Parameter Estimation of Multivariate Adaptive Regression Spline (MARS) with Stepwise Approach to Multi Drug-Resistant Tuberculosis (MDR-TB) Modeling in Lamongan Regency

S D P Yasmirullah¹, B W Otok¹*, J D T Purnomo¹, and D D Prastyo¹

¹Department of Statistics, Faculty of Mathematics, Computation, and Data Science, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

* e-mail: dr.otok.bw@gmail.com

Abstract. Tuberculosis (TB) is an infectious disease, caused by mycobacterium tuberculosis that affects various organs, especially the lungs. If TB treatment is not done thoroughly by the patient, then it can lead to death. Tuberculosis disease could not be healed cause TB bacteria to double immunity against anti-TB drugs, called multi-drug resistant (MDR). One identification issue toward TB infection chain is a TB case distribution analysis using mathematical modeling. The method used in this study was Multivariate Adaptive Regression Spline (MARS). Multivariate Adaptive Regression Spline is one type of non-parametric regression techniques, where the model does not assume the functional relationship between response and predictor variables, and has a flexible functional structure as well. Modeling aims to determine the factors that have the most significant influence on MDR-TB cases in Lamongan regency, as well as predict the incidence of MDR-TB in each sub-regency. The results show, the best model has a combination of BF = 28, MI = 2 and MO = 3, based on the minimum GCV, which is 5.26E-06. Furthermore, the model is statistically proper according to the criteria of APER and Press’s Q.

Keywords: Multivariate, Regression Spline, Tuberculosis

1. Introduction
Tuberculosis (TB) is a contagious infectious disease that can attack various organs, especially the lungs. The cause of TB is Mycobacterium Tuberculosis, which can lead to death if left untreated or treatment is incomplete [1]. Tuberculosis patient can be a source of TB transmission to the surrounding environment because Mycobacterium Tuberculosis is transmitted, through droplets in the air. While coughing or sneezing, TB sufferers spread bacteria into the air in the form of sputum droplets or droplet nuclei, which one cough can produce about 3000 sputum splashes. Tuberculosis disease is not successfully cured, causing the multi-drug TB resistance to anti-TB drugs, namely multi-drug resistant (MDR). Multi-drug resistant tuberculosis is one type of TB bacillary resistance to at least two anti-TB drugs, which is isoniazid and rifampicin [2]. The multi-drug resistant tuberculosis problem further aggravates the condition of TB disease and impedes TB control programs in Indonesia. This situation will eventually lead to a TB epidemic that is difficult to handle.

Tuberculosis is one of the main problems of public health, in the world and Indonesia. The WHO report in 2015 estimated that 480,000 people were suffering from MDR-TB and 190,000 of them died. Roughly 3.3% is estimated MDR-TB cases originate by new TB cases, and 20% of TB cases recur into
MDR-TB. On the other hand, TB is an opportunistic infection. It is infections that occur more often or are more severe in people with weakened immune systems than in people with healthy immune systems. HIV is a kind of disease that attacks the immune system, so people with HIV are more likely to get sick with other infections and diseases. Furthermore, HIV increases the risk of TB in people with HIV. The problem of TB especially MDR-TB not only at the national level but is also faced, by Lamongan Regency. Losses caused by TB were not only from the health but also from the socio-economic, it was being estimated an adult TB patient will lose an average working time of 3 to 4 months, this results in an annual loss of household income around 20-30% [3]. Therefore it is a necessary need to serious way to control TB problems. However, the presence of MDR-TB is a new challenge in TB control programs due to the determination of difficult diagnoses, high rates of treatment failure, and death. MDR-TB patient treatment is more difficult, a success rate of only about 50% with expensive treatment costs, even up to 100 times more expensive than treatment for TB without MD. For developing countries, MDR-TB has become a very burden in its solution [4].

Indonesia, as a developing country, is the third-largest contributor to TB disease and has sought various ways to solve with TB cases and the many costs involved in controlling TB cases. One effort to identify the TB transmission chain is the analysis of the distribution of TB cases by modeling in the form of a mathematical model so that through the mathematical model it can explain the pattern of relationships between response variables and predictors. The pattern of relationships between these variables can be illustrated by regression analysis, which can be predicted by three approaches, which are parametric, semi-parametric, and non-parametric. If there is a known curve pattern assumption, then the parametric regression approach is used. However, not all relationship patterns can be approached with a parametric approach because there is no information about the relationship, between the response, and predictor variables. If the assumptions of the parametric model are not satisfied, then the regression curve can be estimated using the non-parametric regression approach. Non-parametric regression has high flexibility in estimating the regression curve. Data under nonparametric regression are expected to be able to find the pattern of a regression curve that have unknown pattern [5].

Non-parametric regression has an adaptive approach, for instance regression tree, Recursive Partitioning Regression (RPR), and Multivariate Adaptive Regression Spline (MARS). Multivariate Adaptive Regression Spline is one type of non-parametric regression approaches that was introduced by Friedman in 1991. Multivariate Adaptive Regression Spline model does not assume the functional relationship between response and predictor variables and has a flexible, functional structure. Many researchers have been developing MARS and its application, i.e. Multivariate Adaptive Regression Spline (MARS) approach to the grouping of season zones [6], bi-responses nonparametric regression model using MARS and its properties [7], modeling of welfare indicators in Java island using biresponses MARS [8], and parameter estimation and statistical test in multivariate adaptive generalized Poisson regression spline [9]. The purpose of modeling with the MARS method is to predict response variables based on several predictor variables and identify the factors that most influence the response variable. Several basis function coefficients established MARS model that entirely is controlled by regression data. The MARS model is useful for overcoming the problem of high dimensional data known as the curse of dimensionality and producing accurate response predictions. Furthermore, overcoming the weaknesses of Recursive Partitioning Regression (RPR), which is results in a continuous model on knots, based on the Generalized Cross-Validation (GCV) minimum value.

2. Literature

2.1. Multivariate Adaptive Regression Spline
Multivariate Adaptive Regression Spline (MARS) is an approach for multivariate nonparametric regression developed by Friedman (1991), where there is an assumption that the functional relationship between the response and predictor variables is unknown, and has a flexible functional structure. This method aims to overcome the weakness of Recursive Partition Regression (RPR), where the resulting model is not continuous on knots. Determination of knots on MARS is not sought individually from
these combinations but by an adaptive process. The adaptive process in MARS is done using a stepwise algorithm, namely backward and forward. Therefore, MARS is a method used to solve regression and classification problems to predict response variables based on several predictor variables. Multivariate Adaptive Regression Spline can also be used to overcome high-dimensional data that is data with predictor variables \( X_v \) as much 3 \( \leq v \leq 20 \) [10].

The MARS model is a complex combination of the spline method and recursive partitioning, hence, it can be generated an estimation of the continuous regression function. If \( Y \) is a categorical variable with \( m \) predictor variables, \( \mathbf{X} = (x_1, \ldots, x_m) \), then according to Friedman [10] the MARS model generally be written as follow:

\[
\text{logit } \pi(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = a_0 + \sum_{i=1}^{M} a_i \prod_{k=1}^{K_m} [s_{km}(x_{v(i,k)}) - t_{km}] 
\]

where,
- \( a_0 \): main basis function
- \( a_i \): coefficient of \( m \)th basis function
- \( M \): number of maximum basis function
- \( K_m \): degree of interaction in the \( m \)th basis function
- \( s_{km} \): value of 1 or -1 if the data is to the right or left of the knots
- \( x_{v(i,k)} \): \( v \)th predictor variable
- \( t_{km} \): value of knot from predictor variable \( x_{v(i,k)} \)

The matrix form of the MARS model in equation (1) is

\[
\text{logit } \pi(x) = \mathbf{Y} = \mathbf{B}\mathbf{a}
\]

The probability function can be expressed by,

\[
\pi(x) = \sigma_L \left( \mathbf{B}^T \mathbf{a} \right)
\]

If \( Y \sim \text{Bernoulli} \left( \sigma_L \left( \mathbf{B} \mathbf{a} \right) \right) \), \( i = 1, \ldots, n \), stated in the model \( Y = \sigma_L \left( \mathbf{B} \mathbf{a} \right) + e \), then \( e = Y - \sigma_L \left( \mathbf{B} \mathbf{a} \right) \).

Function of \( \sigma_L \) applied to vector \( z = (z_1, \ldots, z_n)^T \in \mathbb{R}^n \) so it can be stated as \( \sigma_L(z) = (\sigma_L(z_1), \ldots, \sigma_L(z_n))^T \) [11].

**Theorem 1**: Using the MARS model and logit model in equation (1) and (2), if \( \mathbf{B} \) is non-singular matrix and smoothing parameter \( \eta > 0 \), then the first derivation of estimator \( \hat{\mathbf{a}} \) using the penalized log-likelihood method, as follow [11]:

\[
\mathbf{B}^T \left[ \mathbf{y} - \sigma_L \left( \mathbf{B} \hat{\mathbf{a}} \right) \right] + n\eta \mathbf{D}\hat{\mathbf{a}} = 0
\]

If smoothing parameter \( \eta = 0 \), then first derivation of estimator \( \hat{\mathbf{a}} \) as follow [11]:

\[
\mathbf{B}^T \left[ \mathbf{y} - \sigma_L \left( \mathbf{B} \hat{\mathbf{a}} \right) \right] = 0
\]

According to Friedman [10], in selecting the MARS model it is necessary to pay attention to the generalized cross-validation (GCV) value of the model, the model that has the lowest GCV value among other models is the best.
\[ GCV(M) = \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \hat{f}_M(x_i) \right]^2 \]
\[ \frac{\left( \text{trace} \left( B (B' B)^{-1} B' \right) + 1 \right)}{N} + dM \]

where,
\( y_i \): response variable
\( x_i \): predictor variable
\( N \): number of observation
\( \hat{f}_M(x_i) \): estimation value of response variable on M basis function at \( x_i \)
\( M \): maximum number of basis function
\( B \): matrix of \( M \) basis function
\( d \): value when each basis function reaches optimization (2 ≤ \( d \) ≤ 4)

Evaluation of classification procedures is an evaluation that looks at the possibility of misclassification by a classification function [12]. The classification accuracy of the grouping results is calculated by the Apparent Error Rate (APER).

Table 1. MARS Model Classification.

| Actual Class | Predicted Class |
|--------------|-----------------|
| Group 1      | \( n_1 \)       |
| Group 2      | \( n_{21} \)    |
|              | \( n_2 \)       |
|              | \( n_{22} \)    |

According to Johnson and Wichern (1992), APER value is calculated by the following formula [12]:

\[ \text{APER} \% = \frac{n_{22} + n_{21}}{n_1 + n_2 + n_{11} + n_{22}} \times 100\% \] (7)

The statistic test used to determine the separation of the classification matrix compared to a chance model is called Press’s Q. The Press's Q value is then compared with the Chi-Square value for 1 degree of freedom at the desired confidence level. If Press's Q is greater than the critical value, then the classification matrix has been considered statistically better than chance [13].

\[ \text{Press's Q} = \frac{\left[ N - (nG) \right]^2}{N(G-1)} \] (8)

where,
\( N \): Total sample size
\( n \): Number of observation correctly classified
\( G \): Number of groups

2.2. Tuberculosis

Tuberculosis is a directly infectious disease caused by TB bacteria, which is Mycobacterium tuberculosis. Mycobacterium tuberculosis attacks mainly the lungs called pulmonary tuberculosis. When attacking organs other than the lung is called extra-pulmonary tuberculosis [14]. Mycobacterium tuberculosis is rod-shaped, measuring 1-4 microns long and 0.3-0.6 microns thick. Tuberculosis bacteria die quickly with direct sunlight but can survive several hours in a dark and damp place. In body tissues,
these bacteria can become dormant within a few years [14]. In figure 1 explains the risk of Mycobacterium tuberculosis in the air [3].

Figure 1. Basic Concepts of Pathogenesis from Tuberculosis Epidemiology.

The incidence of drug resistance has increased since the introduction of the first TB treatment in 1943. Multidrug-resistant tuberculosis is a case of tuberculosis that is resistant to at least two kinds of anti-TB drugs, which are isoniazid and rifampicin or accompanied by other first-line of anti-TB drugs, such as Pyrazinamide, Ethambutol, and Streptomycin [15]. Isoniazid and rifampicin are two important drugs in TB treatment applied in the Directly Observed Treatment Short-course (DOTS) strategy.

3. Research Methodology

3.1. Research Variable

This study used secondary data obtained the Health Profile of Lamongan Regency in 2017. The observation unit used in this study was MDR-TB incidence each sub-regency in the Lamongan Regency. The research variables were shown in Table 2.

| Code | Variables                                      | Scale   |
|------|-----------------------------------------------|---------|
| Y    | Incidence MDR-TB (0 = No, 1=Yes)              | Nominal |
| X₁   | Population density (people/km²)               | Ratio   |
| X₂   | HIV/AIDS prevalence (per 10,000 population)   | Ratio   |
| X₃   | Percentage of households with PHBS (%)        | Ratio   |
| X₄   | Percentage of healthy homes (%)               | Ratio   |
| X₅   | Ratio of primary health facilities (per 10,000 population) | Ratio |
| X₆   | Ratio of health workers (per 10,000 population) | Ratio |
| X₇   | Percentage of the population enrolled in school (%) | Ratio |

3.2. Analysis Method

The steps of analysis in this study are as follows:

1. Descriptive statistical analysis and data exploration of incidence MDR-TB in Lamongan Regency.
2. Parameter estimation MARS model
   a. Determine the possibility of the maximum number of basis functions (BF)
   b. Determine the maximum number of interactions
   c. Determine the minimum observation between knots by trial and error
   d. Determine the best model based on the minimum GCV value
3. Interpreting the best models
4. Testing the accuracy of the model using APER and calculating the stability of the classification with Press’s Q
5. Draw the conclusions and suggestions.

4. Results and Discussion

4.1. Characteristics of Multi-Drug Resistant Tuberculosis (MDR-TB)
The problems faced by Indonesia as a developing country are health, one of which is Tuberculosis (TB). Moreover, TB cases that have not been successfully cured cause another problem, namely multi-drug resistant (MDR). The MDR-TB problem further aggravates the condition of TB disease and impedes TB control programs in Indonesia. This situation will eventually lead to a TB epidemic that is difficult to handle. Lamongan is one of the regencies in East Java Province that faces MDR-TB problems. In 2017 the number of TB cases in Lamongan regency was 2237, of which 26 were MDR-TB. MDR-TB patients spread in several sub-regency in Lamongan Regency it was seen in Figure 2.

![Figure 2. Incidence MDR-TB in Lamongan Regency](image)
The green color indicates MDR-TB incidence in a sub-regency, which are Paciran, Brondong, Kalitengah, Karanggeneng, Turi, Lamongan, Dermolemahbang, Deket, Glagah, Sukodadi, Pucuk, Babat, Sugio, Kedungpring and Sambeng. Furthermore, a general description of the predictor variables has known through descriptive statistics shown by Table 3.

| Variables                        | Mean  | Variance  | Min  | Max  |
|----------------------------------|-------|-----------|------|------|
| Population density               | 743.90| 157314.80 | 245.50| 2053.60|
| HIV/AIDS prevalence              | 3.13  | 1.63      | 0.74 | 5.70 |
| Percentage of households with PHBS| 76.27 | 192.65    | 33.90| 98.30|
| Percentage of healthy homes      | 88.70 | 68.15     | 69.44| 99.55|
| Ratio of primary health facilities| 20.50 | 34.71     | 9.31 | 32.97|
| Ratio of health workers          | 12.13 | 22.62     | 1.00 | 21.08|
| Percentage of the population enrolled in school | 88.20 | 58.76 | 73.03 | 98.95 |
The information obtained based on Table 3 is the average population density of each sub-region in Lamongan Regency of 743.90 people/km², where Paciran has the highest density of 2053.60 people/km². The average HIV risk in Lamongan regency is 3.13 per 10,000 population, with a minimum figure of 0.74 in Brondong and a maximum figure of 5.70 in Demolemahbang. Furthermore, the average percentage of PHBS households in Lamongan Regency is 76.27% and the average percentage of healthy homes of 88.70%. The percentage of the population enrolled in school in Lamongan Regency has an average of 88.20%. Then, the average ratio of primary health facilities in Lamongan Regency is 12.13 and the average ratio of health workers which is 12.13.

4.2. Multi-Drug Resistant Tuberculosis (MDR-TB) Modeling using MARS

MDR-TB incidence modeling in Lamongan regency had used MARS approach. The combined result of BF, MI, and MO in this research have shown in Table 4.

| BF | MI | MO | GCV | R² | BF | MI | MO | GCV | R² | BF | MI | MO | GCV | R² |
|----|----|----|-----|----|----|----|----|-----|----|----|----|----|-----|----|
| 14 | 1  | 0  | 7.83E-02 | 0.682831 | 21 | 1  | 0  | 5.55E-02 | 0.775259 | 28 | 1  | 0  | 4.12E-02 | 0.833081 |
| 14 | 1  | 1  | 3.73E-02 | 0.848902 | 21 | 1  | 1  | 9.33E-03 | 0.9622  | 28 | 1  | 1  | 1.91E-03 | 0.992263 |
| 14 | 1  | 2  | 3.90E-02 | 0.842023 | 21 | 1  | 2  | 1.17E-02 | 0.952709 | 28 | 1  | 2  | 6.96E-04 | 0.997179 |
| 14 | 1  | 3  | 3.29E-02 | 0.866638 | 21 | 1  | 3  | 1.27E-02 | 0.948729 | 28 | 1  | 3  | 4.02E-03 | 0.983701 |
| 14 | 1  | 5  | 3.96E-02 | 0.839791 | 21 | 1  | 5  | 1.81E-02 | 0.926578 | 28 | 1  | 5  | 1.22E-03 | 0.995065 |
| 14 | 1  | 10 | 3.96E-02 | 0.667022 | 21 | 1  | 10 | 7.78E-02 | 0.684812 | 28 | 1  | 10 | 7.78E-02 | 0.684812 |
| 14 | 2  | 0  | 5.15E-02 | 0.791307 | 21 | 2  | 0  | 1.05E-02 | 0.957419 | 28 | 2  | 0  | 1.59E-03 | 0.993549 |
| 14 | 2  | 1  | 2.31E-02 | 0.906279 | 21 | 2  | 1  | 2.18E-03 | 0.99118 | 28 | 2  | 1  | 3.86E-05 | 0.999844 |
| 14 | 2  | 2  | 3.27E-02 | 0.867445 | 21 | 2  | 2  | 8.32E-04 | 0.996631 | 28 | 2  | 2  | 5.10E-05 | 0.999793 |
| 14 | 2  | 3  | 2.37E-02 | 0.90409  | 21 | 2  | 3  | 1.15E-03 | 0.995331 | 28 | 2  | 3  | 5.26E-06* | 0.999979 |
| 14 | 2  | 5  | 3.53E-02 | 0.856971 | 21 | 2  | 5  | 8.40E-03 | 0.965978 | 28 | 2  | 5  | 2.27E-03 | 0.990794 |
| 14 | 2  | 10 | 6.67E-02 | 0.729797 | 21 | 2  | 10 | 2.52E-02 | 0.898086 | 28 | 2  | 10 | 4.97E-03 | 0.979859 |
| 14 | 3  | 0  | 5.15E-02 | 0.791307 | 21 | 3  | 0  | 1.05E-02 | 0.957419 | 28 | 3  | 0  | 1.59E-03 | 0.993549 |
| 14 | 3  | 1  | 2.31E-02 | 0.906279 | 21 | 3  | 1  | 2.18E-03 | 0.99118 | 28 | 3  | 1  | 2.43E-05 | 0.999901 |
| 14 | 3  | 2  | 3.27E-02 | 0.867445 | 21 | 3  | 2  | 8.32E-04 | 0.996631 | 28 | 3  | 2  | 5.10E-05 | 0.999793 |
| 14 | 3  | 3  | 2.37E-02 | 0.90409  | 21 | 3  | 3  | 1.15E-03 | 0.995331 | 28 | 3  | 3  | 6.41E-05 | 0.99974 |
| 14 | 3  | 5  | 3.53E-02 | 0.856971 | 21 | 3  | 5  | 8.40E-03 | 0.965978 | 28 | 3  | 5  | 2.27E-03 | 0.990794 |
| 14 | 3  | 10 | 6.67E-02 | 0.729797 | 21 | 3  | 10 | 2.52E-02 | 0.898086 | 28 | 3  | 10 | 5.29E-03 | 0.978589 |

The combination of BF, MI, and MO in Table 4 shows that the best MARS model has BF = 28, MI = 2 and MO = 3 with GCV values of 5.26E-06. The best MARS models are:

\[
\hat{f}(x) = 0.265 + 0.001BF_1 + 0.005BF_2 - 0.002BF_3 - 0.006BF_4 + 0.208BF_5 + 0.527BF_6 + 0.107BF_7 + 0.014BF_8 - 0.063BF_9 + 0.283BF_{10} - 0.0001BF_{11} - 0.00004BF_{12} + 0.0007BF_{13} + 0.0002BF_{14} - 0.069BF_{15} - 0.794BF_{16} + 0.036BF_{17} + 0.256BF_{18} + 0.002BF_{19} - 0.002BF_{20}
\]

Interpretation for some basis function of the best model.

- BF₆ = h(x_2 - 3.36289)

Coefficient BF₆ will statistically significant if HIV / AIDS prevalence is more than 3.36289. Each BF₆ increase of one unit with the other is considered constant, then an increase in MDR-TB is exp(0.527) = 1.694.
• $BF_{10} = h(x_7 - 97.0902)$

An increase in $BF_{10}$ by one unit can cause an increase in MDR-TB by $\exp(0.283) = 1.327$, assuming the other $BF$ is constant. The coefficient $BF_{10}$ will be significant when the percentage of the population enrolled in school is more than 97.0902%.

• $BF_{18} = h(x_8 - 3.36289) \times h(88.7467 - x_7)$

If $BF_{18}$ has increased by one unit and another $BF$ is considered constant, then the incidence of MDR-TB will increase by $\exp(0.256) = 1.292$. The coefficient $BF_{18}$ statistically significant when the HIV / AIDS prevalence is more than 3.36289 and the percentage of the population enrolled in school is less than 88.7467%.

4.3. Evaluation of Classification Procedures

The level of classification accuracy from the model predictions with MARS to classify MDR-TB incidence are determined, by the APER and Press’s Q values. The results of the classification of MDR-TB has shown in Table 5.

| Actual Class | Predicted Class | MDR-TB | Non MDR-TB |
|--------------|----------------|--------|------------|
| MDR-TB       | 15             | 0      | 0          |
| Non MDR-TB   | 0              | 12     | 0          |

According to equation (7) the APER value calculation result is 0%, which is that the probability of misclassification had done by the classification function is 0. Furthermore, the Press's Q value based on equation 8 obtained 27, which is this value is greater than $\chi^2_{0.05,1} = 3.841$, so the classification matrix has been considered statistically better than the chance model.

5. Conclusion

The best model based on the minimum GCV value, which is 5.26E-06, is a model with a combination of BF = 28, MI = 2 and MO = 3,

$$j(x) = 0.265 + 0.001BF_1 + 0.005BF_2 - 0.002BF_3 - 0.006BF_4 + 0.208BF_5 + 0.527BF_6 + 0.107BF_7 + 0.014BF_8 - 0.063BF_9 + 0.283BF_{10} - 0.001BF_{11} - 0.00004BF_{12} + 0.0007BF_{13} + 0.0002BF_{14} - 0.069BF_{15} - 0.794BF_{16} + 0.036BF_{17} + 0.256BF_{18} + 0.002BF_{19} - 0.002BF_{20}$$

This model has a value of APER = 0 and Press's Q = 27. Based on the criteria of APER and Press’s Q, it concludes that the model is statistically good. The biggest contribution has given by $BF_6 = h(x_7 - 3.6289)$, which is $\exp(0.527) = 1.694$, so the significant predictor variable is HIV/AIDS prevalence $(x_7)$. It shows that HIV/AIDS patients will increase the possibility of MDR-TB. There were also some recommendations to be given; i.e. (i) local government together with the health department should be able to reduce HIV / AIDS prevalence and MDR-TB, especially in areas prone to HIV / AIDS, (ii) conduct bivariate and multivariate analysis of MDR-TB data, and (iii) it was necessary to develop MARS method with other approaches.
Acknowledgement
We would like to express our gratitude to the Ministry of Research and Technology/National Research and Innovation Agency of Republic Indonesia (RISTEK-BRIN) for the financial support of this research. The authors would also like to thank the anonymous reviewers for their valuable comments and suggestions for improving the quality of this paper.

References
[1] Kemenkes. 2016. *Pusat Data dan Informasi Kementerian Kesehatan RI Tuberkulosis Temukan Obati Sampai Sembuh.* (Jakarta: Ministry of Health).
[2] Mulyanto, H. 2014. Hubungan Lima Indikator Perilaku Hidup Bersih dan Sehat dengan Tuberkulosis Multidrug Resistant. *Jurnal Berkala Epidemiologi.* 2, 355-367.
[3] Makalew, L.A., Kuntoro, Otok, B.W., Soenarnatalina, M., & Layuk, S. 2019. Modeling the Number of Cases of Tuberculosis Sensitive Drugs (Tbsd) in East Java using Geographically Weighted Poisson Regression (GWPR). *Indian Journal of Public Health Research & Development.* 10, 398-403.
[4] Kemenkes. 2012. *Pedoman Manajerial Pelayanan Tuberkulosis dengan Strategi DOTS di Rumah Sakit.* (Jakarta: Direktorat Bina Upaya Kesehatan).
[5] Eubank, R. 1999. *Nonparametric Regression and Spline Smoothing. Second Edition.* (New York: Marcel Dekker, Inc.).
[6] Otok, B. W. 2010. Pendekatan Multivariate Adaptive Regression Spline (MARS) pada Pengelompokkan Zona Musim Suatu Wilayah. *Statistika.* 10, 107 – 120.
[7] Ampulembang, A. P., Otok, B. W., Rumiati, A. T., & Budiasih. 2015. Bi-Responses Nonparametric Regression Model Using MARS and Its Properties. *Applied Mathematical Sciences.* 9, 1417 – 1427.
[8] Ampulembang, A. P., Otok, B. W., Rumiati, A. T., & Budiasih. 2015. Modeling of Welfare Indicators in Java Island Using Biresponses MARS. *International Journal of Mathematics and Statistics.* 54, 66 – 75.
[9] Hidayati, S., Otok, B. W., & Purhari. 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* 546, 052051
[10] Friedman, J. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics.* 19, 1-141
[11] Otok, B. W. 2008. *Pendekatan Bootstrap pada Model Multivariate Adaptive Regression Splines (MARS).* (Yogyakarta: Universitas Gadjah Mada).
[12] Johnson, R., & Wichern, D. 1992. *Applied Multivariate Statistical Analysis 6thEdition.* (New Jersey: Prentice Hall International Inc.).
[13] Hair, J. H., Anderson, R. E., Babin, B. J., & Black, W. C. 2010. *Multivariate Data Analysis 7th Edition.* (UK: Pearson).
[14] Depkes. 2002. *Pedoman Nasional Penanggulangan Tuberkulosis.* (Jakarta: Departemen Kesehatan Republik Indonesia).
[15] Widyasrini, E., Probandari, A., & Reviuno. 2017. Factors Affecting the Success of Multidrug Resistance (MDR-TB) Tuberculosis Treatment in Residential Surakarta. *Journal of Epidemiology and Public Health.* 2, 45-57.