Kernel Nonparametric Regression Modeling with the Nadaraya-Watson Estimator (Case Study: Fertility in the Southern Sumatra Region)

Indah Wahyuliani (1), Muhammad Arib Alwansyah (2), Dyah Setyo Rini (3), Winalia Agwil (4)

[1,2,3,4] Statistics Study Program, Bengkulu University, Bengkulu

Road WR. Supratman, Kandang Limun Village, Muara Bangkahulu District, Bengkulu City
e-mail: indahwahyuliani.9e@gmail.com(1), muhammadaribalwanyah232@gmail.com(2), dyah.setyorini@unib.ac.id(3), winaliaagwil@unib.ac.id(4)

ABSTRAK

Fertilitas adalah kelahiran hidup (*live birth*) yaitu terlepasnya bayi dari rahim seorang perempuan dengan adanya tanda-tanda kehidupan seperti berteriak, bernafas, jantung berdenyut, dan sebagainya. Sumber data penelitian ini berasal dari publikasi website resmi Badan Pusat Statistika (BPS). Penelitian ini bertujuan untuk memodelkan dan memprediksi data fertilitas pada tahun 2020 dengan regresi nonparametrik kernel dengan estimator *Nadaraya-Watson*. Model nonparametrik kernel menunjukkan hubungan fertilitas (Y) dengan persentase usia perkawinan pertama dibawah umur (X_1), persentase wanita 15-49 tahun yang tidak menggunakan KB/alat tradisional (X_2) jumlah peserta KB aktif (X_3), jumlah pasangan usia subur (X_4), persentase rata-rata lama sekolah (X_5), dan jumlah pengeluaran perkapita (X_6) berdasarkan nilai *bandwidth* yang hasilkan dan fungsi kernel *Gaussian*. Berdasarkan hasil analisis diperoleh variabel bebas yang berpengaruh signifikan yaitu X_1, X_3, X_4, X_5terhadap variabel terikat (Y) dengan nilai *bandwidth* optimum sebesar 0,490 dan nilai R^2 sebesar 99,6% serta nilai MSE sebesar 0,332. Pemodelan fertilitas penting karena dapat membantu memahami dan memprediksi tren populasi. Hal ini memberikan wawasan tentang potensi tingkat kelahiran dalam suatu populasi di masa depan. Informasi ini dapat digunakan untuk perencanaan kebijakan, termasuk kesehatan, pendidikan, dan kebijakan sosial.

Kata kunci: Fertilitas, Regresi Nonparametrik Kernel, MSE, *Nadaraya-Watson*, R^2.

ABSTRACT

Fertility is a live birth, namely the release of a baby from a woman's womb with signs of life such as screaming, breathing, a throbbing heart, and so on. The source of this research data comes from the publication of the official website of the Central Statistics Agency (BPS). This study aims to model and predict fertility data in 2020 with kernel nonparametric regression using the Nadaraya-Watson estimator. The nonparametric kernel model shows the relationship between fertility (Y) and the percentage of underage women at first marriage (X_1), *the percentage of women 15-49 years who do not use traditional KB or conventional methods* (X_2), *the number of active family planning participants* (X_3), *the number of couples of childbearing age* (X_4), *the percentage of the average length of schooling* (X_5), *and the total expenditure per capita* (X_6) *based on Gaussian kernel function and bandwidth values. Based on the results of the analysis, the independent variables that have a significant effect are* X_1, X_3, X_4, X_5 *on the dependent variable with the optimum bandwidth value of 0.490 and the value of* R^2 *of 99.6%, and the MSE value of 0.332. Modeling fertility is important as it helps understand and predict population trends. It provides insights into the potential number of births in a population in the future. This information can be used for policy planning, including health, educations, and social policies.*

Keywords: *Fertility, Kernel Nonparametric Regression, MSE, Nadaraya-Watson,* R^2

INTRODUCTION

A nonparametric approach is an analysis that does not focus on assuming a particular curve shape, there by providing greater flexibility, the data is expected to find its own estimated shape. In carrying out nonparametric analysis, it is necessary to estimate the nonparametric regression function, which is carried out based on smoothing techniques [1]. One of the estimators used to estimate an unknown regression function is the Nadaraya-Watson estimator. The Nadaraya-Watson estimator has clear generalizations for explanatory variables and kernel functions, and Nadaraya and Watson in 1964 defined a kernel regression estimator called the Nadaraya-Watson estimator, or Nadaraya-Watson kernel estimator [2]. Choosing nonparametric kernel regression for fertility data is advantageous due to the inherent complexity of fertility patterns, which often deviate from linear or parametric assumptions. Nonparametric models, facilitated by kernels, offer increased flexibility in capturing intricate variations within fertility data. This approach is particularly suitable when the relationships influencing fertility exhibit nonlinear characteristics, allowing the model to adapt without being confined to a predefined functional form. Additionally, kernel regression tends to be more robust in the presence of outliers or minor fluctuations commonly found in fertility datasets. Another notable advantage is that nonparametric methods alleviate the need to assume a specific function form, making them well-suited for handling fertility data with unpredictable patterns. However, it is essential to consider challenges such as the proper tuning of kernel parameters and the requirement for an adequate amount of data to support a nuanced model. The selection of nonparametric kernel regression should align with the specific characteristics and complexities inherent in the fertility data under analysis.

The population in the southern Sumatra region continues to increase. This can be seen from the results of the 2000 population census, the population at that time was 18.52 million people. Then, in the 2010 population census, it increased to 21.08 million people [3]. As of the 2020 population census, the population in the Southern Sumatra region reached 24.49 million people [4]. If the population increase is not balanced with an increase in economic capacity, it will have an impact on the welfare of the population in a country, such as high rates of poverty, unemployment, crime, and worsening other social conditions. This is in line with the research conducted by Arum and Haris Which says that population growth is caused by tree components, one of which is fertility [5]. This problem can be limited by controlling the birth rate (fertility). Fertility is a live birth, namely the release of a baby from a woman's womb with signs of life, for example, screaming, breathing, heartbeat, and so on [6]. The relationship between fertility and several factors that influence it has an unknown regression curve shape, so a nonparametric approach is used to estimate the unknown regression function [7].

In the context of fertility data, nonparametric kernel regression offers significant advantages. Fertility, influenced by a myriad of complex factors, often exhibits patterns that are challenging to capture linearly or parametrically. Nonparametric kernel methods provide the flexibility needed to address the intricate relationships among variables in fertility data. This allows the model to adapt dynamically to changes in fertility patterns over time without relying on specific assumptions about the functional form. Moreover, nonparametric kernel regression can identify local patterns in the data, understanding regional or group variations in fertility characteristics. By leveraging kernel techniques, research on fertility data can more effectively manage uncertainty that may arise and gain a deeper understanding of factors influencing fertility rates without sacrificing model complexity [8].

METHOD

Nonparametric regression is one of the methods used to estimate the relationship pattern between the free variable and the bound variable, where the form of its regress curve is unknown [9]. Nonparametric regression does not require assumptions to be met, and the data is expected to find its own form of estimation so that it has high flexibility. Estimation of regression functions is performed based on observational data using certain recovery techniques. Therefore, the systematic model of nonparametric regression can be written as follows:

$$y_i = m(x_i) + \varepsilon_i \qquad i = 1,2, \dots, n.$$

With ε_I is a random variable that is assumed to be independent with an average of zero and a variance σ^2. The function $m(x)$ It is an unknown function called the regression function. In nonparametric regression, there is no assumption of the form of the regression function m(x), thus providing flexibility in possible forms of tre regression function [7]. There are several techniques for reducing the value of a response in response to nonparametric, namely, kernel, spline, local polynomial, fourier dye, and wavelet [10]. Although nonparametric regression is used to overcome data modeling that does not form a particular pattern of relationships, the nonparametric regression model can still be used to model data that has a linear or nonlinear pattern because of the absence of assumptions to be met [9]. One way to estimate the form of relationship between the free variables and the bound variable is by looking at the form of relationship patterns on the scatter diagram or *scatterplot*. By knowing the pattern of relationships that formed, the appropriate approach can be determined to estimate the regression function [11].

The kennel density function is one of the nonparametric methods to determine the probability density function of a random variable. An estimator of kernel density is a development of a histogram estimator. This method is often used because it has a more flexible form. A good density function has smooth functionality variances in sampling are not great and important information from the data is not lost [9]. So it is required to be a constitutive estimator of the function. Kennel density for density functions f(x) and f(x,y) is defined as follows [12]:

$$\hat{f}_x(x) = \frac{1}{nh_x}\sum_{i=1}^{n} K\left(\frac{x - X_i}{nh_x}\right) \tag{1}$$

$$\hat{f}(x, y) = \frac{1}{nh_x h_y}\sum_{i=1}^{n} K_x\left(\frac{x - X_i}{h_x}\right) K_y\left(\frac{y - Y_i}{h_y}\right) \tag{2}$$

The nonparametric regression model of the kernel can be expressed as follows:

$$Y_i = m(x_i) + \varepsilon_i \tag{3}$$

Regarding kernel, is one of the nonparametric techniques to estimate conditional expectations of random variables with the aim of finding relationships between free variables and variables bound by using kernel function weight [13]. Conditional expectations of the bound variable relative to the free variable can be written as follows:

$$m(x) = E(Y|X = x)$$

$$m(x) = \int_{-\infty}^{\infty} y f(y|x) dy$$
$$m(x) = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_x(x)} dy \quad (4)$$

Estimation $m(x)$ obtained from substitution of equation (1) and (2) to equation (4)

$$\hat{m}(x) = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_x(x)} dy$$
$$\hat{m}(x) = \frac{1}{f_x(x)} \int_{-\infty}^{\infty} y f(x,y) dy$$
$$\hat{m}(x) = \frac{1}{f_x(x)} \int_{-\infty}^{\infty} y \frac{1}{nh_x h_y} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) K_y\left(\frac{y-Y_i}{h_y}\right) dy \quad (5)$$
$$\hat{m}(x) = \frac{1}{f_x(x)} \frac{1}{nh_x} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) \int_{-\infty}^{\infty} \frac{y}{h_y} K_y\left(\frac{y-Y_i}{h_y}\right) dy \quad (6)$$

Suppose $u = \frac{y-Y_i}{h_y}$, then:

$$\hat{m}(x) = \frac{1}{f_x(x)} \frac{1}{nh_x} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) \int_{-\infty}^{\infty} (u+Y_i)\, K_y(u) dy \quad (7)$$
$$\hat{m}(x) = \frac{1}{f_x(x)} \frac{1}{nh_x} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) \int_{-\infty}^{\infty} u K_y(u)\, du + Y_i \int_{-\infty}^{\infty} K_y(u)\, du \quad (8)$$

By using the kernel properties $\int_{-\infty}^{\infty} K(u) du = 1$ and $\int_{-\infty}^{\infty} uK(u) du = 0$, so obtained estimator for $m(x)$ as follows:

$$\hat{m}(x) = \frac{\frac{1}{nh_x} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) Y_i}{f_x(x)} \quad (9)$$

On kernel regression, there is a known estimator that can be used to estimate the regression function estimator, *Nadaraya-Watson*. The density function $f_x(x)$ at equation (1) is subtitized to equation (9) it will be obtained by the estimator *Nadaraya-Watson* with $h_x = h$ and $K_x = K$ as follows:

$$\hat{m}(x) = \frac{\frac{1}{nh_x} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) Y_i}{\frac{1}{nh_x} \sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right)}$$
$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} \quad (10)$$

Estimator *Nadaraya-Watson* is an estimator used to estimatead a local weighted average by using the kernel as a weighting function. Estimator *Nadaraya-Watson* multivariate forms can be written in the same equation [14]:

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} \prod_{j=1}^{d} K(\frac{x_{ij} - x_j}{h}) Y_i}{\sum_{i=1}^{n} \prod_{j=1}^{d} K(\frac{x_{ij} - x_j}{h})} \tag{11}$$

The main philosophy of nonparametric regression is to estimate the regression function m(x) using the weighted average of raw data, where the weight is a distance function in space-x. In particular, the weight is a distance decline function. This type of weighting scheme is proposed by *Nadaraya-Watson* (1964), where the weighting is associated with observation Y_I, for predictions x_I obtained from:

$$W_i(x) = \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)} = \frac{K(u)}{\sum_{j=1}^{n} K(u)} \tag{12}$$

where the K(u) is a declining function of u and $h >$ 0 called *bandwidth* or the preparation parameters. K(u) is the kernel function that can be considered as a function of density, like opportunities such as *Gaussian*. $W_I(x)$ is a position of positive weight and has characteristics $\sum_{I=1}^{n} W_I(x) = 1$, then obtained:

$$\widehat{m}(x) = \sum_{i=1}^{n} W_i(x) Y_i \quad , i = 1,2,\ldots,n \tag{13}$$

with each one n data are given different weights [14]. So that *Nadaraya-Watson* is the average weight of Y_I.

RESULT AND DISCUSSION

The presentation of data with descriptive statistics aims to give a general description of the percentage of the birth fertility of people living in the Sumbagsel region (Y) In 2020, the percentage of the age of marriage was first age (X_1), the percentage of women 15-49 years who do not use KB / traditional tools (X_2), the number of KB participants (X_3), the number of feet of fertile age (X_4), the average percentage of the school's long (X_5), the amount of perpaitable expenditure (X_6). The amount of data used in this study was as many as 60, consisting of districts and cities in the Sumbagsel region.

Tabel 1. Statistics Descriptive

Variable	Minimum	Median	*Mean*	Maximum	Standard Deviation
Y	33,490	66,100	64,500	87,540	9,782
X_1	6,180	15,230	15,240	28,090	4,430
X_2	14,830	29,160	31,310	58,160	10,058
X_3	12.308,000	42.727,000	60.964,000	246.345,000	51.565,970
X_4	9,000	55.263,000	74.881,000	345.802,000	73.043,590
X_5	6,350	7,865	8,209	11,790	1,158
X_6	7.892,000	10.220,000	10.499,000	15.663,000	1.699,560

Based on Table 4.1, average birthday living (Y) by 2020 will be 64,500%, with the lowest percentage of 33,490% in the city of Pangkal Pinang and the highest percentage of the baby's birth of the highest life of 87.540% in the city of Tanggamus, with the diversity of the baby's diet living in every district or city in the region of Sumbagsel by 9.782%. Then the average percentage of the age of marriage is first under age (X_1) By 2020 that is 15.240%. The percentage of the first major marriage age was 28.090% in Kerinci District, while the percentage of the first marriage age was below the lowest of 6,180% in the city of Jambi, with a diversity rate of 4,430%.

At the percentage of women 15-49 years old who do not use KB or Traditional tools (X_2) Having a diversity rate of 10.058% and an average of 31.310%, with the lowest percentage value of 14.830% in the rotary of the rotary day and the highest percentage value of 58,160% found in the Regency of the Belitung East. Then the diversity rate on the number of participant KB participants (X_3) 51.565.970, with the average user of KB active of 60.964.000 It has a minimum value of 12.308.000 people in the full river district and a maximum population of 246.345.000 people in Central Lampung Regency.

As for the average PUS (X_4) in the Sumbagsel region of 74.881.000 with the level of diversity of PUS 73.043.590 and It has a minimum value of 9,000 pairs contained in West Bangka regency while the maximum value of 345.802.000 pairs contained in Central Lampung Regency. Then the diversity rate of the RSL percentage (X_5) of 1.156% with a minimum value of 6.350% there is in Tanjung Jabung Timur Regency and the maximum value of 11,790% found in Bengkulu City and the average percentage of RLS in the Sumbagsel region of 8.209%. Average perpendicular expenditure (X_6) In the Sembagsel region of 10.499.000 with minimum perkapita expenditure of 7.892.000 in Pesargan District while the perpendapa expenditure Maximum of 15.663.000 in the Pangkal Pinang city and the diversity of perpenditable expenditure amount 1.699.563.

Tabel 2. Test Linearity and Identification Data Pattern.

Variable	P-*Value*	Information
X_1 terhadap Y	0,086	Nonlinier
X_2 terhadap Y	0,562	Nonlinier
X_3 terhadap Y	0,365	Nonlinier
X_4 terhadap Y	0,233	Nonlinier
X_5 terhadap Y	0,000	Linier
X_6 terhadap Y	0,000	Linier

Based on the result of the linearity test and the identification of data patterns with a *scatterplot*, the variable X_1,X_2, X_3, $X_4$and X_5 it is a nonparametric component because these four variables do not contain previous function-form information that can be shown from data patterns that do not form certain patterns. Where variables X_6 are a parametric component for forming a linear data pattern. Election Optimum Bandwidth

Tabel 3. Test Linearity and Identification Data Pattern.

No	***Bandwidth*** **(h)**	GCV
1	0,4900	0,0000
2	0,5000	0,0001
3	0,0000	0,5100

4	0,0100	0,5200
5	0,0200	0,5300
6	0,0300	0,5400
7	0,0400	0,5500
8	0,0500	0,5600
9	0,0600	0,5700
10	0,0700	0,5800

Based on table 4.5, it can be seen that the minimum GCV value is equal 0,0000 there is on *bandwidth* 0.490 so$h$ =0.490 is the optimal *bandwidth*. The amount of value *bandwidth* Optimum is then used in the kernel method by means of subtituting the value *bandwidth* on the estimator *Nadaraya-Watson*. Based on the election *bandwidth* Optimum using GCV approach then obtained value *bandwidth* Optimum of 0.490 and by using variables Nonparameteric is the fertility of living babies of life, the percentage of age of marriage is first age (X_1) ,Percentage of women 15-49 years who do not use KB / Traditional tools (X_2), the number of participant KB participants (X_3), the number of feet of fertile age (X_4), the estimated result is obtained by the mesignitual kennel function of estimator as *Nadaraya-Watson*as follows:

$$\widehat{m}(x_i) = \frac{\sum_{i=1}^{n} \prod_{j=1}^{d} (\frac{1}{\sqrt{2\pi}} exp\,(-\frac{1}{2}(\frac{x_{ij}-x_j}{h_j})^2) Y_i}{\sum_{i=1}^{n} \prod_{j=1}^{d} (\frac{1}{\sqrt{2\pi}} exp\,(-\frac{1}{2}(\frac{x_{ij}-x_j}{h_j})^2)}$$

So it is obtained the following equation:

$$\widehat{m}(x_i) = \frac{\sum_{i=1}^{60} \prod_{j=1}^{5} (\frac{1}{\sqrt{2\pi}} exp\,(-\frac{1}{2}(\frac{x_{ij}-x_j}{0,490})^2) Y_i}{\sum_{i=1}^{60} \prod_{j=1}^{5} (\frac{1}{\sqrt{2\pi}} exp\,(-\frac{1}{2}(\frac{x_{ij}-x_j}{0,490})^2)}$$

After the kernel estimation, *Nadaraya-Watson* With the Gaussian function, the form of the nonparametric regression model as follows:

$$Y_i = m(x_i) + \varepsilon_i$$

$$Y_i = \frac{\sum_{i=1}^{60} \prod_{j=1}^{5} (\frac{1}{\sqrt{2\pi}} exp\,(-\frac{1}{2}(\frac{x_{ij} - x_j}{0,490})^2) Y_i}{\sum_{i=1}^{60} \prod_{j=1}^{5} (\frac{1}{\sqrt{2\pi}} exp\,(-\frac{1}{2}(\frac{x_{ij} - x_j}{0,490})^2)} + \varepsilon_i$$

$$Y_i = \frac{\sum_{i=1}^{60}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i1}-x_1}{0{,}490}\right)^2\right)Y_i\right)}{\sum_{i=1}^{60}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i1}-x_1}{0{,}490}\right)^2\right)} + \frac{\sum_{i=1}^{60}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i2}-x_2}{0{,}490}\right)^2\right)Y_i\right)}{\sum_{i=1}^{60}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i2}-x_2}{0{,}490}\right)^2\right)}$$
$$+ \frac{\sum_{i=1}^{60}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i3}-x_3}{0{,}490}\right)^2\right)Y_i\right)}{\sum_{i=1}^{60}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i3}-x_3}{0{,}490}\right)^2\right)}$$
$$+ \frac{\sum_{i=1}^{60}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i4}-x_4}{0{,}490}\right)^2\right)Y_i\right)}{\sum_{i=1}^{60}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i4}-x_4}{0{,}490}\right)^2\right)}$$
$$+ \frac{\sum_{i=1}^{60}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i5}-x_5}{0{,}490}\right)^2\right)Y_i\right)}{\sum_{i=1}^{60}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_{i5}-x_5}{0{,}490}\right)^2\right)} + \varepsilon_i$$

The kernel nonparametric regression model shows the form of the weight of the Gaussian kernel function with a scaling size of the weighting function of 0.490, which means that the estimated curve fits the fertility data pattern in the optimum weight measure of 0.490. If we do not use the optimum weight, the following conditions will occur: if the weight size used is less than 0.490 then the estimated curve will undersmooth, if the weight size used is greater than 0.490 then the estimated curve will oversmooth.

CONCLUSION

Based on the non-parametric regression analysis of the kernel estimated with the estimator *Nadaraya-Watson* In the case of fertility in the Sumbagsel region by 2020, we will obtain a value *bandwidth* optimum of 0.490 and the value of the determination coefficient of 99.65%, and the independent variable that significantly influences on the bound variable is variable X_1,X_3,X_4, and X_5that is the percentage of women who had married first age, the number of participant KB participants, the number of feet of fertile age, and the average percentage of the school's sstudent's.

REFERENCE

[1] N. E. Chandra, S. Haryatmi, and Z. Zulaela, "Regresi Nonparametrik Kernel Adjusted," *J. Ilm. Mat. dan Pendidik. Mat.*, vol. 7, no. 1, p. 1, 2015, doi: 10.20884/1.jmp.2015.7.1.2894.

[2] W. Hardle, "Applied Nonparametric Methods," *Handb. Econom.*, vol. IV, no. 26, pp. 2295–2339, 1992.

[3] B. 2010, *Penduduk Indonesia Penduduk Indonesia Hasil SP2010*. Jakarta: Badan Pusat Statistik, 2023.

[4] BPS Indonesia, *Statistik Indonesia Statistical Yearbook Of Indonesia 2020*, vol. 1101001. Jakarta: Badan Pusat Statistik, 2020.

[5] P. rismawati Arum, "Analisis Faktor-Faktor yang Mempengaruhi Jumlah Penduduk di Kota Semarang Menggunakan Metode Regresi Data Panel," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 12, no. 2, pp. 36–41, 2019, doi: 10.36456/jstat.vol12.no2.a2227.

[6] B. 2010, *Fertilitas Penduduk Indonesia*. Jakarta: Badan Pusat Statistik.

[7] M. Abdy, "Tinjauan Singkat Tentang Regresi Parametrik dan Regresi non Parametrik,"

Saintifik, vol. 5, no. 1, pp. 58–62, 2019, doi: 10.31605/saintifik.v5i1.199.
[8] S. Alfiani and P. R. Arum, “Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression,” *J Stat.*, vol. 15, no. 2, pp. 219–227, 2022.
[9] N. Salam, Y. Sukmawaty, and A. Halida, “Estimasi Model Regresi Nonparametrik Dengan Metode B-Spline,” vol. 3, no. 7, pp. 6843–6848, 2022.
[10] I. Puspitasari, Suparti, and Y. Wilandari, “Analisis Indeks Harga Saham Gabungan (IHSG) dengan Menggunakan Model Regresi Kernel,” *J. Gaussian*, vol. 1, no. 1, pp. 93–102, 2012.
[11] A. H. Maharani, H. Yozza, and Y. Asdi, “Pemodelan Berat Badan Balita dengan Menggunakan Regresi Kernel,” *J. Mat. UNAND*, vol. 4, no. 3, p. 31, 2019, doi: 10.25077/jmu.4.3.31-40.2015.
[12] R. Blundell, A. Duncan, and R. Blundell, “Board of Regents of the University of Wisconsin System Kernel Regression in Empirical Microeconomics Kernel Regression in Empirical Microeconomics,” vol. 33, no. 1, pp. 62–87, 2015.
[13] Istiqomatul Fajriyah Yuliati and P. 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, 2020, doi: 10.21009/jsa.04105.
[14] W. Hardle, “Applied Nonparametric Regression.,” *Biometrics*, vol. 50, no. 2, p. 592, 1994, doi: 10.2307/2533418.