedoman Nasional Pengendalian Tuberkulosis, Jakarta: Kementerian Kesehatan Republik Indonesia, 2020.

[3] I. Yuniar, A. Wahyono and H. Purnomo, "Relationship of House Building Materials, Lighting and Occupational Density to the Incidence of Tuberculosis," in *Proceedings of the 3rd Borobudur International Symposium on Humanities and Social Science 2021*, vol. 3, Magelang, Indonesia, Atlantis Press, 222, pp. 378-382.

[4] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, "Laporan Program Penanggulangan Tuberkulosis Tahun 2021," Kementerian Indonesia Republik Indonesia, Jakarta, 2022.

[5] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, Strategi Nasional Penanggulangan Tuberkulosis di Indonesia 2020-2024, Jakarta: Kementerian Kesehatan Republik Indonesia, 2020.

[6] Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan, Pedoman Nasional Pengendalian Tuberkulosis, Jakarta: Kementerian Kesehatan Republik Indonesia, 2014.

[7] J. H. Friedman, "Multivariate Adaptive Regression Splines," *The Annals od Statistics*, vol. 19, no. 1, pp. 1-67, 1991.

[8] N. Evitasari., S. Handajani. and H. Pratiwi, "Pemodelan Faktor-Faktor yang Memengaruhi Kesembuhan Tuberkulosis di Provinsi Jawa Timur dengan Regresi Nonparametrik Spline Truncated," in *ProSandika Universitas Pekalongan*, Pekalongan, 2022.

[9] D. F, Shabrina, N. A, T. M, H. N and Wahyono, "Prediction of Indonesia Inflation Rate Using Regression Model Based on Genetic Algorithm," *Journal Online Informatika (JOIN),* vol. 5, no. 1, pp. 45-52, 2020.

[10] P. Cizek and S. Sadikoglu, "Robust Nonparametric Regression: A Review," *Wiley Interdiscipinary Reviews: Computational Statistics,* vol. 12, no. 3, pp. 1-16, 2019.

[11] M. S. Abed, F. J. Kadhim, J. K. Almusawi, H. Imran, L. F. A. Bernardo and S. N. Henedy, "Untilizing Multivariate Adaptive Regression Splines (MARS) for Precise Estimation of Soil Compaction Parameters," *Applied Sciences,* vol. 13, no. 21, 2023.

[12] S. Celik, A. Bakoglu and M. I. Catal, "Examination of the Relationship Among Plant Characteristics Affecting Yield in Pea Plants with MARS Algorithm," *Journal of Animal & Plant Sciences,* vol. 31, no. 6, pp. 1622-1631, 2021.

[13] H. Kuswanto, "Multilevel Spline Linear Model in Nonparametric Regression using the Method Restricted Maximum Likelihood," in *Doctoral Disertation, Universitas Hasanuddin*, Makassar, 2022.

[14] S. D. P. Yasmirullah, B. W. Otok, J. D. T. Purnomo and D. D. Prastyo, "Multivariate Adaptive Regression Spline (MARS) Methods with Application to Multi Drug Resistance Tuberculosis (MDR-TB) Prevalence," in *AIP Conference Proceedings*, Surabaya, 2021.

[15] J. Friedman. and C. Roose, "An Introduction to Multivariate Adaptive Regression Splines," *Statistical Methods in Medical Research,* vol. 4, pp. 197-217, 1995.

[16] R. Biswas, B. Rai, P. Samui and S. S. Roy, "Estimating Concrete Compressive Strength Using MARS, LSSVM and GP," *Engineering Journal,* vol. 24, no. 2, pp. 41-52, 2020.

[17] S. Rissambesy, S. Aulele and F. Lembang, "Misclassiification Analysis of Elementary School Accreditation Data in Ambon City Using Multivariate Adaptive Regression Spline," *Jurnal Matematika, Statistik, dan Komputasi,* vol. 18, no. 3, pp. 394-406, 2022.

[18] Sekretariat Jenderal, Profil Kesehatan Indonesia tahun 2022, Jakarta: Kementerian Kesehatan Republik Indonesia, 2023.