BAREKENG: Journal of Mathematics and Its Applications
December 2023 Volume 17 Issue 4 Page 2135–2146
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss4pp2135-2146

COMPARING GAUSSIAN KERNEL AND QUADRATIC SPLINE OF NONPARAMETRIC REGRESSION IN MODELING INFECTIOUS DISEASES

Amanda Adityaningrum[1*], Sri Indriani Ladjali[2], Ismail Djakaria[3], Lailany Yahya[4], Muhammad Rezky Friesta Payu[5], La Ode Nashar[6], Herlina Jusuf[7]

[1,2,3,4,5,6]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gorontalo State University*
[7]*Department of Public Health, Faculty of Sports and Health, Gorontalo State University*
St. Jenderal Sudirman, Gorontalo, 96128, Indonesia

*Corresponding author's e-mail: ** ***amanda@ung.ac.id***

ABSTRACT

Article History:

Received: 2nd July 2023
Revised: 1st October 2023
Accepted: 2nd November 2023

Keywords:

Gaussian Kernel; Quadratic Spline; Tuberculosis; Diarrhoeal; Pneumonia; COVID-19.

The regression curve for nonparametric regression is assumed to belong to some infinite-dimensional collection of functions, which allows great flexibility in the form of the curve. This research intends to compare the Gaussian Kernel and Quadratic Spline regressions in four infectious diseases in Indonesia by 2021. The data used is secondary data from the Central Bureau of Statistics and the Ministry of Health, Indonesia, and the sample consists of four infectious diseases in Indonesia by 2021 (Tuberculosis, Diarrhoeal, Pneumonia, and COVID-19). Considering the correlation value, it was found that the independent and dependent variables of the four infectious diseases are all highly correlated (r values are more than 0.7). Furthermore, the scatter plots for four infectious diseases do not follow a particular pattern; due to this, parametric regression cannot be used to analyze the data. Therefore, nonparametric regression was applied in this research. According to the analysis, the Gaussian Kernel is the best regression technique for modeling four infectious diseases in Indonesia by 2021, which its R^2 *values are 99.85% (Tuberculosis), 100% (Diarrhoeal), 99.91% (Pneumonia), and 99.99% (COVID-19).*

How to cite this article:

A. Adityaningrum, S. I. Ladjali, I. Djakaria, L. Yahya, M. R. F. Payu, L O. Nashar and H. Jusuf., " COMPARING GAUSSIAN KERNEL AND QUADRATIC SPLINE OF NONPARAMETRIC REGRESSION IN MODELING INFECTIOUS DISEASES," *BAREKENG: J. Math. & App.*, vol. 17, iss. 4, pp. 2135-2146, December, 2023.

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

Research Article · Open Access

1. INTRODUCTION

Sustainable Development Goals (SDGs) contain several objectives, one of which is to "ensure healthy lives and promote well-being for all at all ages", and one of the targets is to "end infectious diseases" [1], [2]. Due to the pandemic and other ongoing crises, progress toward universal health coverage is being hampered, exacerbating existing health inequalities and jeopardizing the realization of this goal. Health systems in low- and middle-income countries have been particularly challenged by this since they already had inadequate resources prior to the outbreak [1]. Since Indonesia is one of the middle-income countries, it is mandatory to pay attention to infectious diseases [3]. It has been reported that Tuberculosis, Diarrhoeal, and Pneumonia (lower respiratory infections) are the top three infectious diseases that cause death in Indonesia for both sexes and all ages [4]. In addition to the three previous diseases, Coronavirus disease (COVID-19) is another infectious disease included in the disease control category by the Indonesian Ministry of Health, where this disease was declared a pandemic by WHO on March 11, 2020 [5]. Nonparametric regression can be used to analyze infectious diseases since they do not follow specific assumptions or patterns.

Nonparametric regression can be used to analyze infectious diseases since they do not follow specific assumptions or patterns. Regression analysis is a statistical technique to explore and model relationships between variables in a dataset [6]. In regression, the functions of estimate parameters are referred to as regression functions or regression curves, and it describes the relationships between dependent and independent variables [7], [8]. According to [9], for the proper interpretation of data, observational errors must be minimized to focus on the essential details of mean dependence between independent and dependent variables. The mean function can be approximated in two ways, namely parametric and nonparametric techniques [9].

Based on [8], the parametric regression model presumes that the form of *f(x)* is known and can depend on the parameters in a linear or nonlinear fashion. A fitted curve is generated by using this regression, which has been chosen from the family of curves that the model allows. Essentially, experimenters select a set of curves from a collection of all curves and input them into the inferential process. Consequently, the data provide limited information regarding the development of a model based on its assumed parametric form [8].

In contrast, as stated by [8], nonparametric regression techniques emphasize data to determine *f(x)*. Regression curves are assumed to belong to some infinite-dimensional collection of functions, such as a linear function; for instance, *f(x)* may be considered to differ from a square-integrable second derivative. It allows great flexibility in the form of the regression curve and, in particular, makes no assumptions regarding a parametric model. When specifying this regression model, the experimenter chooses a function space that contains the unknown regression, which is usually motivated by the assumption that the regression function will possess smoothness (i.e., continuity and differentiability). The data is then utilized to determine an element of this function space that is representative of the unknown curve [8]. Kernel, Spline, and Fourier Series are often considered nonparametric regression approaches [10].

Several previous studies have used nonparametric regression to study infectious diseases, including: (1) using Linear Spline [11]–[16], using Gaussian Kernel [17], and using Fourier Series [18] for modeling tuberculosis; (2) using Linear Spline [19] and using Qubic Spline [20] for modeling diarrhoeal; (3) using Linear Spline [21], B-Spline [22], and using Fourier Series and Linear Spline [23] for modeling pneumonia; and (4) using Linear Spline [24], [25], using Cubic Spline [26], using Cubic B-Spline [27], using Penalized Spline [28], Cubic Spline, Smoothing Spline, and Multivariate Adaptive Regression Spline (MARS) [29], using Linear, Quadratic, and Qubic Spline [30], using Polynomial Gaussian Kernel [31], using Regular Linear Kernel Regression (RLKR) [32], using Local Polynimial Gaussian Kernel [33], using Gaussian, Epanechnikov, Triangular, Biweight, and Triweight Kernel [34], and using Fourier Series [35] for modeling COVID-19. Considering the above description, it is evident that kernel and spline regressions are commonly used to model infectious diseases.

Compared to other kernel functions, gaussian is considered smoother than other kernel functions and has optimal criteria for selecting bandwidth for normal data or close to normal, which enables a fine and precise balance between fitting and smoothing [33], [36]. Meanwhile, for spline, [37] indicates that linear splines are generally used for simple data patterns, whereas quadratic splines are usually used for data with complex data patterns. The statement is consistent with research by [38], which found quadratic splines more appropriate for modeling toddler growth data (a complex set of data with outliers) than linear splines. Until this date, no research has been comparing nonparametric regressions, namely Gaussian Kernel and Quadratic Spline, in infection studies. Thus, this research intends to compare both regressions in four infectious diseases in Indonesia by 2021. The research aims to provide insight into nonparametric regression in infectious disease case studies.

2. RESEARCH METHODS

2.1 Data

The data is secondary data from the Ministry of Health and the Central Bureau of Statistics (BPS) in Indonesia. This study aimed to examine infectious diseases in Indonesia by 2021, and the sample consists of four infectious diseases in Indonesia by 2021 (Tuberculosis, Diarrhoeal, Pneumonia, and COVID-19).

a. Tuberculosis

Tuberculosis (T) is caused by a bacterium called Mycobacterium Tuberculosis which attacks the lungs and other organs in the body, and pulmonary tuberculosis is responsible for 80% of all cases [39], [40]. Gender is one of the factors that influence the number of tuberculosis cases, with men having twice the potential to contract this disease as women due to their high mobility, so they are more likely to be exposed, and there is also the knowledge that Indonesia will have a higher number of men than women in 2021, both nationally and provincially [5], [41], [42]. Thus, tuberculosis cases are the dependent variable (Y_T), while the male population is the independent variable (X_T).

b. Diarrhoeal

Diarrhoeal (D), the second leading cause of death in children under five years of age, is defined as the passing of loose stools as well as an increase in stool frequency, weight, or volume, and it is usually associated with long-term health issues such as malnutrition, stunted growth, and immune system defects [43]–[45]. According to previous research, breastfeeding up to 6 months of age lowers diarrhea risk since most children under 2 consume infant formula, and exclusively breastfeeding lowers diarrhea risk [44], [46]. In addition, infants with the lowest and medium breastfeeding performance index (BPI) categories were 2.22 and 2.15 times more likely to develop diarrhea than infants with the highest BPI, respectively [47]. Thus, diarrhoeal cases are the dependent variable (Y_D), while exclusively breastfed infants are the independent variable (X_D).

c. Pneumonia

Pneumonia (P), the leading cause of death for children under five each year, is a viral, bacterial, or fungal respiratory tract infection that attacks lung tissues [48], [49]. Previous research found that low birth weight was associated with severe pneumonia and increased mortality risk for children under five [50]–[52]. Thus, pneumonia cases are the dependent variable (Y_P), while low birth weight infants are the independent variable (X_p).

d. COVID-19

COVID-19 (C) is a highly contagious virus caused by a type of Coronavirus that has been spreading worldwide, forcing most countries to recommend or require restrictive measures such as home isolation and mask-wearing [53]–[55]. Unlike males, females account for most COVID-19 cases in many countries, including Indonesia [5], [56]–[58]. In addition, it was found that females were more likely than males to suffer from long COVID [59], [60]. Thus, COVID-19 cases are the dependent variable (Y_C), while the female population is the independent variable (X_C).

2.2 Methods

2.2.1. Correlation and Scatter Plot

A correlation analysis was conducted to determine whether there was a significant relationship between the dependent (Y) and independent (X) variables. According to [61]–[63], the correlation coefficient, commonly expressed as Pearson product-moment correlation, measures the strength of an association in which two variables play a similar role and can be substituted for each other. It takes values between -1 and 1, where -1 or 1 implies a perfect linear relationship. Moreover, its characterization is weak $\left(r < \pm 0.4\right)$, moderate $\left(r \text{ between } \pm 0.4 \text{ to } \pm 0.7\right)$, and strong $\left(r > \pm 0.7\right)$ [61]–[63]. The Pearson correlation for random variables X and Y is written as follows:

$$r = \frac{\sum_{j=1}^{k}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{k}(x_j - \bar{x})^2 \sum_{j=1}^{k}(y_j - \bar{y})^2}} \tag{1}$$

where x_j, y_j, $j=1,2,\ldots,k$ are observed bivariate data and $\bar{x} = \sum_{j=1}^{k} x_j \big/ k$, $\bar{y} = \sum_{j=1}^{k} y_j \big/ k$ are sample means [64].

A scatter plot was constructed to determine whether regression function $f(x)$ is known (or not). It is a pattern of X and Y relationships that showed by a plot [65], [66].

2.2.2. Nonparametric Regression

According to [8], [9], if n points of data $(t_1, y_1),\ldots,(t_n, y_n)$ have already been collected, it is possible to model the regression relationship as follows:

$$y_i = \mu(t_i) + \varepsilon_i, \quad i = 1,\ldots,n \tag{2}$$

with μ as an unknown function (or regression function/curve) and ε_i as an observation error. Regression analysis aims to obtain a reasonable solution to the unknown response function m, and one of the ways to do so is through nonparametric techniques, such as kernel or spline regression [8], [9].

2.2.3. Gaussian Kernel

Kernel regression is a nonparametric regression that uses kernel weighting functions (or kernel function) to estimate conditional expectations of random variables (which are dependent and independent), and the estimation value is expressed as a weighted sum of the responses at any point t [8], [67]. Based on [8], [9], the kernel function (in one dimension x) that are constructed using the following formula:

$$K(t, t_i; \lambda) = \frac{1}{\lambda} K\left(\frac{t - t_i}{\lambda}\right), \tag{3}$$

with K defining the kernel functions, while λ defining the size (referred to as bandwidth or smoothing parameter) [8], [9]. There are several kernel functions, including the Gaussian kernel. According to [9], [68], the gaussian kernel is written as follows:

$$K(u) = (2\pi)^{-1/2} \exp\left(-u^2 \big/ 2\right), -\infty < u < \infty \tag{4}$$

where **Equation (4)** represents the kernel with u as $(t, t_i;\lambda)$. Meanwhile, the estimator obtained from **Equation (3)** is written as follows:

$$\mu_\lambda(t) = (n\lambda)^{-1} \sum_{i=1}^{n} K\left(\lambda^{-1}(t - t_i)\right) y_i \tag{5}$$

with its kernel estimators proposed by Nadaraya and Watson in 1964. It is written as follows:

$$\mu_\lambda(t) = \frac{\sum_{i=1}^{n} K\left(\lambda^{-1}(t - t_i)\right) y_i}{\sum_{i=1}^{n} K\left(\lambda^{-1}(t - t_j)\right)}. \tag{6}$$

Equation (6) is called the Nadaraya-Watson Estimator (NWE) [8], [9], [69], [65], [66].

2.2.4. Quadratic Spline

Spline regression is a nonparametric regression that employs continuous segmented (truncated) polynomials to estimate the behavior of data that encounters variance at various intervals [67]. Generally, truncated polynomial spline functions with function μ and order m $(g+1)$ are defined as functions with knots (joint points indicating changes in the data's behavior) that are constructed using the following formula:

$$\mu(t)=\sum_{j=0}^{g}\beta_j t^j+\sum_{k=1}^{q}\beta_{k+g}\left(t-\xi_k\right)_+^g . \tag{7}$$

where the truncated function is written as follows:

$$\left(t-\xi_k\right)_+^g=\begin{cases}\left(t-\xi_k\right)^g & ;t-\xi_k\geq 0\\ 0 & ;t-\xi_k<0\end{cases} \tag{8}$$

with β_j as a polynomial coefficient, β_{k+g} as a truncated coefficient, and ξ_k is knot $\xi_1,\xi_2,\ldots,\xi_v$ $\left(a<\xi_1<\xi_2<\ldots<\xi_v<b\right)$ [8], [67], [70], [71]. An example of a spline regression is the Quadratic spline. This regression has three as its order (g=2) [67]. From Equation (7), quadratic spline can be written as follows:

$$\mu(t)=\beta_0+\beta_1 t+\beta_2 t^2+\sum_{k=1}^{q}\beta_{k+2}\left(t-\xi_k\right)_+^2 . \tag{9}$$

and Equation (9) can be substituted into Equation (2); hence, it can be written as follows:

$$y_i=\beta_0+\beta_1 t+\beta_2 t^2+\sum_{k=1}^{q}\beta_{k+2}\left(t-\xi_k\right)_+^2+\varepsilon_i \tag{10}$$

with $i=1,\ldots,n$ [37], [72]. Coefficient β is estimated using the Ordinary Least Squares (OLS) [67].

2.2.5. Generalized Cross Validation

The regression curve can be obtained by selecting the optimum bandwidth for the kernel and optimum knots for the spline (number and location of knot points) [14], [36], [69]. Generalized Cross Validation (GCV) can be used to calculate both, where a model with the lowest GCV score will be selected for the subsequent analysis [69], [73]. GCV is written as follows:

$$GCV=\frac{MSE}{\left[n^{-1}\text{trace}\left(I-H\right)\right]^2} \tag{11}$$

where $H=X\left(X^tX\right)^{-1}X^t$ and $MSE=n^{-1}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$ [69], [74]. GCV is simple and efficient in its calculations, asymptotically optimal, invariant to transformation, and does not require σ^2 information compared to other techniques, such as Cross Validation, Unbiassed Risk, and Generalized Maximum Likelihood [14], [75].

2.2 Analyses Steps

The analysis steps for this research are: (i) doing data exploration; (ii) conducting the model of the gaussian kernel; (iii) conducting the model of the quadratic spline; and (iv) comparing both models. All steps are performed using R software.

3. RESULTS AND DISCUSSION

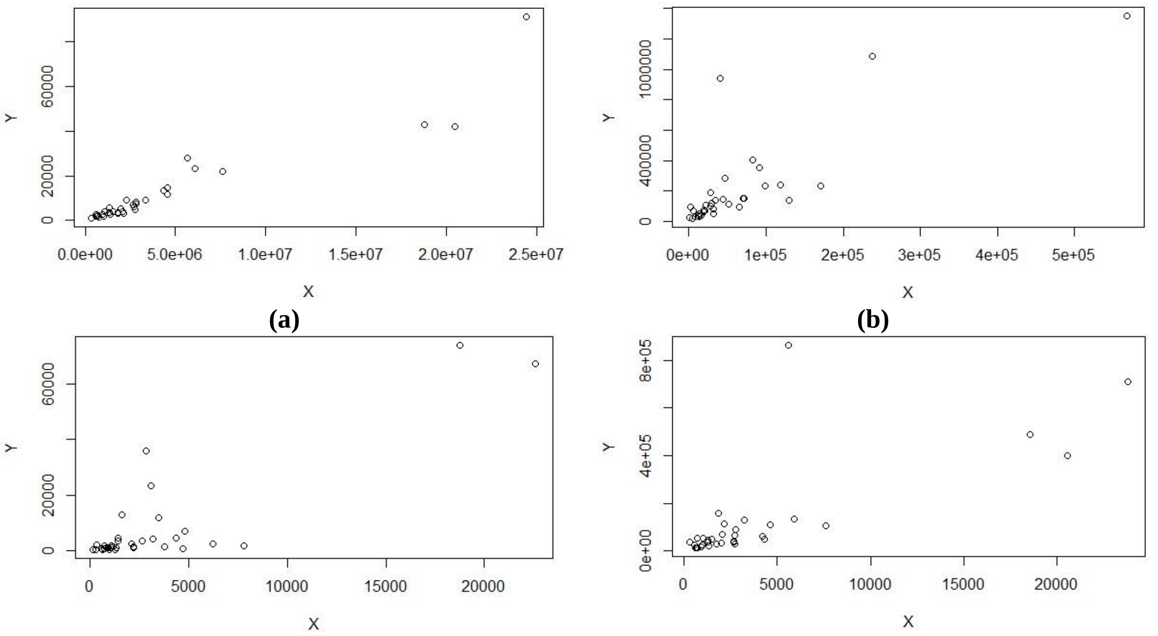
3.1 Data Exploration

Data exploration in this research consists of performing a correlation analysis and making scatter plots. Table 1 presents the result of the correlation analysis using Equation (1). As illustrated in Table 1, four infectious diseases demonstrate a strong correlation between the X and Y variables; hence, indicating that the dependent and independent variables are indeed correlated and can be used to perform a regression analysis.

Table 1. **Correlate Analysis**

Dependent Variables	Independent Variables	Correlate Values
Y_T	X_T	0.947
Y_D	X_D	0.811
Y_P	X_P	0.864
Y_C	X_C	0.739

The scatter plot for four infectious diseases showed in Figure 1. As shown in Figure 1, all the scatter plots do not follow a certain pattern. Therefore, nonparametric regression is used to analyze the data.



(a) (b)

(c) (d)

Figure 1. Scatter plot of *Y* and *X* for four infectious diseases,
(a) Scatter plot of *Y* and *X* of Tuberculosis, (b) Scatter plot of *Y* and *X* of Diarrhoeal,
(c) Scatter plot of *Y* and *X* of Pneumonia, and (d) Scatter plot of *Y* and *X* of COVID-19

3.2 Nonparametric Regression Analysis

3.2.1 Gaussian Kernel

Optimal bandwidth (λ) values must be selected for the gaussian kernel modeling. The curve is unsmooth in models with a small bandwidth value, whereas in models with a large bandwidth value, the curve is over-smooth [69]. Due to this, the optimal bandwidth should be selected using the GCV score in Equation (11); the optimal bandwidth is the one that has the lowest GCV score. Table 2 presents the results of GCV scores for four infectious diseases, sorted by lowest GCV (only three iterations are shown as an example).

Table 2. Optimal Bandwidth Value and GCV Score for Four Infectious Diseases

Tuberculosis		Diarrhoeal		Pneumonia		COVID-19	
λ	GCV	λ	GCV	λ	GCV	λ	GCV
73.5	**2952012**	**10.5**	**139697364578**	**28**	**2301250**	**1**	**143795350**
74	2952028	50	139704020422	28.5	2302837	35.5	960385186
73	2952140	49.5	139704963146	27.5	2303825	36	960544120

As illustrated in Table 2, the optimum bandwidth for tuberculosis, diarrhoeal, pneumonia, and COVID-19 are 73.5, 10.5, 28, and 1, respectively, these bandwidth are the smoothing parameters in kernel. Therefore, as an example of tuberculosis, a model for the gaussian kernel can be written as follows:

$$\mu_\lambda(t)=\frac{\sum_{i=1}^{n}(2\pi)^{-1/2}\exp\left(\frac{-\left(73.5^{-1}(t-t_i)\right)^2}{2}\right)y_i}{\sum_{i=1}^{n}(2\pi)^{-1/2}\exp\left(\frac{-\left(73.5^{-1}(t-t_j)\right)^2}{2}\right)} \tag{12}$$

3.2.2 Quadratic Spline

A spline model can be obtained by selecting the optimal knot (ξ) points based on the number and location of knots. If there are too many knots in a model, the curve tends to be overfitted and unsmooth, whereas if there are too few knots, the curve becomes over smooth and cannot describe the data distribution [76]. Accordingly, the optimal knot should be selected using the GCV score in Equation (11); the optimal knot is the one that has the lowest GCV score. Table 3, Table 4, Table 5, and Table 6 presents the results of GCV scores for four infectious diseases, sorted by smallest GCV (only three iterations are shown as an example).

Table 3. Optimal Knot Location and GCV Score for Tuberculosis

ξ Total	ξ Location			GCV
1	**7238.839**			**33489234**
	6093.2			33582366
	8384.477			33691610
2	**6093.2**	**7238.839**		**32319622**
	7238.839	8384.477		32456517
	6093.2	8384.477		32537461
3	**2656.286**	**3801.924**	**4947.562**	**27479532**
	2656.286	3801.924	19840.858	27840705
	2656.286	3801.924	18695.22	27875637

First, for tuberculosis, as illustrated in Table 3, the lowest GCV scores are located at knot 7238.839 for knot 1, 6093.2 and 7238.839 for knot 2, and 2656.286, 3801.924, and 4947.562 for knot 3.

Table 4. Optimal Knot Location and GCV Score for Diarrhoeal

ξ Total	ξ Location			GCV
1	**28284.38**			**67593762523**
	55247.76			67593762523
	82211.14			67593762523
2	**28284.38**	**55247.76**		**67593762522**
	28284.38	82211.14		67593762522
	28284.38	109174.52		67593762522
3	**28284.38**	**55247.76**	**82211.14**	**67593762522**
	28284.38	55247.76	109174.52	67593762522
	28284.38	55247.76	136137.9	67593762522

Second, as illustrated in Table 4, for diarrhoeal, the lowest GCV scores are located at knot 28284.38 for knot 1, 28284.38 and 55247.76 for knot 2, and 28284.38, 55247.76, and 82211.14 for knot 3.

Table 5. Optimal Knot Location and GCV Score for Pneumonia

ξ Total	ξ Location			GCV
1	**18307.905**			**74937683**
	17241.381			74973468
	19374.429			74998186
2	**17241.38**	**18307.9**		**74748807**
	18307.9	19374.43		74772535
	16174.86	18307.9		74800854
3	**2310.048**	**3376.571**	**8709.19**	**74413466**
	2310.048	3376.571	9775.714	74483512
	17241.381	18307.905	19374.429	74579478

Next, as illustrated in Table 5, the lowest GCV scores for pneumonia are located at knot 18307.905 for knot 1, 17241.38 and 18307.9 for knot 2, and 2310.048, 3376.571, and 8709.19 for knot 3.

Table 6. Optimal Knot Location and GCV Score for COVID-19

ξ Total	ξ Location			GCV
1	**5919.583**			**18837387550**
	7036.901			18870510842
	8154.22			18910095401
2	**5919.583**	**7036.901**		**18720903282**
	5919.583	8154.22		18761891165
	5919.583	9271.539		18783707188
3	**5919.583**	**7036.901**	**8154.22**	**18650714941**
	5919.583	7036.901	9271.539	18671654360
	5919.583	7036.901	10388.857	18691220598

Last, for COVID-19, as illustrated in Table 6, the lowest GCV scores are located at knot 5919.583 for knot 1, 5919.583 and 7036.901 for knot 2, and 5919.583, 7036.901, and 8154.22 for knot 3.

In order to determine the best quadratic regression model, all knot points are then calculated for their R^2 values. Table 7 presents the results of R^2 for four infectious diseases. As illustrated in Table 7, the best quadratic spline model for tuberculosis has three knots and six regression coefficients, with the highest R^2 of 92.17%.

Table 7. R^2 Values of Quadratic Spline Model for Four Infectious Diseases

ξ Total	R^2 (%) Tuberculosis	 Diarrhoeal	 Pneumonia	 COVID-19	Total Coefficient
1	90.46	**78.14**	**58.71**	**55.43**	4
2	90.75	78.19	58.71	55.61	5
3	**92.17**	78.78	58.71	55.76	6

Meanwhile, for diarrhoeal, pneumonia, and COVID-19, models with one knot have R^2 values similar to two or three knots. Furthermore, according to [65], [72], [75], [77], [78], models with low oscillation and parsimony (a model that consists of a few parameters and is capable of producing a high R^2 value) are recommended for modeling. Therefore, for parsimony reasons, the best quadratic spline model in diarrhoeal, pneumonia, and COVID-19 consists of one knot and has four regression coefficients with R^2 of 78.14%, 58.71%, and 55.43%, respectively. The quadratic spline model for four infectious diseases can be written in Equation (13), Equation (14), Equation (15), and Equation (16).

a) Tuberculosis

$$\mu(t)=\beta_0+\beta_1 t+\beta_2 t^2+\beta_3(t-2656)_+^2+\beta_4(t-3802)_+^2+\beta_5(t-4948)_+^2 \tag{13}$$

b) Diarrhoeal

$$\mu(t)=\beta_0+\beta_1 t+\beta_2 t^2+\beta_3(t-28284)_+^2 \tag{14}$$

c) Pneumonia

$$\mu(t)=\beta_0+\beta_1 t+\beta_2 t^2+\beta_3(t-18308)_+^2 \tag{15}$$

d) COVID-19

$$\mu(t)=\beta_0+\beta_1 t+\beta_2 t^2+\beta_3(t-5920)_+^2 \tag{16}$$

3.3 Comparison of Kernel and Spline Regression

Compared to SMAPE, MAPE, MAE, MSE, and RMSE, R^2 appears to be the most informative rate in many cases based on our experience and the results of this study [79]. Thus, the R^2 value is used to compare the best nonparametric regression.

Table 8. R^2 Values of Three Regression Techniques for Four Infectious Diseases

Regression Techniques	R^2 (%) Tuberculosis	 Diarrhoeal	 Pneumonia	 COVID-19
Gaussian Kernel	**99.85**	**100**	**99.91**	**99.99**
Quadratic Spline	92.17	58.71	78.78	55.76
Linear	89.61	65.74	74.72	54.55

Table 8 presents the results of R^2 for the gaussian kernel, quadratic spline, and linear regression for four infectious diseases. As illustrated in **Table 8**, the gaussian kernel is the most suitable regression technique for modeling four infectious diseases in Indonesia by 2021, as it has high R^2 values (for each disease) compared to quadratic spline and linear regression.

4. CONCLUSIONS

Based on the analysis findings, the R^2 values for tuberculosis, diarrhoeal, pneumonia, and COVID-19 are 99.85%, 100%, 99.91%, and 99.99%, respectively. Therefore, the most suitable regression technique for modeling four infectious diseases in Indonesia by 2021 is the gaussian kernel, since it has high R^2 values compared to two other regression techniques. Research was limited to one independent variable for each infectious disease. Therefore, independent variables can be added to compare gaussian kernel and quadratic spline for further research.

REFERENCES

[1] UN, “SDG Progress Report 2023,” 2023. [Online]. Available: https://sdgs.un.org/documents/sdg-progress-report-2023-advance-unedited-version-52208

[2] WHO, “SDG Target 3.3 Communicable diseases,” *THE GLOBAL HEALTH OBSERVATORY*. https://www.who.int/data/gho/data/themes/topics/sdg-target-3_3-communicable-diseases (accessed Jul. 02, 2023).

[3] World Bank, “Indonesia,” *The World Bank Data*, 2023. https://data.worldbank.org/country/ID (accessed Jul. 02, 2023).

[4] W. H. O. WHO, “Global health estimates: Leading causes of death,” 2023. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death (accessed May 10, 2023).

[5] R. Kemenkes, “Profil Kesehatan Indonesia 2021,” Indonesia, 2022. [Online]. Available: https://www.kemkes.go.id/downloads/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-2021.pdf

[6] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 5th ed. United Kingdom: Wiley, 2012.

[7] A. Schneider, G. Hommel, and M. Blettner, “Linear Regression Analysis,” *Dtsch. Arztebl. Int.*, Nov. 2010, doi: 10.3238/arztebl.2010.0776.

[8] R. L. Eubank, *Nonparametric regression and spline smoothing*, 2nd ed. New York: Marcel Dekker, Inc, 1999.

[9] W. Härdle, *Applied Nonparametric Regression*. Berlin: Cambridge University Press, 1990. doi: 10.1017/CCOL0521382483.

[10] P. P. Gabrela, J. D. T. Purnomo, and I. N. Budiantara, “Mixed spline truncated, kernel, and Fourier series estimator in biresponse nonparametric regression,” 2023, p. 080015. doi: 10.1063/5.0105823.

[11] M. H. Pratama, Sifriyani, and D. A. Nohe, “Regresi Nonparametrik Multivariabel dengan Pendekatan Spline Truncated pada Kasus Tuberculosis,” *STATISTIKA*, vol. 22, no. 1, pp. 87–93, 2022, [Online]. Available: https://journals.unisba.ac.id/index.php/statistika/article/view/506

[12] R. Asriyanti, I. Yahya, B. Abapihi, G. N. A. Wibawa, and L. Laome, “Penerapan Regresi Nonparametrik Spline dalam Memodelkan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Tuberkulosis di Sulawesi Tenggara,” *Semin. Nas. Sains dan Terap. VI*, vol. 6, no. April, pp. 100–109, 2022, [Online]. Available: https://ejournal.unsrat.ac.id/v3/index.php/sinta6/article/view/41875

[13] N. Evitasari, S. S. Handajani, and H. Pratiwi, “Pemodelan faktor-faktor yang memengaruhi angka kesembuhan tuberkulosis di Jawa Barat menggunakan regresi spline truncated,” *Maj. Ilm. Mat. dan Stat.*, vol. 22, no. 2, p. 118, Sep. 2022, doi: 10.19184/mims.v22i2.30356.

[14] I. G. P. S. Darma, M. Ratna, and I. N. Budiantara, “Pemodelan Faktor yang Memengaruhi Angka Kasus Tuberculosis di Kota Surabaya Menggunakan Regresi Nonparametrik Spline Truncated,” *J. Sains dan Seni ITS*, vol. 8, no. 2, Feb. 2020, doi: 10.12962/j23373520.v8i2.45404.

[15] D. F. W. Hapsari and L. Khikmah, “Modeling of Tuberculosis Case in Central Java 2018 with Three Knot Point,” *J. Intell. Comput. Heal. Informatics*, vol. 1, no. 2, p. 53, Sep. 2020, doi: 10.26714/jichi.v1i2.6348.

[16] N. P. R. Anggreni, N. L. P. Suciptawati, and I. G. A. M. Srinadi, “Model Regresi Nonparametrik Spline Truncated pada Jumlah Kasus Tuberkulosis di Provinsi Bali Tahun 2016,” *E-Jurnal Mat.*, vol. 7, no. 3, p. 211, Sep. 2018, doi: 10.24843/MTK.2018.v07.i03.p205.

[17] E. Andrianto, “Pendugaan Regresi Nonparametrik Dengan Fungsi Kernel Gaussian,” Universitas Brawijaya, Malang, 2014. [Online]. Available: http://repository.ub.ac.id/id/eprint/154034/

[18] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanto, and D. P. Rahmawati, “Nonparametric Regression Model for Longitudinal Data with Mixed Truncated Spline and Fourier Series,” *Abstr. Appl. Anal.*, vol. 2020, pp. 1–11, Dec. 2020, doi: 10.1155/2020/4710745.

[19] M. S. Khudzaifi, “Pemodelan Kasus Penyakit Diare pada Balita di Kota Surabaya Menggunakan Regresi Nonparametrik Spline Truncated,” Institut Teknologi Sepuluh Nopember, Surabaya, 2017. [Online]. Available: https://repository.its.ac.id/47753/

[20] S. A. Richard *et al.*, “Diarrhea in Early Childhood: Short-term Association with Weight and Long-term Association with Length,” *Am. J. Epidemiol.*, vol. 178, no. 7, pp. 1129–1138, Oct. 2013, doi: 10.1093/aje/kwt094.

[21] C. N. Afisah, “Faktor-faktor yang Mempengaruhi Pneumonia pada Balita di Jawa Barat Menggunakan Regresi Nonparametrik Spline Truncated,” Institut Teknologi Sepuluh Nopember, Surabaya, 2020. [Online]. Available: https://repository.its.ac.id/73622/

[22] I. G. A. M. V. Dewi, I. G. A. M. Srinadi, and M. Susilawati, “Pemodelan Kasus Pneumonia pada Balita di Provinsi Bali

Menggunakan Metode Regresi Nonparametrik B-Spline," *E-Jurnal Mat.*, vol. 9, no. 3, p. 197, Sep. 2020, doi: 10.24843/MTK.2020.v09.i03.p299.

[23] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanto, and D. P. Rahmawati, "Pneumonia cases modeling in Java Island using two estimators of nonparametric regression for longitudinal data," 2021, p. 060029. doi: 10.1063/5.0042763.

[24] C. Imama *et al.*, "Mask Compliance Modeling Related COVID-19 in Indonesia Using Spline Nonparametric Regression," *J. Varian*, vol. 5, no. 2, pp. 179–190, May 2022, doi: 10.30812/varian.v5i2.1895.

[25] K. Demertzis, D. Tsiotas, and L. Magafas, "Modeling and Forecasting the COVID-19 Temporal Spread in Greece: An Exploratory Approach Based on Complex Network Defined Splines," *Int. J. Environ. Res. Public Health*, vol. 17, no. 13, p. 4693, Jun. 2020, doi: 10.3390/ijerph17134693.

[26] R. Yamamoto *et al.*, "Threshold of increase in oxygen demand to predict mechanical ventilation use in novel coronavirus disease 2019: A retrospective cohort study incorporating restricted cubic spline regression," *PLoS One*, vol. 17, no. 7, p. e0269876, Jul. 2022, doi: 10.1371/journal.pone.0269876.

[27] O. Polsen and P. Kamoljitprapa, "Cubic B-spline and Generalised Linear Models for COVID-19 Patients in Thailand," *J. Appl. Sci.*, vol. 20, no. 2, pp. 23–33, Sep. 2021, doi: 10.14416/j.appsci.2021.02.003.

[28] N. Chamidah, M. Rifada, and D. Amelia, "A theoretical discussion on modeling the number of COVID-19 death cases using penalized spline negative binomial regression approach," *Commun. Math. Biol. Neurosci.*, vol. 2022, 2022, doi: 10.28919/cmbn/7518.

[29] R. Diana and R. Rory, "PEMODELAN KASUS COVID-19 MENGGUNAKAN MODEL REGRESI NONPARAMETRIK," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 108–115, Jan. 2021, doi: 10.34123/semnasoffstat.v2020i1.588.

[30] C. Liang and L. Shen, "An analysis of coronavirus disease 2019 with spline regression at province level during first-level response to major public health emergency out of Hubei, China," *Epidemiol. Infect.*, vol. 149, p. e10, Jan. 2021, doi: 10.1017/S095026882000312X.

[31] T. W. Utami and A. Lahdji, "MODELING OF LOCAL POLYNOMIAL KERNEL NONPARAMETRIC REGRESSION FOR COVID DAILY CASES IN SEMARANG CITY, INDONESIA," *MEDIA Stat.*, vol. 14, no. 2, pp. 206–215, Jan. 2022, doi: 10.14710/medstat.14.2.206-215.

[32] W. Chao, L. Bo, L. Huan, and P. Pai, "Consistence Condition of Kernel Selection in Regular Linear Kernel Regression and Its Application in COVID-19 High-risk Areas Exploration," *Res. Sq.*, pp. 0–20, 2022, doi: https://doi.org/10.21203/rs.3.rs-1430675/v1.

[33] R. Rory and R. Diana, "Pemodelan Data COVID-19 Menggunakan Regresi Polinomial Lokal," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 91–98, Jan. 2021, doi: 10.34123/semnasoffstat.v2020i1.554.

[34] S. Saidi, N. Herawati, K. Nisa, and E. Setiawan, "Nonparametric Modeling Using Kernel Method for the Estimation of the Covid-19 Data in Indonesia During 2020," *Int. J. Math. Trends Technol.*, vol. 67, no. 6, pp. 136–144, Jun. 2021, doi: 10.14445/22315373/IJMTT-V67I6P516.

[35] M. D. Pasarella, S. Sifriyani, and F. D. T. Amijaya, "NONPARAMETRIK REGRESSION MODEL ESTIMATION WITH THE FOURIER SERIES THE FOURIER SERIES APPROACH AND ITS APPLICATION TO THE ACCUMULATIVE COVID-19 DATA IN INDONESIA," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 4, pp. 1167–1174, Dec. 2022, doi: 10.30598/barekengvol16iss4pp1167-1174.

[36] L. R. Cheuruiyot, G. O. Orwa, and O. R. Otieno, "Kernel Function and Nonparametric Regression Estimation: Which Function is Appropriate?," *African J. Math. Stat. Stud.*, vol. 3, no. 3, pp. 51–59, 2020, [Online]. Available: https://www.researchgate.net/publication/342832915_KERNEL_FUNCTION_AND_NONPARAMETRIC_REGRESSION_ESTIMATION_WHICH_FUNCTION_IS_APPROPRIATE

[37] A. Tripena, "Analisis Regresi Spline Kuadratik," *Semin. Nas. Mat. dan Pendidik. Mat.*, 2011, [Online]. Available: https://eprints.uny.ac.id/7288/

[38] D. Yosefanny, H. Yozza, and I. Rahmi HG, "Model Spline Kuadratik Untuk Merancang Kurva Pertumbuhan Balita Di Kota Padang," *J. Mat. UNAND*, vol. 7, no. 1, pp. 33–42, 2018, doi: 10.25077/jmu.7.1.52-58.2018.

[39] S. S. Rajak, S. Ismail, and R. Resmawan, "Metode Conditional Autoregressive dalam Analisis Penyebaran Kasus Penyakit Tuberculosis," *Jambura J. Probab. Stat.*, vol. 2, no. 1, pp. 28–34, Apr. 2021, doi: 10.34312/jjps.v2i1.9771.

[40] H. Jusuf, M. Sakti, I. Husein, M. Elveny, R. Syah, and S. Tuba, "Modelling optimally to the treatment of TB patients for increase medical knowledge," *Syst. Rev. Pharm.*, vol. 11, no. 4, pp. 742–748, 2020.

[41] K. Kristy, J. Jajang, and N. Nurhayati, "Analisis Regresi Count Data untuk Pemodelan Jumlah Kasus Penyakit Tuberkulosis di Kabupaten Banyumas," *J. Ilm. Mat. dan Pendidik. Mat.*, vol. 13, no. 2, p. 57, Dec. 2021, doi: 10.20884/1.jmp.2021.13.2.4919.

[42] T. D. Hartanto, L. D. Saraswati, M. S. Adi, and A. Udiyono, "Analisis Spasial Persebaran Kasus Tuberkulosis Paru di Kota Semarang Tahun 2018," *J. Kesehat. Masy.*, vol. 7, no. 4, pp. 2356–3346, 2019, [Online]. Available: http://ejournal3.undip.ac.id/index.php/jkm

[43] W. H. O. WHO, "Diarrhoeal disease," 2017. https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease (accessed May 10, 2023).

[44] V. Mogharab and S. Rajput, "Factors associated with diarrhea in children under 12 years of age referred to Ostad Motahari hospital of Jahrom in 2020," *J. Fam. Med. Prim. Care*, vol. 11, no. 10, p. 6170, 2022, doi: 10.4103/jfmpc.jfmpc_342_22.

[45] A. Sokic-Milutinovic, A. Pavlovic-Markovic, R. S. Tomasevic, and S. Lukic, "Diarrhea as a Clinical Challenge: General Practitioner Approach," *Dig. Dis.*, vol. 40, no. 3, pp. 282–289, 2022, doi: 10.1159/000517111.

[46] T. Mulatu, N. B. Yimer, B. Alemnew, M. Linger, and M. L. Liben, "Exclusive breastfeeding lowers the odds of childhood diarrhea and other medical conditions: evidence from the 2016 Ethiopian demographic and health survey," *Ital. J. Pediatr.*, vol. 47, no. 1, p. 166, Dec. 2021, doi: 10.1186/s13052-021-01115-3.

[47] D. Haile and S. Biadgilign, "Higher breastfeeding performance index is associated with lower risk of illness in infants under six months in Ethiopia," *Int. Breastfeed. J.*, vol. 10, no. 1, p. 32, Dec. 2015, doi: 10.1186/s13006-015-0057-2.

[48] R. S. Imran, R. Resmawan, N. Achmad, and A. R. Nuha, "SEIPR-Mathematical Model of the Pneumonia Spreading in Toddlers with Immunization and Treatment Effects," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 2, pp. 202–218, Dec. 2020, doi: 10.20956/jmsk.v17i2.11166.

[49] N. Achmad, M. R. F. Payu, and Y. Rahim, "Pemodelan Pneumonia Berat Menggunakan Regresi Zero Inflated Negative Binomial di Gorontalo," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 10, no. 1, pp. 45–53, May 2022, doi:

10.34312/euler.v10i1.13990.

[50] A. L. O. Martins, D. da S. F. Nascimento, I. J. C. Schneider, and F. Schuelter-Trevisol, “Incidence of community-acquired infections of lower airways among infants,” *Rev. Paul. Pediatr. (English Ed.*, vol. 34, no. 2, pp. 204–209, Jun. 2016, doi: 10.1016/j.rppede.2015.10.005.

[51] P. A. W. Karmany, S. S. Rahardjo, and B. Murti, “Effect of Low Birth Weight on the Risk of Pneumonia in Children Under Five: Meta-Analysis,” 2020. doi: 10.26911/the7thicph.01.61.

[52] V. N. Sutriana, M. N. Sitaresmi, and A. Wahab, “Risk factors for childhood pneumonia: a case-control study in a high prevalence area in Indonesia,” *Clin. Exp. Pediatr.*, vol. 64, no. 11, pp. 588–595, Nov. 2021, doi: 10.3345/cep.2020.00339.

[53] A. Adityaningrum and H. Jusuf, “Forecasting Simulation of Indonesia COVID-19 Confirmed Cases,” *J. Heal. Sci. Gorontalo J. Heal. Sci. Community*, vol. 5, no. 3, pp. 10–16, Mar. 2022, doi: 10.35971/gojhes.v5i3.13029.

[54] I. Djakaria and S. E. Saleh, “Covid-19 forecast using Holt-Winters exponential smoothing,” *J. Phys. Conf. Ser.*, vol. 1882, no. 1, p. 012033, May 2021, doi: 10.1088/1742-6596/1882/1/012033.

[55] V. Galasso, V. Pons, P. Profeta, M. Becher, S. Brouard, and M. Foucault, “Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries,” *Proc. Natl. Acad. Sci.*, vol. 117, no. 44, pp. 27285–27291, Nov. 2020, doi: 10.1073/pnas.2012520117.

[56] Y. Aleksanyan and J. P. Weinman, “Women, men and COVID-19,” *Soc. Sci. Med.*, vol. 294, p. 114698, Feb. 2022, doi: 10.1016/j.socscimed.2022.114698.

[57] N. Dehingia and A. Raj, “Sex differences in COVID-19 case fatality: do we know enough?,” *Lancet Glob. Heal.*, vol. 9, no. 1, pp. e14–e15, Jan. 2021, doi: 10.1016/S2214-109X(20)30464-2.

[58] A. Doerre and G. Doblhammer, “The influence of gender on COVID-19 infections and mortality in Germany: Insights from age- and gender-specific modeling of contact rates, infections, and deaths in the early phase of the pandemic,” *PLoS One*, vol. 17, no. 5, p. e0268119, May 2022, doi: 10.1371/journal.pone.0268119.

[59] M. M. Robertson *et al.*, “The Epidemiology of Long Coronavirus Disease in US Adults,” *Clin. Infect. Dis.*, vol. 76, no. 9, pp. 1636–1645, May 2023, doi: 10.1093/cid/ciac961.

[60] A. Subramanian *et al.*, “Symptoms and risk factors for long COVID in non-hospitalized adults,” *Nat. Med.*, vol. 28, no. 8, pp. 1706–1714, Aug. 2022, doi: 10.1038/s41591-022-01909-w.

[61] S. le Cessie, R. H. H. Groenwold, and O. M. Dekkers, “Correlation or regression, that’s the question,” *Eur. J. Endocrinol.*, vol. 184, no. 6, pp. E15–E18, Jun. 2021, doi: 10.1530/EJE-21-0251.

[62] P. Schober, C. Boer, and L. A. Schwarte, “Correlation Coefficients: Appropriate Use and Interpretation,” *Anesth Analg*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.

[63] S. E. Tanni, C. M. Patino, and J. C. Ferreira, “Correlation vs. regression in association studies,” *J. Bras. Pneumol.*, vol. 46, no. 1, 2020, doi: 10.1590/1806-3713/e20200030.

[64] F. Z. Okwonu, B. L. Asaju, and F. I. Arunaye, “Breakdown Analysis of Pearson Correlation Coefficient and Robust Correlation Methods,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 917, no. 1, p. 012065, Sep. 2020, doi: 10.1088/1757-899X/917/1/012065.

[65] A. T. R. Dani, N. Y. Adrianingsih, and A. Ainurrochmah, “Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi,” *Jambura J. Probab. Stat.*, vol. 1, no. 2, pp. 98–106, Nov. 2020, doi: 10.34312/jjps.v1i2.7755.

[66] N. Y. Adrianingsih, I. N. Budiantara, and J. D. T. Purnomo, “Mixture Model Nonparametric Regression and Its Application,” *J. Phys. Conf. Ser.*, vol. 1842, no. 1, p. 012044, Mar. 2021, doi: 10.1088/1742-6596/1842/1/012044.

[67] S. Suparti, R. Santoso, A. Prahutama, and A. R. Devi, *Regresi Nonparametrik : Buku ajar*, 1st ed. Ponorogo: WADE Group, 2018. [Online]. Available: https://opac.perpusnas.go.id/DetailOpac.aspx?id=1262978#

[68] S. C. Junianingsih and Y. Karyana, “Regresi Nonparametrik Kernel dalam Pemodelan Jumlah Kelahiran Bayi di Jawa Barat Tahun 2017,” *Bandung Conf. Ser. Stat.*, vol. 1, no. 1, pp. 30–37, Dec. 2021, doi: 10.29313/bcss.v1i1.39.

[69] H. I. Zebua, “Pemodelan Kemiskinan di Sumatera Utara Menggunakan Regresi Nonparametrik Kernel dan Splines,” *Semin. Nas. Off. Stat.*, vol. 2021, no. 1, pp. 899–907, Nov. 2021, doi: 10.34123/semnasoffstat.v2021i1.1087.

[70] I. N. Budiantara, “Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika Yang Mandiri Dan Berkarakter,” *Pros. Semin. Nas. FMIPA Undiksha*, pp. 9–28, 2011, [Online]. Available: https://ejournal.undiksha.ac.id/index.php/semnasmipa/article/view/2736

[71] B. Suwasono, S. Widjaja, A. Zubaydi, Z. Yuliadi, and I. N. Budiantara, “Development of Technology Parameter Towards Shipbuilding Productivity Predictor Using Cubic Spline Approach,” *Makara J. Technol.*, vol. 14, no. 2, May 2010, doi: 10.7454/mst.v14i2.704.

[72] I. A. S. Intansari and I. N. Budiantara, “Inferensi Statistik untuk Kurva Regresi Nonparametrik Spline Kuadratik dan Aplikasinya pada Data ASFR (Age Spesific Fertility Rate) di Bali,” Institut Teknologi Sepuluh Nopember, 2016. [Online]. Available: https://repository.its.ac.id/51371/

[73] I. Sriliana, I. N. Budiantara, and V. Ratnasari, “A Truncated Spline and Local Linear Mixed Estimator in Nonparametric Regression for Longitudinal Data and Its Application,” *Symmetry (Basel).*, vol. 14, no. 12, p. 2687, Dec. 2022, doi: 10.3390/sym14122687.

[74] I. F. Yuliati and P. R. Sihombing, “Pemodelan Fertilitas Di Indonesia Tahun 2017 Menggunakan Pendekatan Regresi Nonparametrik Kernel dan Spline,” *J. Stat. dan Apl.*, vol. 4, no. 1, pp. 48–60, Jun. 2020, doi: 10.21009/JSA.04105.

[75] S. M. T. Marina and I. N. Budiantara, “Pemodelan Faktor-Faktor yang Mempengaruhi Persentase Kriminalitas di Jawa Timur dengan Pendekatan Regresi Semiparametrik Spline,” *J. Sains dan Seni ITS*, vol. 2, no. 2, pp. 147–152, 2013, [Online]. Available: https://ejurnal.its.ac.id/index.php/sains_seni/article/view/4512

[76] S. Wulandary and D. I. Purnama, “Perbandingan Regresi Nonparametrik Kernel dan B-Splines pada Pemodelan Rata-Rata Lama Sekolah dan Pengeluaran Perkapita di Indonesia,” *Jambura J. Probab. Stat.*, vol. 1, no. 2, pp. 89–97, Nov. 2020, doi: 10.34312/jjps.v1i2.7501.

[77] A. T. R. Dani, F. B. Putra, M. A. Zen, V. Ratnasari, and Q. Q. A’yun, “Fourier Series Application for Modeling ‘Chocolate’ Keyword Search Trends in Google Trends Data,” *J. Stat. Univ. Muhammadiyah Semarang*, vol. 11, no. 2, pp. 1–9, 2023, [Online]. Available: https://jurnal.unimus.ac.id/index.php/statistik/article/view/11895

[78] Kuzairi, Miswanto, and I. N. Budiantara, “Three form fourier series estimator semiparametric regression for longitudinal data,” *J. Phys. Conf. Ser.*, vol. 1538, no. 1, May 2020, doi: 10.1088/1742-6596/1538/1/012058.

[79] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, no. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.