Check 13

by Sifri Sifriyani

AIP Conference Proceedings

RESEARCH ARTICLE | MAY 09 2023

Application of nonparametric truncated spline regression on infant mortality rate in Kalimantan

Sifriyani; Rahmat Hidayat

AIP Conference Proceedings 2480, 030004 (2023)
https://doi.org/10.1063/5.0104011

Application of Nonparametric Truncated Spline Regression on Infant Mortality Rate in Kalimantan

Sifriyani[1*], Rahmat Hidayat[2]

[1]*Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia*
2 Mathematics Department, Cokroaminoto University, Palopo, Indonesia

**sifriyani@fmipa.unmul.ac.id*

Abstract. Infant Mortality Rate (IMR) reflects the level of health development of a country, as well as the quality of life of its people. This rate is used to monitor and evaluate the population and health policies of countries around the world. The infant mortality rate in Kalimantan in 2020 reached an average of 23 deaths per 1000 births, which was quite high when compared to the infant mortality rate in Indonesia: 21 deaths per 1000 births. The research objective was to determine the factors that influence the causes of increased infant mortality in Kalimantan by using nonparametric truncated spline regression model. The results showed that the best model was nonparametric truncated spline regression with a polynomial order of 2 and a number of knot points of 2. The criteria for the best model had an R^2 value of 88% and AIC of 0.056. The factors that influence the increase in infant mortality in Kalimantan were the rate of economic growth, the households that do not have a clean and healthy lifestyle, the intensity of exclusive breastfeeding, and the childbirth process that was not assisted by medical personnel. This research is useful for developing health and quality of life for people in Kalimantan, as well as increasing knowledge innovation in statistical methods.

Keywords: Infant Mortality Rate, Nonparametric Regression Model, Truncated Spline

INTRODUCTION

Regression analysis is used to determine the pattern of the relationship between the predictor variables and the response variables. Regression analysis consists of three approaches in estimating the model, namely parametric regression, nonparametric regression, and semiparametric regression. Parametric regression is used when the pattern of the relationship between the predictor variable and the response variable (the shape of the regression curve) is known, while semiparametric regression is used when the relationship pattern between the predictor variable and the response variable is partially known. If the pattern of the relationship between the predictor variable and the response variable is unknown, then the nonparametric regression approach is a solution to solve this problem[1].

Infant Mortality Rate (IMR) reflects the level of health development of a country, as well as the quality of life of its people. The infant mortality rate in Kalimantan in 2020 reached 23 deaths per 1000 births which was quite high when compared to the infant mortality rate in Indonesia, which was 21 deaths per 1000 births[2]. The government has always made efforts to reduce the number of infant deaths on the island of Kalimantan every year. In order for the right efforts, it is necessary to identify the factors causing infant mortality.

Based on the description on this background, we applied nonparametric truncated spline regression to the data on infant mortality rate in Kalimantan. The model is used to identify the factors that lead to an increase in infant mortality

The 4th International Conference on Science and Technology (ICST) 2021
AIP Conf. Proc. 2480, 030004-1–030004-10; https://doi.org/10.1063/5.0104011
Published by AIP Publishing. 978-0-7354-4463-8/$30.00

rates in Kalimantan. This research is expected to be useful for the development of health and quality of life for people in Kalimantan as well as to increase knowledge innovation in statistical methods.

LITERATURE REVIEW

Nonparametric Regression

Nonparametric regression is a technique used to overcome difficulties in parametric regression where the function form of the regression curve f must be known[1][2]. The nonparametric regression approach to estimate regression curves has the objective of providing a method that can be used to explore the relationship between two variables and providing predictions for observations that cannot be made without a reference to a particular parametric model. The nonparametric regression approach is used if the pattern is not known for the shape of the regression curve or there is no complete past information about the shape of the data pattern[4]. Visually, this type of regression takes the form of an unidentifiable pattern. The nonparametric regression approach has high flexibility because the data is expected to find its own form of regression curve estimation without being influenced by the researcher's subjectivity factors[1].

Given data (x_i, y_i) where the relationship between x_i and y_i is assumed to follow a nonparametric regression model as follows:

$$y_i = f(x_i) + \varepsilon_i,\ i = 1,2,\ldots,n \tag{1}$$

Where ε_i is assumed independent and normally distributed with mean zero and variance σ^2. Furthermore, $f(x_i)$ is a regression curve which is assumed to be smooth in the sense that it is a member of a certain function space (2). The shape of regression curve $f(x_i)$ is assumed to be unknown and is contained in the Sobolev space $W_2^m[a,b]$ especially $f(x_i) \in W_2^m[a,b]$ with

$$W_2^m[a,b] = \left\{ f : f, f^{(1)}, f^{(2)}, \ldots, f^{(m-1)} \text{absolute continues on } [a,b], \int_a^b \left[f^{(m)}(x)\right]^2 dx < \infty \right\} \tag{2}$$

for a segmented m, this provides better flexibility properties than ordinary polynomials and this property allows the model to adapt more effectively to the local characteristics of the data.

Truncated Spline In Nonparametric Regression

In nonparametric regression modeling, one of the approaches used to estimate the regression curve is the truncated spline approach. Truncated spline in nonparametric regression is a model that has good statistical and visual interpretations. In addition, truncated spline has high flexibility. Truncated spline is a segmented polynomial model that provides better flexibility properties than ordinary polynomials and this property allows the model to adapt more effectively to the local characteristics of the data[1][4].

Truncated spline of order m (3) with knots point $K_1, K_2, \ldots, K_r$ with $a < K_1 < \ldots < K_r < b$ is a function as follows:

$$f(x_i) = \sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h} (x_i - K_h)_+^m, \tag{3}$$

which β_k, β_{m+h} are real constants with $k = 0,1,2,\ldots,m$ and $h = 1,2,\ldots,r$ then the truncated function:

$$(x_i - K_h)_+^m = \begin{cases} (x_i - K_h)^m, & x_i \geq K_h \\ 0, & x_i < K_h \end{cases} \tag{4}$$

Based on the definition of a truncated spline, it is obtained::

(i) The function f is the slice of polynomial with degree m on the subinterval $[K_h, K_{h+1}]$.

(ii) The function f has a continuous derivative of the level m - 1

(iii) $f^{(m)}$ is a ladder function with jump points $K_1, K_2, \ldots, K_r$.

The nonparametric regression model given in Equation (5) with $f(x)$ is a regression curve which is approximated by a truncated spline function.

$$y_i = \sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h} (x_i - K_h)_+^m + \varepsilon_i,\ \ i = 1,2,\ldots,n \tag{5}$$

where y_i is the i-th response variable, x_i is the i-th predictor variable, β_k, β_{m+h} are real contants, $K_1, K_2, \ldots, K_r$ are knot points and ε_i is random error assumed to be identical, independent, and normally distributed with mean zero and variance σ^2 [1][3].

Estimation of Nonparametric Truncated Spline Regression Model

The estimation of nonparametric truncated spline regression model can be performed by using the Maximum Likelihood Estimator (MLE). Given Equation (6) is assumed $\varepsilon_i \sim N(0, \sigma^2)$ and Y_i is normally distributed:

$$y_i \sim N\left(\sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h}(x_i - K_h)_+^m, \sigma^2\right) \tag{6}$$

The first step in the MLE method is to form the probability density function (7) from $y_1, y_2, \ldots, y_n$ as follows:

$$f(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2)$$

$$= \prod_{i=1}^{n}\left[(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \left(\sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h}(x_i - K_h)_+^m\right)\right)^2\right)\right] \tag{7}$$

Next, the likelihood function is given in Equation (8)

$$L\,(\boldsymbol{\beta}, \sigma^2|Y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} T\right) \tag{8}$$

with the component T (9)

$$T = \sum_{i=1}^{n}\left(y_i - \left(\sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h}(x_i - K_h)_+^m\right)\right)^2 \tag{9}$$

Then the natural logarithmic operation is performed to facilitate mathematical operations and to obtain the parameter estimators $\boldsymbol{\beta}$ and σ^2 obtained from the results of the partial derivatives in Equation (7). Furthermore, the optimal knot point is determined to obtain the best nonparametric truncated spline regression model.

The spline estimator in nonparametric regression can handle smooth functions[5]. The estimation of the nonparametric truncated spline regression is obtained using the maximum likelihood estimation.

$$f(y|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - f(x_i))^2}{2\sigma^2}\right], \boldsymbol{\beta} > 0,\ \sigma^2 > 0 \tag{10}$$

The joint density function is given in Equation (11).

$$f(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2\sigma^2}(y_i - f(x_i))^2\right)}{(2\pi\sigma^2)^{\frac{1}{2}}} \tag{11}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f(x_i))^2\right)$$

As a result, the likelihood function is obtained in Equation (12).

$$L(\beta, \sigma^2 | y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f(x_i))^2\right) \tag{12}$$

The point estimate for the function f is obtained by solving the likelihood optimization in Equation (13).

$$MaxL(\beta, \sigma^2 | y) = \underset{\beta}{Max}\left\{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f(x_i))^2\right)\right\} \tag{13}$$

If a logarithmic transformation is taken, Equation (14) is obtained.

$$LnL(\beta,\sigma^2 \mid y) = Ln\left\{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f(x_i))^2\right)\right\}$$
$$= -\frac{n}{2}Ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f(x_i))^2 \quad (14)$$
$$= -\frac{n}{2}Ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \left(\sum_{k=0}^{p}\beta_k x_i^k + \sum_{h=1}^{r}\beta_{h+m}(x_i - k_h)_+^m\right)\right)$$
$$LnL(\beta,\sigma^2 \mid y) = -\frac{n}{2}Ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[\left(\mathbf{y} - \mathbf{X}(\mathbf{k})\beta\right)^T(\mathbf{y} - \mathbf{X}(\mathbf{k})\beta)\right]$$

Therefore, the likelihood estimation $\hat{\beta}(k)$ for nonparametric truncated spline regression is shown in Equation (15).

$$\hat{\beta}(k) = (\mathbf{X}(\mathbf{k})^{\mathrm{T}}\mathbf{X}(\mathbf{k}))^{-1}\mathbf{X}(\mathbf{k})^{\mathrm{T}}\mathbf{y} \quad (15)$$

Where $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ is the response variable, $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_p, \beta_{p+1}, ..., \beta_{p+m}]^T$ is a vector of the known parameters, $\mathbf{X}[\mathbf{k}] = \mathbf{X}[k_1, k_2, ..., k_p]$ is a matrix sized $n \times (p+m+1)$ with $[k_1, k_2, ..., k_p]$.

Selection of Optimal Knot Points

Nonparametric truncated spline regression has the advantage of having knot points in the model. With the existence of knot points, the model will follow the relationship pattern according to the data behavior. However, the number of knot points also affects the complexity of the model with the number of parameters used, therefore, a precise method is needed to determine the optimal knot point. One method that is widely used in selecting optimal knot points is Generalized Cross Validation (16). The optimal knot point is indicated by the smallest GCV value[3][6].

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1,K_2,\dots,K_r)}{(n^{-1}trace[I-A(K_1,K_2,\dots,K_r)])^2} \quad (16)$$

where

$$MSE(K_1, K_2, \dots, K_r) = n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 \quad (17)$$

Simultaneous and Partial Significance Test of Parameters

The hypothesis used in the simultaneous significance test[7]:

$H_0 : \beta_1 = \dots = \beta_p = \delta_1 = \dots = \delta_p = 0$

H_1 : There is at least one $\beta_k = \delta_k \neq 0; k = 1,2,\dots,p$

and the test statistics used are given by:

$$F_{value} = \frac{MS_{regression}}{MS_{residual}} \quad (18)$$

The solution to Equation (18) is obtained by completing the table of analysis of variance.

TABLE 1. Analysis of Variance (ANOVA)

Source of Variation	degree of freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F_{value}
Regression	p(1+r)	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$\frac{SS_{regression}}{df_{regression}}$	$\frac{MS_{regression}}{MS_{residual}}$

Residual	n-(p+r))-1	$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$\frac{SS_{residual}}{df_{residual}}$
Total	n-1	$\sum_{i=1}^{n}(y_i-\bar{y})^2$	

Partial significance test of parameters are as follows[8]:

$H_0: \beta_k = \delta_k = 0$

H_1: There is at least one $\beta_k = \delta_k \neq 0; k = 1,2,\ldots,p$

Test statistics for partial significance of parameters:

$$t_{value} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \quad (19)$$

where:

$\hat{\beta}_k$: the regression coefficient of the k-th independent variable (k=1,2,....p)

$se(\hat{\beta}_k)$: standard error of $\hat{\beta}_k$, $se(\hat{\beta}_k) = \sqrt{\text{var}(\hat{\beta}_k)}$

METHODOLOGY

Data Sources and Research Variables

Data was taken from the Central Bureau of Statistics and the Health Office in East Kalimantan [9][13], West Kalimantan [10][14], North Kalimantan [11][15], and Central Kalimantan [12][16]. The research variables used were the infant mortality rate (y), the rate of economic growth (x_1), the household with a clean and healthy lifestyle (x_2), the exclusive breastfeeding, (x_3) and the childbirth that wasn't assisted by medical personnel (x_4).

Data Analysis Stages

The data analysis steps in applying the nonparametric truncated spline regression model in this study are as follows:

1. Analyzing the pattern of relationships between variables using scatter plot.
2. Determining the optimal knot point using the GCV method given in Equation (16).
3. Estimating the nonparametric truncated spline regression model using Equation (15).
4. Testing the simultaneous hypothesis for the parameters of nonparametric truncated spline regression using Equation (18).
5. Testing the partial hypothesis on each predictor variable of the nonparametric truncated spline regression model using Equation (19).
6. Interpreting the nonparametric truncated spline regression model.

RESULT AND DISCUSSION

Data Exploration

Data exploration was carried out by looking at the pattern of the relationship between each predictor variable and response variable using scatter plot. If the pattern of the relationship does not follow a parametric curve then for parameter estimation a nonparametric regression approach is used[1][17]. Figure 1 shows the relationship pattern of each predictor variable to infant mortality rate.

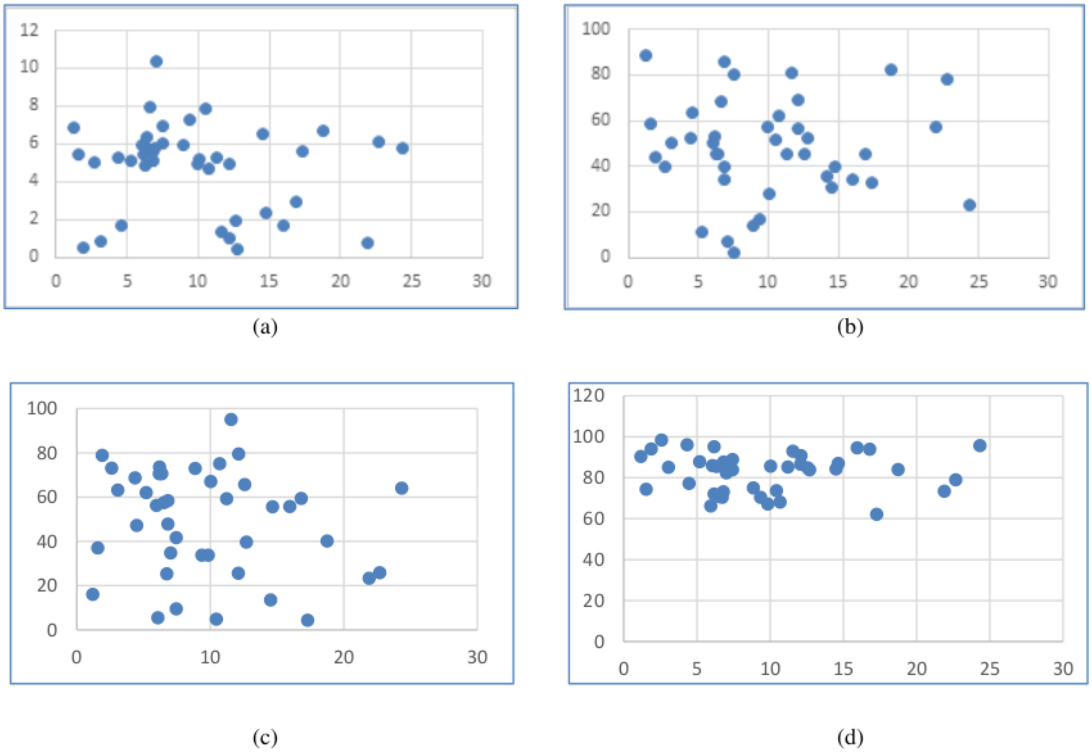

FIGURE 1. Pattern of the relationship between response variables and predictor variables

Figure 1 (a) shows the pattern of the relationship between IMR and the rate of economic growth, Figure 1 (b) shows the pattern of the relationship between IMR and the household with a clean and healthy lifestyle, Figure 1 (c) shows the pattern of the relationship between IMR and the exclusive breastfeeding, and Figure 1(d) shows the pattern of the relationship between IMR and the childbirth that wasn't assisted by medical personnel. It can be seen that the the relationship between the infant mortality rate and each predictor variable does not show a tendency for the pattern or the pattern of the regression curve is unknown., hence we used nonparametric regression approach in estimating.

Nonparametric Truncated Spline Regression

Nonparametric truncated spline regression modeling includes knot point selection, regression model estimation, simultaneous test, and partial test. The first step in the analysis is the selection of knot points obtained from the smallest GCV value.

Selection of Knot Point

The selection of the optimal knot point used the GCV method in Equation (16). The calculation results are presented in Table 1.

TABLE 1. Selection of Optimal Knot Point with Two Knot Points

Obs	x_1	x_2	x_3	x_4	GCV
1	3,151	25,736	28,862	71,864	41,952
	9,506	81,288	86,846	95,128	
2	3,151	25,736	28,862	71,864	41,971
	9,605	82,156	87,752	95,492	
3	3,250	26,604	29,768	72,228	42,974
	9,506	81,288	86,846	95,128	
4	3,250	26,604	29,768	72,228	42,026
	9,605	82,156	87,752	95,492	
5	3,051	24,868	27,956	71,501	42,036
	9,605	82,156	87,752	95,492	
6	3,051	24,868	27,956	71,501	42,046
	9,506	81,288	86,846	95,128	
7	3,349	27,472	30,674	72,591	42,091
	9,506	81,288	86,846	95,128	
8	3,349	27,472	30,674	72,591	42,180
	9,605	82,156	87,752	95,492	
9	2,952	24,000	27,050	71,137	42,242
	9,605	82,156	87,752	95,492	
10	2,952	24,000	27,050	71,137	42,242
	9,506	81,288	86,846	95,1285	

Based on Table 1 the optimal knot points using the minimum GCV are:

$k_{11} = 3{,}151 \quad k_{21} = 25{,}736 \quad k_{31} = 28{,}862 \quad k_{41} = 71{,}864$

$k_{12} = 9{,}506 \quad k_{22} = 81{,}288 \quad k_{32} = 86{,}846 \quad k_{42} = 95{,}128$

After obtaining the optimal knot points, the second step in the analysis is to estimate the parameters so that nonparametric truncated spline regression with geographically weighting is obtained.

Parameter Estimation of Nonparametric Truncated Spline Regression Model

The estimation of nonparametric truncated spline regression using 2 point knots and order of 2 is as follows:

$$\begin{aligned} \hat{y} = {} & 526{,}7461 + 9{,}0515x_1 + 1{,}5775x_1^2 + 0{,}7868(x_1 - 3{,}151) - 15{,}1201(x_1 - 9{,}506) \\ & - 1{,}9431x_2 - 0{,}0382x_2^2 - 0{,}0197(x_2 - 25{,}736) \\ & + 0{,}106(x_2 - 81{,}288) + 2{,}5174x_3 - 15{,}8224x_3^2 \\ & + 0{,}0458(x_3 - 28{,}862) - 0{,}3982(x_3 - 86{,}846) + 0{,}0246x_4 \\ & - 0{,}1719x_4^2 - 0{,}1021(x_4 - 71{,}864) - 0{,}7886(x_4 - 95{,}128) \end{aligned} \tag{20}$$

Furthermore, to determine whether the variables have an effect on infant mortality, it is necessary to test the significance of the parameters simultaneously and partially.

Simultaneous Test

The hypothesis of simultaneous parameter test:

$H_0 : \beta_1 = \ldots = \beta_4 = \delta_1 = \ldots = \delta_4 = 0$

$H_1 :$ There is at least one $\beta_k = \delta_k \neq 0;\ k = 1,2\ldots,4$

TABLE 2. Simultaneous Test

P-value	Decision
0,00001	Reject H_0

Table 2 states that the rate of economic growth, the household with a clean and healthy lifestyle, the exclusive breastfeeding and the childbirth process that wasn't assisted by medical personnel simultaneously affect the infant mortality rate.

Partial Test

The hypothesis of the partial parameter test:

$$H_0 : \beta_1 = ... = \beta_4 = \delta_1 = ... = \delta_4 = 0$$

$$H_1 : \beta_k = \delta_k \neq 0; k = 1,2,\dots,4$$

TABLE 3. Partial Test

Variable	Parameter	P-Value	Decision
	β_0	0,2046	Significant
	β_1	0,0745	
x_1	β_2	0,0123	Significant
	δ_1	0,1693	
	δ_2	0,2046	
	β_3	0,0569	
x_2	β_4	0,0087	Significant
	δ_3	0,1079	
	δ_4	0,2095	
	β_5	0,0538	
	β_6	0,2748	
x_3	δ_5	0,0064	Significant
	δ_6	0,0006	
	β_5	0,0538	
	β_7	0,1017	
x_4	β_8	0,1751	Not significant
	δ_7	0,2741	
	δ_8	0,1324	

Table 3 shows that there are 3 significant variables, namely the rate of economic growth, the household with a clean and healthy lifestyle, and the exclusive breastfeeding, while the childbirth process that wasn't assisted by medical personnel is not significant.

Model Interpretation

The final step performed is interpreting the results of the model formed. For x_1, the areas that have a percentage of economic growth rate of less than 3.151%, namely West Kutai, Kutai Kartanegara, East Kutai, Berau, North Penajam Paser, Samarinda, Bontang, Malinau, Nunukan, Tarakan , and Bulungan, The areas that have a percentage of the economic growth rate between 3.151% and 9.506% are Tana Tidung, East Kotawaringin, Kapuas, Katingan, West Kotawaringin, Murung Raya, Seruyan, Lamandau, Sukamara, Gunung Mas, Pulang Pisau, North Barito, South

Barito, Palangka Raya, East Barito, Balikpapan, Paser, Singkawang, Pontianak, Kubu Raya, Kayong Utara, Melawi, Sekadau, Sintang, Ketapang, Mempawah, Porcupine, Bengkayang and Sambas. Meanwhile, the area that has a percentage of the economic growth rate of more than 9.506% is Sanggau.

For x_2, the areas that have a percentage of the household with a clean and healthy lifestyle of less than 25.736% are Mempawah, Singkawang, Kapuas, and Tana Tidung. The areas that have a percentage of the household with a clean and healthy lifestyle between 25.736% and 81.288% are Sambas, Landak, Ketapang, Sintang, Sekadau, Melawi, North Kayong, Kubu Raya, Pontianak, Paser, West Kutai, Kutai Kartanegara, East Kutai, Berau, North Penajam Paser, Samarinda, West Barito, South Barito, North Barito, Pulang Pisau, Sukamara, Seruyan, Murung Raya, West Kotawaringin, Katingan, East Kotawaringin, Malinau, Nunukan, Tarakan, and Bulungan. Meanwhile, the areas that have a percentage of the household with a clean and healthy lifestyle of more than 81.288% are Bengkayang, Balikpapan, Bontang, and Palangka Raya.

For x_3, the areas that have a percentage of exclusive breastfeeding less than 28.862% are West Kutai, East Kutai, Palangka Raya, North Barito, Gunung Mas, Murung Raya, Katingan, and East Waringin City. The areas that have a percentage of exclusive breastfeeding between 28.862% and 86.846% are Tana Tidung, Bulungan, Tarakan, Nunukan, Malinau, Kapuas, West Kotawaingin, Seruyan, Lamandau, Pulang Pisau, East Barito, Samarinda, Balikpapan, North Penajam Paser, Berau, Kutai Kartanegara, Paser, Singkawang, Pontianak, Kubu Raya, North Kayong, Melawi, Sekadau, Sintang, Ketapang, Sanggau, Mempawah, Landak, Bengkayang, Sambas. Meanwhile, the area that has a percentage of exclusive breastfeeding of more than 86.846% is Bontang.

For x_4, the areas that have a percentage of the childbirth process that wasn't assisted by medical personnel of less than 71.864% are Ketapang, Melawi, North Kayong, South Barito, Seruyan, Kapuas and Murung Raya. The areas with a percentage of the childbirth process that wasn't assisted by medical personnel between 71.864% and 95.128% are Bulungan, Tarakan, Nunukan, Malinau, East Kotawaringin, Katingan, West Kotawaringin, Lamandau, Sukamara, Gunung Mas, Pulang Pisau, North Barito, Palangka Raya, Bontang. , Samarinda, Balikpapan, North Penajam Paser, Berau, East Kutai, Kutai Kartanegara, West Kutai, Paser, Singkawang, Kubu Raya, Sekadau, Sintang, Sanggau, Mempawah, Bengkayang, and Sambas. Meanwhile the areas that have a percentage of the childbirth process that wasn.t assisted by medical personnel of more than 86.846% are Pontianak, Landak, and Tana Tidung.

CONCLUSION

Based on the optimal nonparametric truncated spline regression model at order of 2 and 2 knot points, the R-square value was 88% and RMSE was 4.1076 and the assumptions had been fulfilled hence it was indicated as a good model. And the variables that affected the infant mortality rate were the rate of economic growth (x_1), the household with a clean and healthy lifestyle (x_2) and the exclusive breastfeeding (x_3).

ACKNOWLEDGMENTS

The authors would like to thank Mulawarman University. Thank for the reviewers for their careful reading of the manuscript and the fruitful suggestions that helped to improve this work

REFERENCES

[1]. Budiantara, I.N. (2009). *Spline dalam Regresi Nonparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Surabaya: FMIPA ITS.

[2]. Kementrian Pemberdayaan Perempuan dan Perlindungan Anak (KPPPA) & BPS. (2017). *Profil Anak Indonesia 2017*.

[3]. Eubank, R.L. (1998). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker

[4]. Sifriyani, S.H. Kartiko, I.N. Budiantara and Gunardi, 2018, *Development Of Nonparametric Geographically Weighted Regression Using Truncated Spline Approach*. Songklanakarin Journal Of Science And Technology, 40(4).

[5]. Hardle, W. (1990). *Applied Nonparametric Regression* . NewYork: Cambridge University Press.

[6]. Budiantara, I.N. dan Purnomo, J.D.T., 2011, *infants' Weight Growth Model in Surabaya (Indonesia) by using Weighted Spline Regression*, International Journal of Basic and Applied Sciences, Vol. 11, No. 2, hal. 119-123.
[7]. Draper & Smith. (1992). *Applied Regression Analysis*. Second Edition. New York: John Wiley & Sons.
[8]. Fotheringham, A.S., Brundson, C., & Charlton, M. (2002). *Geoghraphically Weighted Regression*. New York: John Wiley and Sons.
[9]. Badan Pusat Statistik Kalimantan Timur. Provinsi Kalimantan Timur Dalam Angka 2019. Badan Pusat Statistik Provinsi Kalimantan Timur. 2020.
[10]. Badan Pusat Statistik Kalimantan Barat.Provinsi Kalimantan Barat Dalam Angka 2019. Badan Pusat Statistik Provinsi Kalimantan Barat. 2020.
[11]. Badan Pusat Statistik Kalimantan Utara.Provinsi Kalimantan Utara Dalam Angka 2019. Badan Pusat Statistik Provinsi Kalimantan Utara. 2020.
[12]. Badan Pusat Statistik Kalimantan Tengah.Provinsi Kalimantan Tengah Dalam Angka 2019. Badan Pusat Statistik Provinsi Kalimantan Tengah. 2020.
[13]. Dinas Kesehatan Kalimantan Timur. (2020). *Profil Kesehatan Provinsi Kalimantan Timur*
[14]. Dinas Kesehatan Kalimantan Barat. (2020). *Profil Kesehatan Provinsi Kalimantan Barat*.
[15]. Dinas Kesehatan Kalimantan Tengah. (2020). *Profil Kesehatan Provinsi Kalimantan Tengah*.
[16]. Dinas Kesehatan Kalimantan Utara. (2020). *Profil Kesehatan Provinsi Kalimantan Utara*.
[17]. Anna Islamiyati, 2017. *Spline Polynomial Truncated dalam Regresi Nonparametrik*. Jurnal Matematika Statistika dan Komputasi. Vol 14, No. 1, 54-60

Check 13

ORIGINALITY REPORT

10%	2%	11%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Long Pang, Zhiwei Zhao, Xian-Yin Ma, Wen-Bin Cai, Limin Guo, Shaojun Dong, Chuntai Liu, Zhangquan Peng. " Hyphenated DEMS and ATR-SEIRAS techniques for multidimensional analysis of lithium-ion batteries and beyond ", The Journal of Chemical Physics, 2023 Publication	4%
2	A Prahutama, A Rusgiyono. "Robust regression with MM-estimator for modelling the number maternal mortality of pregnancy in Central Java, Indonesia", Journal of Physics: Conference Series, 2021 Publication	2%
3	A Halida, N Salam, A S Lestia. "Modeling of poverty in Kalimantan with nonparametric spline regression approach", Journal of Physics: Conference Series, 2021 Publication	2%
4	Dewi Retno Sari Saputro, Kornelius Ronald Demu, Purnami Widyaningsih. "Nonparametric truncated spline regression	2%

model on the data of human development index (HDI) in indonesia", Journal of Physics: Conference Series, 2018
Publication

5 Faisol, Tony Yulianto, Moh. Yaqin, Achmad Basuki, Muhammad Agus Zainuddin. "Prediction of seawater salinity using truncated spline regression method", AIP Publishing, 2022 2%
Publication

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		

Check 13

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11