

PEMODELAN KADAR HEMOGLOBIN PADA PASIEN DEMAM BERDARAH DI KOTA SAMARINDA MENGGUNAKAN REGRESI SEMIPARAMETRIK SPLINE TRUNCATED

Andrea Tri Rian Dani¹, Fachrian Bimantoro Putra², Muhammad Aldani Zen³, Sifriyani⁴,
Meirinda Fauziyah⁵, Vita Ratnasari⁶, Narita Yuri Adrianingsih⁷

^{1,2,3,4,5} Program Studi Statistika, Jurusan Matematika, Fakultas MIPA, Universitas Mulawarman

⁶ Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember

⁷ Jurusan Matematika, Fakultas MIPA, Universitas Tribuana Kalabahi

e-mail: andreatriandani@fmipa.unmul.ac.id

Abstrak

Artikel ini membahas inovasi pemodelan statistika dalam analisis regresi dengan pendekatan semiparametrik yang diaplikasikan pada data Kesehatan. Regresi merupakan metode dalam ilmu statistik yang mengambil peranan dalam pemodelan statistika. Analisis regresi digunakan memodelkan variabel bebas dengan variabel terikat. Pada regresi, terdapat tiga metode pendekatan yang umum digunakan, regresi nonparametrik, parametrik, dan gabungan keduanya yaitu semiparametrik. Regresi semiparametrik merupakan pendekatan yang digunakan ketika variabel terikat memiliki hubungan yang diketahui bentuknya dengan sebagian variabel bebas, dan memiliki pola hubungan yang tidak diketahui bentuknya dengan sebagian variabel bebas yang lain. Tujuan dari penelitian ini adalah memodelkan kadar hemoglobin pada pasien demam berdarah, dengan variabel bebas yang digunakan adalah jumlah hematokrit (x_1) dan jumlah leukosit (x_2). Regresi semiparametrik spline truncated dipilih sebagai alternatif pemodelan, dikarenakan memiliki banyak kelebihan dalam pemodelan, salah satunya dapat memodelkan pola yang tidak diketahui bentuk hubungannya. Estimasi parameter yang digunakan adalah pendugaan maksimum. Berdasarkan analisis, diperoleh model regresi semiparametrik spline truncated yang diaplikasikan pada data kadar hemoglobin adalah model dengan tiga titik knot yang memiliki nilai koefisien determinasi sebesar 89,074 %. Berdasarkan hasil pengujian hipotesis secara simultan, diperoleh kesimpulan jika secara simultan dan parsial, variabel bebas signifikan berpengaruh.

Kata Kunci: Regresi; Regresi Semiparametrik; Pendugaan Maksimum; Hemoglobin

Abstract

This article discusses the innovation of statistical modeling in regression analysis with a semiparametric approach applied to health data. Regression is a method in statistics that takes a role in statistical modeling. Regression analysis is used to model the independent variable with the dependent variable. In regression, there are three approaches that are commonly used, nonparametric, parametric, and a combination of the two, namely semiparametric. Semiparametric regression is an approach used when the dependent variable has a known relationship with some of the independent variables, and has an unknown pattern of relationship with some of the other independent variables. The purpose of this study was to model hemoglobin levels in dengue fever patients, with the independent variables used being the number of hematocrit (x_1) and the number of leukocytes (x_2). The truncated spline semiparametric regression was chosen as an alternative modeling, because it has many advantages in modeling, one of which is being able to model patterns whose relationship form is unknown. The parameter estimation used is the maximum estimation. Based on the analysis, it was obtained that the truncated spline semiparametric regression model applied to the hemoglobin level data was a three-point knot model with a coefficient of determination of 89.074%. Based on the results of simultaneous

hypothesis testing, it can be concluded that simultaneously and partially, the independent variables have a significant effect.

Keywords: *Regression; Semiparametric Regression; Maximum Likelihood; Hemoglobin*

1. PENDAHULUAN

Metode statistika untuk memodelkan antara variabel terikat dengan variabel bebas adalah regresi (Dani, Ratnasari and Budiantara, 2021). Pada umumnya, terdapat tiga metode pemodelan analisis regresi, diantaranya parametrik, nonparametrik, dan semiparametrik (Igustin and Budiantara, 2021). Regresi parametrik, umumnya digunakan ketika variabel terikat dengan variabel bebas diketahui bentuk hubungannya, katakanlah bentuk linear, eksponensial, kuadratik, dan lainnya. Namun, pendekatan regresi nonparametrik adalah alternatif dari regresi parametrik, yang dimana diterapkan ketika pola hubungannya tidak diketahui dan atau tidak memiliki informasi yang lengkap dan jelas mengenai pola hubungannya (Dani and Ni'matuzzahroh, 2022).

Pada kasus riil-nya dengan variabel bebas lebih dari 1 (multivariabel), seringkali dijumpai kasus yang dimana sebagian pola hubungan antara variabel terikat dengan bebas diketahui, dan sebagian lainnya tidak diketahui, maka regresi semiparametrik adalah alternatif yang dapat digunakan (Nurchayani, Budiantara and Zain, 2021). Apabila menggunakan hanya satu pendekatan saja, maka diindikasikan akan diperoleh hasil estimasi pola hubungan yang tidak tepat dan berakibat kepada akurasi dari model yang terbentuk (Ahmed, Aydin and Yilmaz, 2022).

Kombinasi pendekatan regresi nonparametrik regresi parametrik dengan adalah regresi semiparametrik. Pada kasus-kasus pemodelan, sangat memungkinkan jika variabel terikat memiliki pola hubungan linear dengan beberapa variabel bebas, dan memiliki pola hubungan yang tidak diketahui bentuknya dengan beberapa variabel bebas yang lainnya. Pada kasus seperti ini, Wahba (Wahba and Wang, 2014), Eubank (Eubank, 1999), dan Budiantara (Budiantara *et al.*, 2012) mengusulkan penggunaan regresi semiparametrik. Mengingat pada regresi semiparametrik terdapat komponen parametrik dan nonparametrik. Maka bentuk estimator untuk komponen parametrik dan nonparametrik perlu dibatasi terlebih dahulu. Pada penelitian ini, estimator untuk komponen parametrik adalah regresi linear, sedangkan estimator untuk komponen nonparametrik adalah spline truncated.

Regresi linear adalah salah satu estimator dalam pendekatan parametrik yang sangat umum dan mendasar. Regresi linear digunakan ketika pola hubungan antara variabel terikat dan bebas cenderung membentuk pola hubungan garis linear yang diidentifikasi melalui diagram pencar. Sedangkan estimator spline truncated adalah estimator dalam pendekatan nonparametrik yang memiliki banyak kelebihan, diantaranya dapat memodelkan pola yang tidak diketahui bentuk hubungannya dengan menggunakan bantuan titik knot (Dani, Adrianingsih and Ainurrochmah, 2020). Bentuk kurva regresi dari estimator spline truncated cenderung mencari sendiri bentuk estimasinya. Beberapa penelitian terdahulu yang mengkaji estimator spline truncated diantaranya: (Budiantara *et al.*, 2015; Sifriyani *et al.*, 2018, 2023; Dani, Ratnasari and Budiantara, 2021; Hidayat *et al.*, 2021). Penelitian-penelitian yang telah dirincikan, menggunakan estimator spline truncated pada pendekatan regresi nonparametrik. Namun menjadi hal yang inovatif dan aplikatif jika juga menggeneralisasikan pada pendekatan regresi semiparametrik.

Pendekatan regresi semiparametrik spline truncated tentunya dapat diaplikasikan pada berbagai bidang, khususnya bidang kesehatan. Pada artikel ini, pendekatan regresi semiparametrik digunakan untuk memodelkan kadar hemoglobin pada pasien sebagai variabel terikat. Hemoglobin atau biasa disingkat sebagai Hb merupakan suatu protein yang

ada pada sel darah merah dan yang membuat darah berwarna merah. Hemoglobin tentunya memiliki banyak fungsi yang dibutuhkan oleh tubuh, sedemikian sehingga kadar normal hemoglobin perlu selalu dijaga. Ketika kadar hemoglobin yang ada pada manusia memiliki tingkatan yang lebih tinggi atau lebih rendah dari pada jumlah normal, maka hal ini menjadi adanya tanda adanya gangguan Kesehatan. Pemodelan kadar hemoglobin (y) tentunya dengan mempertimbangkan variabel yang diduga mempengaruhi (x).

2. METODE PENELITIAN

Penelitian ini merupakan penelitian dengan pendekatan kuantitatif. Data yang dianalisis, diperoleh dari salah satu rumah sakit di Samarinda dengan unit observasi pada penelitian ini adalah Pasien. Banyaknya pasien yang dijadikan sampel pada penelitian ini sebanyak 116. Variabel penelitian terdiri atas variabel terikat dan bebas yang dirincikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Notasi	Satuan
Kadar Hemoglobin	y	g/dL
Jumlah Hematokrit	x_1	g/dL
Jumlah Leukosit	x_2	g/dL

Tahapan penelitian pada penelitian ini:

1. Melakukan eksplorasi dengan menampilkan diagram pencar antara variabel terikat dengan semua variabel bebas.
2. Memilih variabel bebas yang akan dimodelkan sebagai komponen parametrik dan komponen nonparametrik.
3. Memodelkan kadar hemoglobin (y) menggunakan model regresi semiparametrik spline truncated. Titik knot yang dicobakan adalah 1, 2, dan 3 titik knot. Misalkan diberikan variabel bebas dan terikat (x_{1i}, x_{2i}, y_i) , sedemikian sehingga model regresi semiparametrik dituliskan pada Persamaan (8).

$$y_i = f(x_{1i}, x_{2i}) + e_i \quad (1)$$

Dengan kurva regresi diasumsikan aditif, sehingga:

$$y_i = f(x_{1i}) + f(x_{2i}) + e_i \quad (2)$$

Dimana y_i adalah variabel terikat, $f(x_{1i})$ adalah bentuk parametrik, $f(x_{2i})$ adalah bentuk nonparametrik spline truncated, e_i adalah *error* random (IIDN $\sim 0, \sigma^2$) (Adrianingsih and Dani, 2021).

$$\tilde{y} = X\tilde{\beta} + G(k)\tilde{\theta} + \tilde{e} \quad (3)$$

Dengan menggunakan metode estimasi parameter pendugaan maksimum, diperoleh estimasi untuk parameter $\tilde{\beta}$ dan $\tilde{\theta}$ sebagai berikut:

$$\tilde{\vartheta} = (M^T M)^{-1} M^T \tilde{y} \quad (4)$$

dimana:

$$\tilde{\vartheta} = \begin{bmatrix} \tilde{\beta} \\ \tilde{\theta} \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} & x_1^1 & \dots & x_1^m & (x_1 - K_1)_+^1 & \dots & (x_1 - K_r)_+^m \\ 1 & x_{12} & \dots & x_{p2} & x_2^1 & \dots & x_2^m & (x_2 - K_1)_+^1 & \dots & (x_2 - K_r)_+^m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{pn} & x_n^1 & \dots & x_n^m & (x_n - K_1)_+^1 & \dots & (x_n - K_r)_+^m \end{bmatrix}$$

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

4. Memilih banyaknya dan lokasi titik knot yang optimal. Poin utama yang berperan penting dalam mendapatkan model terbaik regresi semiparametrik spline truncated adalah proses pemilihan titik knot. Wahba (Wahba and Wang, 1990) memperkenalkan salah satu metode yang dapat digunakan adalah *Generalized Cross-Validation* (GCV). Formula umum GCV adalah:

$$GCV = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n^{-1} \text{trace} [(I - M)])^2} \quad (5)$$

5. Membuat visualisasi y dengan \hat{y} dari model regresi semiparametrik terbaik.
6. Pengujian hipotesis dari model regresi semiparametric terbaik, baik secara simultan dan parsial. Rumusan hipotesis pengujian simultan:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0$$

Dan rumusan hipotesis pengujian parsial:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

7. Interpretasi.

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi Data

Sebelum pemodelan regresi semiparametrik dengan menggunakan spline truncated, statistika deskriptif disajikan pada Tabel 2 di bagian ini.

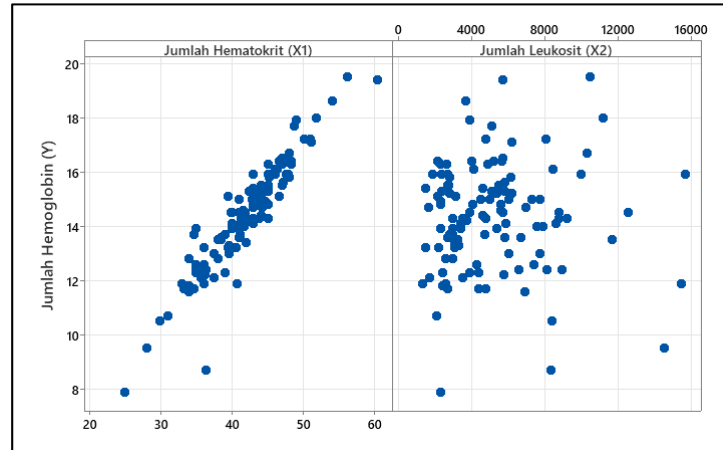
Tabel 2. Statistika Deskriptif

Variabel	Rata-Rata	Minimum	Maksimum
y	14,3	7,9	19,5
x_1	41,8	25	60,4
x_2	5172,7	1310	15680

Pada Tabel 2, memperlihatkan jika masing-masing unit observasi yang dalam hal ini adalah pasien memiliki karakteristik yang cukup bervariasi. Variabel respon (y) yaitu Jumlah Hemoglobin memiliki rata-rata sebesar 14.3 g/dL dengan nilai minimum 7.9 g/dL dan maksimum sebesar 19.5 g/dL. Secara umum, nilai normal kadar hemoglobin didalam tubuh ditentukan berdasarkan usia dan jenis kelamin. Kadar hemoglobin pada Wanita dewasa, normalnya berada pada interval 12-15 g/dL, namun berbeda pada Pria dewasa, normalnya berada pada interval 13-17 g/dL.

3.2 Diagram Pencar

Diagram pencar dapat digunakan sebagai alat identifikasi awal mengenai pola hubungan. Adapun diagram pencar antara variabel (x_1) dan (x_2) terhadap (y) divisualisasikan pada Gambar 1.



Gambar 1. Pola hubungan antara variabel x_1 dan x_2 terhadap variabel y

Pada Gambar 1, pola hubungan antara variabel (x_1) dengan (y) cenderung membentuk suatu garis lurus yang linear. Sedemikian sehingga, pendekatan yang digunakan adalah parametrik dengan estimator regresi linear. Hal yang berbeda terjadi pada variabel (x_2) dengan (y) dimana menunjukkan adanya pola yang acak dan cenderung menyebar, sehingga pendekatan yang digunakan adalah estimator spline truncated.

3.3 Pemodelan Regresi Semiparametrik Spline Truncated

Model regresi semiparametrik spline truncated dibangun karena terdapat komponen parametrik dan nonparametrik pada kasus penelitian ini. Model regresi semiparametrik terbaik dipilih berdasarkan kriteria GCV yang minimum. Titik knot yang digunakan untuk komponen regresi nonparametrik sebanyak 1 - 3 titik knot.

a. Estimasi Parameter dan Pemilihan Titik Knot Optimal

Estimasi parameter menggunakan metode pendugaan maksimum. Selanjutnya, tabel 3 menyajikan lokasi titik knot dan banyaknya titik knot yang optimal dengan nilai GCV minimum.

Tabel 3. Nilai GCV Regresi Semiparametrik

No	Banyaknya Titik Knot	Spline Truncated	GCV
		Titik Knot	
1	1 titik knot	$K_1 = 3106,25$	0,484
2	2 titik knot	$K_1 = 3705,00$ $K_2 = 7896,25$	0,486
3	3 titik knot	$K_1 = 3705,00$ $K_2 = 8495,00$ $K_3 = 9692,50$	0,474

Berdasarkan Tabel 3, nilai GCV minimum diperoleh dengan banyaknya titik knot yang digunakan sebanyak tiga titik knot. Lokasi titik knot pada variabel x_2 yang dimodelkan dengan estimator spline truncated dirincikan sebagai berikut:

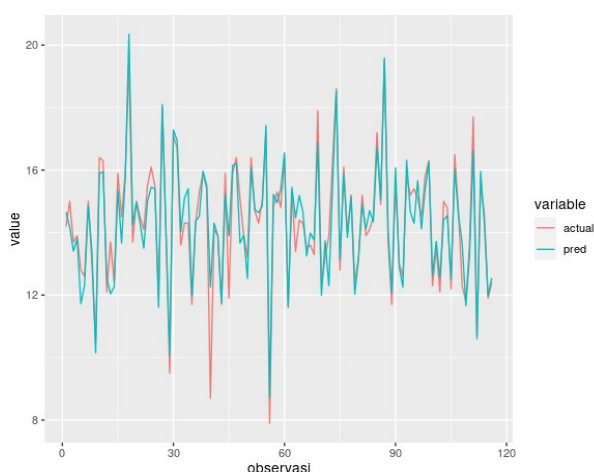
$$K_1 = 3705,00 ; K_2 = 8495,00 ; K_3 = 9692,50$$

Model terbaik dituliskan pada Persamaan (14).

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\xi}_{21}(x_{2i} - K_{12})_+ + \hat{\xi}_{22}(x_{2i} - K_{22})_+ + \hat{\xi}_{23}(x_{2i} - K_{32})_+ \quad (14)$$

Menggunakan metode pendugaan maksimum, diperoleh hasil estimasi parameter untuk Persamaan (14), yang dituliskan pada Persamaan (15).

$$\hat{y}_i = -0,0644 + 0,3231x_{1i} + 0,0003x_{2i} - 0,0004(x_{2i} - K_{12})_+ + 0,0009(x_{2i} - K_{22})_+ - 0,0009(x_{2i} - K_{32})_+ \quad (15)$$



Gambar 2. Visualisasi nilai prediksi (\hat{y}) dari variabel terikat

Gambar 2, merupakan visualisasi nilai prediksi (\hat{y}) dari variabel terikat berdasarkan model terbaik regresi semiparametrik dengan aktual. Pada grafik, terlihat jika nilai prediksi cenderung mengikuti pola dari data aktual, yang diindikasikan jika hasil prediksi untuk variabel terikat cenderung mendekati nilai dari data aktual. Koefisien determinasi sebesar 89,074 %.

b. Pengujian Hipotesis

Pengujian hipotesis secara simultan, dilakukan dengan taraf signifikansi sebesar 10%. Menggunakan statistik uji F dengan rumusan hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0$$

Tabel ANOVA hasil pengujian hipotesis simultan ditampilkan pada Tabel 4.

Tabel 4. Tabel ANOVA

Sumber	df	SS	MS	Fhit
Regresi	5	403,325	80,665	189,144
Error	110	49,470	0,426	
Total	115	452,796		

Berdasarkan Tabel 4, didapatkan $F_{hit} = 189,144$ yang lebih besar dari $F_{(0,1,5,110)} = 1,900$ sehingga diperoleh keputusan tolak H_0 , yang berarti minimal ada satu $\beta_k \neq 0$ atau secara simultan variabel bebas yang digunakan berpengaruh terhadap variabel terikat. Selanjutnya, pengujian hipotesis secara parsial dilakukan. Menggunakan statistik uji t_{hit} dengan rumusan hipotesis:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Hasil pengujian hipotesis secara parsial ditampilkan pada Tabel 5.

Tabel 5. Pengujian Parsial

Parameter	t_{hit}	$P\text{-Value}$	Kesimpulan
$\hat{\beta}_0$	-0,11	$8,39 \times 10^{-1}$	Signifikan
$\hat{\beta}_1$	28,84	$6,56 \times 10^{-53}$	Signifikan
$\hat{\beta}_2$	2,55	1×10^{-2}	Signifikan
$\hat{\xi}_{21}$	-2,74	$6,52 \times 10^{-3}$	Signifikan
$\hat{\xi}_{22}$	2,68	$9,10 \times 10^{-3}$	Signifikan
$\hat{\xi}_{23}$	-2,30	$2,48 \times 10^{-2}$	Signifikan

Berdasarkan Tabel 5, terlihat jika semua parameter pada model regresi semiparametric spline truncated signifikan, hal ini berarti jika variabel bebas memberikan pengaruh yang signifikan terhadap variabel terikat y , yang dalam hal ini adalah kadar hemoglobin pasien.

4. KESIMPULAN

Model regresi nonparametrik menggunakan 3 titik knot untuk komponen spline truncated merupakan model yang terbaik, dengan nilai GCV yang paling minimum sebesar 0,474 dan nilai R^2 sebesar 89,074 %. Berdasarkan hasil pengujian hipotesis, baik secara simultan dan parsial, diperoleh kesimpulan jika minimal ada satu $\beta_k \neq 0$ yang berarti secara simultan variabel bebas signifikan berpengaruh terhadap variabel terikat. Kemudian pada pengujian parsial, diperoleh kesimpulan jika variabel (x_1) dan (x_2) memang benar memberikan pengaruh yang signifikan terhadap variabel terikat (y) .

DAFTAR PUSTAKA

- Adrianingsih, N.Y. and Dani, A.T.R. (2021) 'Estimasi Model Regresi Semiparametrik Spline Truncated Menggunakan Metode Maximum Likelihood Estimation (MLE)', *Jambura Journal of Probability and Statistics*, 2(2), pp. 56–63.
- Ahmed, S.E., Aydin, D. and Yilmaz, E. (2022) 'Semiparametric Time-Series Model Using Local Polynomial: An Application on the Effects of Financial Risk Factors on Crop Yield', *Journal of Risk and Financial Management*, 15(3).
- Budiantara, I.N. *et al.* (2012) 'Modeling the percentage of poor people in Indonesia using spline nonparametric regression approach', *International Journal of Basic and Applied Sciences*, 12(06), pp. 119–124.
- Budiantara, I.N. *et al.* (2015) 'The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties', *Applied Mathematical Sciences*, 9(122), pp. 6083–6094.
- Dani, A.T.R., Adrianingsih, N.Y. and Ainurrochmah, A. (2020) 'Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi (Studi Kasus: Gini Ratio Kabupaten/Kota Di Provinsi Jawa Timur Tahun 2017)', *Jambura Journal of Probability and Statistics*, 1(2), pp. 99–106.
- Dani, A.T.R. and Ni'matuzzahroh, L. (2022) 'Penerapan Keluarga Model Spline Truncated Polinomial pada Regresi Nonparametrik', *Inferensi*, 5(1), p. 37.
- Dani, A.T.R., Ratnasari, V. and Budiantara, I.N. (2021) 'Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression', *IOP Conference Series: Materials Science and Engineering*, 1115(1), p. 012020.
- Eubank, R.L. (1999) 'Nonparametric Regression and Spline Smoothing'. New York: Marcel Dekker.
- Hidayat, R. *et al.* (2021) 'The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model', *Communications in Statistics - Theory and Methods*, 50(17), pp. 3942–3953.
- Igustin, E.D. and Budiantara, I.N. (2021) 'Pemodelan Faktor-Faktor yang Mempengaruhi Total Fertility Rate di Indonesia Menggunakan Regresi Nonparametrik Spline Truncated', *Jurnal Sains dan Seni ITS*, 9(2), pp. 178–185.
- Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012) *Introduction to Linear Regression Analysis*. 5th edn, Wiley. 5th edn. Canada: John Wiley and Sons Inc.
- Nurchayani, H., Budiantara, I.N. and Zain, I. (2021) 'The semiparametric regression curve estimation by using mixed truncated spline and fourier series model', in *AIP Conference Proceedings*. American Institute of Physics Inc.
- Sifriyani *et al.* (2018) 'Development of Nonparametric Geographically Weighted Regression using truncated spline approach', *Songklanakarin Journal of Science and Technology*, 40(4), pp. 909–920.
- Sifriyani, S. *et al.* (2023) 'Spline And Kernel Mixed Estimators In Multivariable Nonparametric Regression For Dengue Hemorrhagic Fever Model', *Commun. Math. Biol. Neurosci.*, 2023, pp. 1–15.

Wahba, G. and Wang, Y. (1990) *Spline Function: Overview*.

Wahba, G. and Wang, Y. (2014) 'Spline Function', *Encyclopedia of Statistical Sciences*, pp. 1–27.