

Comparative Analysis of Parametric and Nonparametric Methods in Modeling Under-Five Malnutrition Prevalence Across Indonesia

Dita Amelia*, Sulyanto, Adelia Putri Andini, Faya Najwatus Silma, Nafla Nara Yonay, Slavina, dan Dinnara Chairana Aisha

Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya, Indonesia

*Corresponding author: dita.amelia@fst.unair.ac.id

Submitted : 10 Juli 2026
Revised : 16 Agustus 2026
Accepted : 30 Agustus 2026

ABSTRACT

Malnutrition prevalence among children under five in Indonesia varied widely across provinces in 2024, ranging from 8.70% to 37.00%. Addressing this issue supports the global Sustainable Development Goals (SDGs) agenda, particularly SDG 2 (Zero Hunger) and SDG 3 (Good Health and Well-being). While most studies rely on multiple linear regression, its strict assumption of linearity is often violated in practice. This study conducts a comparative analysis between multiple linear regression and nonparametric penalized spline regression to model the effects of professional-assisted deliveries, safe sanitation access, complete basic immunization, and households at risk of stunting across 36 Indonesian provinces. Cross-sectional data for 2024 from the Central Bureau of Statistics (BPS) were analyzed, using MSE, R^2 , and GCV for model comparison. The multiple linear regression model yielded an MSE of 18.812 and an R^2 of 53.03%. Conversely, the penalized spline model (second-order polynomial, three knot points, $\lambda = 0.0004$) achieved a substantially lower MSE of 2.653 and a higher R^2 of 92.31%, demonstrating its superior capability in capturing nonlinear patterns. Regarding predictor significance, both models consistently identified safe sanitation, complete basic immunization, and households at risk of stunting as influential factors. However, professional assisted deliveries showed no significant effect in the nonparametric model. These findings confirm that the penalized spline approach provides more accurate estimates and is better suited for modeling provincial level malnutrition determinants.

Keywords: Children Malnutrition, Malnutrition Determinants, Multiple Linear Regression, Penalized Spline, Stunting Risk Family



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Prevalensi gizi buruk balita antarprovinsi di Indonesia tahun 2024 masih menunjukkan kesenjangan yang signifikan, dengan rentang nilai antara 8,7% hingga 37%, di mana provinsi-provinsi di kawasan timur Indonesia secara konsisten mencatatkan angka yang lebih tinggi dibandingkan wilayah barat (Pangaribuan et al., 2026). Kondisi gizi yang buruk pada masa balita berdampak langsung terhadap pertumbuhan fisik, perkembangan kognitif, serta daya tahan tubuh anak dalam jangka panjang (Rani et al., 2025). Penanganan permasalahan ini berkontribusi terhadap agenda global *Sustainable Development Goals* (SDGs), khususnya SDG ke-2 dan SDG ke-3 yang menargetkan penghapusan kelaparan serta penjaminan kesehatan bagi seluruh lapisan masyarakat (Hadjarati et al., 2022). Kesenjangan antarprovinsi tersebut mengindikasikan bahwa determinan gizi buruk balita beroperasi secara berbeda di setiap wilayah, sehingga diperlukan pendekatan analisis yang mampu mengakomodasi heterogenitas pola hubungan antarvariabel secara lebih komprehensif.

Sejumlah penelitian terdahulu telah mengidentifikasi berbagai faktor yang berkorelasi dengan kondisi gizi balita, di antaranya cakupan persalinan oleh tenaga kesehatan yang berperan menjamin keselamatan ibu dan bayi sejak awal kehidupan (Wati et al., 2025), sanitasi lingkungan yang terbukti menekan risiko paparan penyakit infeksi (Budge et al., 2022), serta imunisasi dasar lengkap sebagai bentuk perlindungan terhadap penyakit yang berpotensi menurunkan status gizi anak (Oktavia et al., 2025). Selain itu, proporsi keluarga berisiko *stunting* turut mencerminkan kondisi sosial dan kesehatan yang kurang kondusif bagi tumbuh kembang optimal balita (Maryani et al., 2024). Keterkaitan antara keempat faktor tersebut dengan prevalensi gizi buruk balita telah dikonfirmasi oleh Afdhal dan Arsi (2023), meskipun kajian

tersebut masih mengandalkan regresi linier berganda yang mensyaratkan asumsi linearitas hubungan antarvariabel. Analisis awal data mengindikasikan adanya hubungan non-linear antara keempat variabel prediktor dengan prevalensi balita bergizi buruk di tingkat provinsi, yang mengisyaratkan perlunya pendekatan pemodelan yang lebih fleksibel, sehingga penerapan model parametrik konvensional berpotensi menghasilkan estimasi yang kurang akurat (James et al., 2021).

Berdasarkan kesenjangan tersebut, penelitian ini mengajukan pendekatan regresi nonparametrik *penalized spline* sebagai alternatif pemodelan yang lebih fleksibel dan tidak terikat pada asumsi bentuk fungsi tertentu (Dani et al., 2021). Metode ini telah terbukti menghasilkan estimasi yang lebih baik dibandingkan pendekatan parametrik pada data dengan pola hubungan yang kompleks (Sarmilah & Amalita, 2026). Kinerja model tersebut selanjutnya dibandingkan secara objektif dengan regresi linier berganda berdasarkan kriteria nilai MSE dan koefisien determinasi. Tujuan utama dari penelitian ini adalah menghasilkan model estimasi terbaik dengan tingkat akurasi yang lebih tinggi, sehingga mampu menyediakan basis empiris yang kokoh bagi perumusan strategi intervensi kesehatan yang efektif dan tepat sasaran di tingkat provinsi (Desmita et al., 2025).

II. METODE PENELITIAN

A. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) tahun 2024. Data yang digunakan berupa data *cross section* dengan unit pengamatan seluruh provinsi di Indonesia. Variabel yang dianalisis mencakup prevalensi gizi buruk balita sebagai variabel respon serta beberapa indikator kesehatan balita sebagai variabel prediktor. Pemilihan data tahun 2024 didasarkan pada ketersediaan data yang lengkap dan merupakan data terbaru yang dapat menggambarkan kondisi prevalensi gizi buruk balita di Indonesia secara aktual. Selain itu, BPS dipilih sebagai sumber data karena merupakan lembaga resmi pemerintah yang menyediakan data statistik yang kredibel dan dapat dipertanggungjawabkan.

B. Variabel Penelitian

Penelitian ini menggunakan satu variabel respons dan beberapa variabel prediktor yang diduga memiliki pengaruh terhadap prevalensi gizi buruk balita di Indonesia. Variabel-variabel tersebut disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala	Definisi
X_1	Persalinan oleh Tenaga Kesehatan (%)	Rasio	Persentase persalinan yang dibantu oleh tenaga kesehatan profesional pada setiap provinsi
X_2	Sanitasi Akses Aman (%)	Rasio	Persentase rumah tangga yang memiliki akses sanitasi layak pada setiap provinsi
X_3	Imunisasi Dasar Lengkap (%)	Rasio	Persentase balita yang memperoleh imunisasi dasar lengkap pada setiap provinsi
X_4	Keluarga Berisiko <i>Stunting</i> (%)	Rasio	Persentase keluarga yang berisiko <i>stunting</i> pada setiap provinsi
Y	Prevalensi Gizi Buruk Balita (%)	Rasio	Persentase balita yang mengalami gizi buruk pada setiap provinsi

C. Tahapan Analisis

Prosedur analisis dalam penelitian ini dikembangkan secara sistematis agar selaras dengan rumusan masalah dan target capaian yang ada. Garis besar langkah pelaksanaan penelitian tersebut diuraikan sebagai berikut.

1. Identifikasi variabel dan pengumpulan data dilakukan dengan menetapkan variabel respons prevalensi gizi buruk (Y) serta empat variabel prediktor (X_1, X_2, X_3, X_4) kemudian mengumpulkan data sekunder *cross section* 36 provinsi tahun 2024 dari BPS.
2. Pembersihan data (*data cleaning*) dijalankan melalui proses input dan pemeriksaan kelengkapan data dari seluruh wilayah pengamatan guna memastikan tidak terdapat data yang hilang.
3. Analisis deskriptif dan eksplorasi diterapkan untuk menghitung karakteristik numerik data seperti nilai minimum, maksimum, rata-rata, dan standar deviasi guna mengetahui gambaran umum dari seluruh variabel penelitian.
4. Estimasi parameter regresi linear berganda diperoleh dengan menghitung koefisien regresi menggunakan metode kuadrat terkecil atau *Ordinary Least Square* (OLS). Jika matriks $X^T X$ bersifat non-singular, maka estimator titik untuk parameter β dapat diperoleh secara tunggal melalui operasi invers matriks (Wooldridge, 2020):
$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (1)$$
5. Pengujian signifikansi model parametrik mencakup uji multikolinearitas, uji simultan (uji F), dan uji parsial (uji t) untuk menyusun persamaan akhir berdasarkan variabel prediktor yang terbukti signifikan. Untuk mendeteksi

multikolinearitas, salah satu metode yang digunakan adalah *Variance Inflation Factor* (VIF) yang dirumuskan (Mardiatmoko, 2020):

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2)$$

Selanjutnya, pengujian statistik uji simultan (uji F) diterapkan untuk mengevaluasi pengaruh seluruh variabel prediktor secara bersama-sama terhadap variabel respons:

$$F = \frac{MSR}{MSE} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - k - 1)} \quad (3)$$

Sementara itu, kontribusi signifikan dari masing-masing parameter secara parsial diidentifikasi menggunakan uji parsial atau uji t (James, dkk., 2021):

$$t = \frac{\hat{\beta}_i}{SE_{\hat{\beta}_i}} \quad (4)$$

6. Pemeriksaan asumsi klasik dilakukan melalui rangkaian uji asumsi residual model OLS yang meliputi uji normalitas, homoskedastisitas, dan autokorelasi. Pengujian asumsi residual berdistribusi normal dapat dievaluasi salah satunya melalui statistik uji Shapiro-Wilk (Silalahi dkk., 2024):

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

7. Optimasi komponen *Penalized Spline* diawali dengan menentukan orde polinomial, jumlah dan posisi titik *knot*, serta mencari nilai parameter penghalus (λ) yang optimal. Estimasi vektor koefisien regresi β pada tahap ini diturunkan dari fungsi *Penalized Least Square* atau PLS (Octavia dkk., 2024):

$$\hat{\beta} = (X^T X + \lambda^2 D)^{-1} X^T Y \quad (6)$$

8. Pemilihan model *Penalized Spline* terbaik dilakukan dengan mengestimasi vektor parameter model nonparametrik dan menentukan kombinasi komponen terbaik berdasarkan nilai *Generalized Cross Validation* (GCV) paling minimum. Penentuan parameter penghalus (λ) yang menghasilkan keseimbangan optimal dihitung melalui fungsi GCV (Ega Tanaty, 2023):

$$GCV(\lambda) = \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{f}_\lambda(x_i))^2}{[n^{-1} \text{tr}(I - H(\lambda))]^2} \quad (7)$$

9. Komparasi dan evaluasi kebaikan suai menguji performa model parametrik dan nonparametrik secara objektif dengan membandingkan nilai *Mean Square Error* (MSE) terkecil serta koefisien determinasi (R^2) terbesar. Nilai deviasi prediksi dihitung berdasarkan rata-rata kuadrat selisih data aktual (Caroline Wongkar & Fahmuddin, 2023):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

sedangkan pengukuran proporsi keragaman variabel respons yang mampu dijelaskan oleh model dievaluasi melalui rumus koefisien determinasi (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

10. Interpretasi final dan penarikan kesimpulan menjadi tahap akhir untuk menjelaskan kontribusi spesifik setiap prediktor berdasarkan model terbaik terpilih, sekaligus merumuskan rekomendasi kebijakan praktis serta saran penelitian.

III. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Berdasarkan hasil analisis statistik deskriptif terhadap 36 provinsi di Indonesia tahun 2024, diperoleh ringkasan karakteristik variabel penelitian yang disajikan pada tabel berikut:

Tabel 2. Statistika Deskriptif Variabel Penelitian

Variabel	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Prevalensi Gizi Buruk Balita (%)	36	8,70	37,00	22,27	5,96	35,47
Persalinan Oleh Tenaga Kesehatan (%)	36	78,17	100,00	95,59	5,52	30,48
Sanitasi Akses Aman (%)	36	0,80	20,40	6,40	5,05	25,50
Imunisasi Dasar Lengkap (%)	36	5,40	89,50	48,56	17,75	315,12
Keluarga Berisiko <i>Stunting</i> (%)	36	2,50	18,80	7,50	4,64	21,59

Berdasarkan Tabel.2, prevalensi gizi buruk balita (Y) di Indonesia tahun 2024 memiliki nilai minimum sebesar 8,70% dan maksimum sebesar 37,00%, dengan rata-rata sebesar 22,27% dan standar deviasi sebesar 5,96. Rentang nilai yang cukup lebar menunjukkan adanya variasi prevalensi gizi buruk balita antarprovinsi. Variabel persentase persalinan oleh tenaga kesehatan (X_1) memiliki rata-rata tertinggi, yaitu 95,59%, dengan standar deviasi 5,52 yang menunjukkan bahwa cakupan layanan persalinan tenaga kesehatan relatif merata di sebagian besar provinsi.

Sementara itu, persentase sanitasi akses aman (X_2), imunisasi dasar lengkap (X_3), dan keluarga berisiko *stunting* (X_4) masing-masing memiliki rata-rata sebesar 6,40%, 48,56%, dan 7,51%. Variabel imunisasi dasar lengkap (X_3) memiliki standar deviasi terbesar, yaitu 17,75, yang mengindikasikan adanya perbedaan cakupan imunisasi yang cukup tinggi antarprovinsi. Secara umum, hasil statistika deskriptif menunjukkan bahwa indikator kesehatan balita di Indonesia masih bervariasi antarwilayah, sehingga diperlukan analisis lebih lanjut untuk mengidentifikasi faktor-faktor yang memengaruhi prevalensi gizi buruk balita.

B. Pemodelan Regresi Linier Berganda

Estimasi parameter untuk model regresi linier berganda biasanya dilakukan dengan metode *Ordinary Least Square* (OLS). Sebelum menginterpretasi model, dilakukan uji multikolinearitas dengan menggunakan nilai *Variance Inflation Factor* (VIF) untuk memastikan tidak terdapat hubungan linear kuat antarvariabel prediktor. Nilai VIF dari masing-masing variabel ditunjukkan pada Tabel 3.

Tabel 3. Nilai VIF Variabel

Variabel	Coef	VIF
(Constant)	55,378	—
Persalinan oleh Tenaga Kesehatan (X_1)	−0,307	1,26
Sanitasi Akses Aman (X_2)	−0,382	1,43
Imunisasi Dasar Lengkap (X_3)	−0,102	1,32
Keluarga Berisiko <i>Stunting</i> (X_4)	0,476	1,07

Berdasarkan Tabel 3, seluruh nilai VIF < 10 sehingga tidak terdapat multikolinearitas antarvariabel prediktor. Selain itu, diperoleh pula nilai koefisien yang membentuk persamaan model regresi linier berganda sebagai berikut:

$$\hat{Y} = 55,378 - 0,307X_1 - 0,382X_2 - 0,102X_3 + 0,476X_4 \quad (10)$$

Persamaan (10) menunjukkan bahwa persalinan oleh tenaga kesehatan (X_1), sanitasi akses aman (X_2), dan imunisasi dasar lengkap (X_3) memiliki pengaruh negatif terhadap prevalensi gizi buruk balita, sedangkan keluarga berisiko *stunting* (X_4) memiliki pengaruh positif. Dengan asumsi variabel lain konstan, setiap peningkatan satu satuan pada X_1 , X_2 , dan X_3 akan menurunkan prevalensi gizi buruk balita masing-masing sebesar 0,307; 0,382; dan 0,102 satuan. Sebaliknya, setiap peningkatan satu satuan pada X_4 akan meningkatkan prevalensi gizi buruk balita sebesar 0,476 satuan. Nilai konstanta sebesar 55,378 menunjukkan perkiraan prevalensi gizi buruk balita ketika seluruh variabel prediktor bernilai nol.

Selanjutnya, akan dilakukan uji asumsi klasik yang bertujuan untuk memastikan bahwa model regresi linier berganda yang diperoleh memenuhi asumsi-asumsi yang diperlukan sehingga hasil estimasi dan pengujian statistik yang dihasilkan dapat dipercaya. Uji asumsi yang dilakukan meliputi uji normalitas, uji heteroskedastisitas, serta uji autokorelasi. Hasil pengujian asumsi klasik disajikan pada Tabel 4.

Tabel 4. Hasil Uji Asumsi Klasik Model

Uji Asumsi Klasik	Hasil
Uji Normalitas (<i>Anderson-Darling</i>)	AD = 0,390; <i>p-value</i> = 0,365
	X_1 <i>p-value</i> = 0,837
	X_2 <i>p-value</i> = 0,600
Uji Heteroskedastisitas (<i>Glejser</i>)	X_3 <i>p-value</i> = 0,854
	X_4 <i>p-value</i> = 0,075
Uji Autokorelasi (<i>Durbin-Watson</i>)	DW = 1,836

Berdasarkan Tabel 4, model regresi linier berganda telah memenuhi asumsi klasik. Hal ini ditunjukkan oleh residual yang berdistribusi normal ($p\text{-value} = 0,365 > 0,05$), tidak terdapat gejala heteroskedastisitas (seluruh $p\text{-value} > 0,05$), serta tidak ada autokorelasi pada residual ($DW = 1,836$). Dengan demikian, model regresi yang diperoleh layak untuk dianalisis lebih lanjut, yaitu dengan pengujian signifikansi model menggunakan uji simultan (uji F) dan uji parsial (uji t), serta evaluasi *goodness of fit*. Uji F digunakan untuk mengetahui apakah seluruh variabel prediktor secara bersama-sama memengaruhi variabel respons, sedangkan uji t digunakan untuk mengidentifikasi pengaruh masing-masing variabel prediktor secara individual. Kemudian evaluasi *goodness of fit* menggunakan *Mean Squared Error* (MSE) dan nilai koefisiensi determinasi (R^2) dilakukan untuk menilai kemampuan model dalam menjelaskan variasi data yang diamati. Hasil pengujian tersebut disajikan pada Tabel 5.

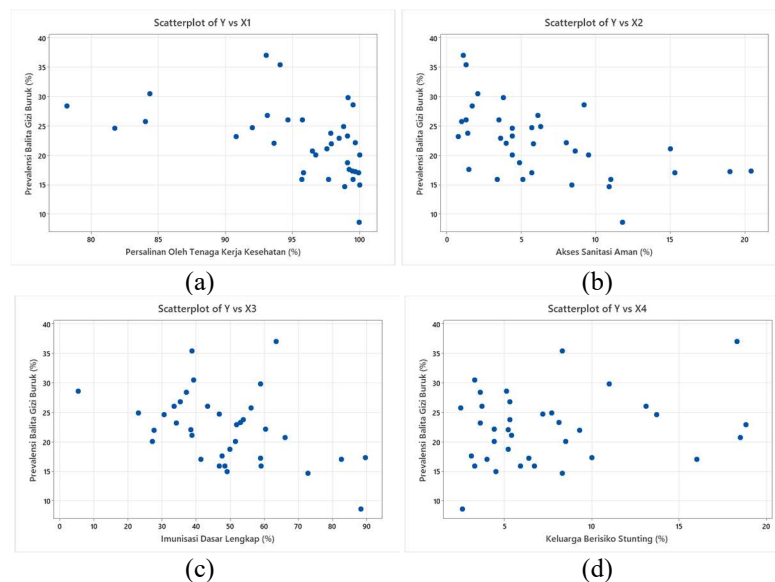
Tabel 5. Hasil Uji Signifikansi dan *Goodness of Fit* Model

Uji	Hasil
Uji Signifikansi Simultan (Uji F)	$p\text{-value} = < 0,001$
	X_1 $p\text{-value} = 0,048$
	X_2 $p\text{-value} = 0,035$
Uji Signifikansi Parsial (Uji t)	X_3 $p\text{-value} = 0,040$
	X_4 $p\text{-value} = 0,006$
<i>Goodness of Fit</i>	MSE = 18,812; $R^2 = 53,03\%$

Berdasarkan Tabel 5, hasil dari uji simultan (uji F) menghasilkan nilai $p\text{-value} < 0,001$ sehingga dapat disimpulkan bahwa variabel persalinan oleh tenaga kesehatan, sanitasi akses aman, imunisasi dasar lengkap, dan keluarga berisiko *stunting* secara bersama-sama berpengaruh signifikan terhadap prevalensi gizi buruk balita. Hasil uji parsial (uji t) menunjukkan bahwa seluruh variabel prediktor memiliki nilai $p\text{-value} < 0,05$, sehingga dapat disimpulkan bahwa masing-masing variabel prediktor berpengaruh secara signifikan terhadap prevalensi gizi buruk balita. Selain itu, nilai MSE 18,812 menunjukkan bahwa variansi simpangan antara nilai prevalensi gizi buruk aktual di setiap provinsi dengan nilai yang diprediksi oleh model terhitung cukup tinggi, serta nilai koefisiensi determinasi 53,03% menunjukkan bahwa seluruh variabel prediktor secara bersama-sama mampu menjelaskan variabilitas dari prevalensi gizi balita sebesar 52,03%, sedangkan 46,97% lainnya dijelaskan oleh faktor atau variabel lain di luar model.

C. Pemodelan Regresi Nonparametrik *Penalized Spline*

Sebelum membangun model regresi nonparametrik *Penalized Spline*, dilakukan eksplorasi awal untuk mengamati pola hubungan antara prevalensi gizi buruk balita dengan setiap variabel prediktor. Visualisasi hubungan masing-masing variabel disajikan pada Gambar 1.



Gambar 1. *Scatter plot* hubungan prevalensi gizi buruk balita dengan (a) persalinan oleh tenaga kesehatan (X_1), (b) sanitasi akses aman (X_2), (c) imunisasi dasar lengkap (X_3), dan (d) keluarga berisiko *stunting* (X_4).

Sebaran data pada Gambar 1 menunjukkan bahwa hubungan antara prevalensi gizi buruk balita dan masing-masing variabel prediktor belum membentuk pola linier yang konsisten. Titik-titik observasi tersebar cukup luas tanpa mengikuti kecenderungan garis lurus yang jelas. Pada variabel persalinan oleh tenaga kesehatan (X_1), sanitasi akses aman (X_2), dan imunisasi dasar lengkap (X_3) terlihat kecenderungan hubungan negatif, sedangkan keluarga berisiko *stunting* (X_4) menunjukkan pola yang lebih lemah. Kondisi tersebut mengindikasikan bahwa hubungan antarvariabel bersifat lebih kompleks sehingga diperlukan pendekatan yang fleksibel, yaitu regresi nonparametrik *Penalized Spline*.

Untuk mengeksplorasi kemungkinan hubungan yang tidak sepenuhnya linier antara prevalensi gizi buruk balita dan faktor-faktor yang memengaruhinya, digunakan pendekatan regresi nonparametrik *Penalized Spline* yang memberikan fleksibilitas lebih tinggi dibandingkan regresi linier berganda karena bentuk hubungan mengikuti karakteristik data yang diamati. Di sisi lain, keberadaan parameter penghalus (λ) berperan dalam mengendalikan kompleksitas model sehingga keseimbangan antara kecocokan model dan kehalusan kurva tetap terjaga.

Penentuan model dilakukan melalui evaluasi beberapa kombinasi orde polinomial, jumlah titik knot, dan parameter penghalus (*smoothing parameter*). Evaluasi dilakukan pada model spline orde 1 dan orde 2 dengan variasi jumlah titik knot sebanyak satu, dua, dan tiga pada setiap variabel prediktor. Untuk setiap kombinasi tersebut ditentukan nilai parameter penghalus yang optimal, kemudian kinerja model dibandingkan menggunakan nilai *Mean Squared Error* (MSE), *Generalized Cross Validation* (GCV), dan koefisien determinasi (R^2). Dari seluruh kombinasi yang dievaluasi, model terbaik diperoleh pada spline orde 2 dengan tiga titik knot, sehingga spesifikasi model tersebut disajikan pada Tabel 6.

Tabel 6. Ringkasan Model *Penalized Spline* Terbaik

Orde Polinomial	Jumlah Knot	Titik Knot X1	Titik Knot X2	Titik Knot X3	Titik Knot X4	Lambda (λ)	MSE	GCV	R^2
2	3	92,257	1,425	51,5	5,3	0,0004	2,652538	14,9389	92,31%
		95,17	8,75	58,825	7,2				
		97,353	13,4	69,4	8,7				

Hasil pada Tabel 6 menunjukkan bahwa spesifikasi terbaik diperoleh pada spline orde 2 dengan tiga titik knot dan parameter penghalus sebesar 0,0004. Kombinasi tersebut menghasilkan MSE sebesar 2,652538 dan GCV sebesar 14,9389, disertai koefisien determinasi mencapai 92,31%. Temuan ini mengindikasikan bahwa sebagian besar keragaman prevalensi gizi buruk balita dapat dijelaskan oleh model dengan tingkat kesalahan prediksi yang relatif rendah. Berdasarkan spesifikasi terpilih, selanjutnya diperoleh estimasi parameter yang membentuk fungsi regresi *Penalized Spline* berikut.

$$\begin{aligned} \hat{Y} = & -274,0528 + 5,3993X_1 - 0,0305X_1^2 - 0,5458(X_1 - 92,257)_+^2 + 1,5374(X_1 - 95,170)_+^2 \\ & - 1,4063(X_1 - 97,353)_+^2 + 54,9969X_2 - 19,7858X_2^2 + 19,8058(X_2 - 1,425)_+^2 \\ & + 0,4881(X_2 - 8,750)_+^2 - 1,2461(X_2 - 13,400)_+^2 - 1,1975X_3 + 0,0152X_3^2 + 0,0211(X_3 - 51,500)_+^2 \\ & - 0,1676(X_3 - 58,825)_+^2 - 1,2461(X_2 - 13,400)_+^2 + 24,6571X_4 - 2,8843X_4^2 \\ & + 6,5424(X_4 - 5,300)_+^2 - 6,6751(X_4 - 7,200)_+^2 + 3,0793(X_4 - 8,700)_+^2 \end{aligned} \quad (11)$$

Persamaan (11) memperlihatkan bahwa pola hubungan antara prevalensi gizi buruk balita dan variabel prediktor tidak mengikuti satu bentuk fungsi yang seragam pada seluruh rentang pengamatan. Kehadiran komponen *truncated spline* mengisyaratkan adanya perubahan karakter hubungan pada titik-titik knot tertentu, sehingga respons model terhadap perubahan variabel prediktor dapat berbeda antarinterval data. Karakteristik tersebut memungkinkan model menangkap variasi lokal yang sulit diakomodasi oleh pendekatan linier, sekaligus mempertahankan kestabilan estimasi melalui mekanisme *penalization*.

Untuk mengevaluasi kelayakan model yang diperoleh, dilakukan pengujian signifikansi baik secara simultan maupun parsial. Pengujian ini bertujuan untuk mengidentifikasi kontribusi variabel prediktor dalam menjelaskan variasi prevalensi gizi buruk balita serta memastikan bahwa model yang terbentuk memiliki makna statistik. Hasil uji signifikansi simultan disajikan pada Tabel 7.

Tabel 7. Hasil Uji Signifikansi Simultan Model *Penalized Spline*

F	p-value	Keputusan	Kesimpulan
9,180	0,000032	Tolak H_0	Signifikan

Dilihat dari Tabel 7, didapatkan nilai F-hitung sebesar 9,180 dan p-value sebesar 0,000032 yang lebih kecil dari taraf signifikansi 5%. Hasil tersebut menunjukkan bahwa model regresi *Penalized Spline* yang terbentuk signifikan secara statistik. Dengan demikian, variabel persalinan oleh tenaga kesehatan, sanitasi akses aman, imunisasi dasar lengkap, dan keluarga berisiko *stunting* secara simultan berpengaruh terhadap prevalensi gizi buruk balita di Indonesia

tahun 2024. Untuk mengidentifikasi kontribusi masing-masing variabel dalam model, pengujian dilanjutkan secara parsial terhadap parameter yang diestimasi. Ringkasan hasil uji signifikansi parsial disajikan pada Tabel 8.

Tabel 8. Hasil Uji Signifikansi Parsial Model *Penalized Spline*

Variabel	Parameter Signifikan ($p\text{-value} < 0,05$)	Keputusan	Kesimpulan
Persalinan oleh tenaga kesehatan (X_1)	Tidak ada	Terima H_0	Tidak Signifikan
Sanitasi akses aman (X_2)	$\beta_{X_2}, \beta_{X_2^2}, \beta_{K_{21}}, \beta_{K_{23}}$	Tolak H_0	Signifikan
Imunisasi dasar lengkap (X_3)	$\beta_{X_3}, \beta_{X_3^2}, \beta_{K_{32}}, \beta_{K_{33}}$	Tolak H_0	Signifikan
Keluarga beresiko <i>stunting</i> (X_4)	$\beta_{X_4}, \beta_{X_4^2}, \beta_{K_{41}}, \beta_{K_{42}}, \beta_{K_{43}}$	Tolak H_0	Signifikan

Berdasarkan Tabel 8, seluruh parameter yang berkaitan dengan variabel persalinan oleh tenaga kesehatan (X_1) tidak menunjukkan signifikansi statistik pada taraf 5%. Sebaliknya, variabel sanitasi akses aman (X_2), imunisasi dasar lengkap (X_3), dan keluarga berisiko *stunting* (X_4) memiliki beberapa parameter signifikan, baik pada komponen polinomial maupun *truncated spline*. Temuan ini mengindikasikan bahwa variasi prevalensi gizi buruk balita lebih banyak dipengaruhi oleh kondisi sanitasi, cakupan imunisasi dasar lengkap, dan keberadaan keluarga berisiko *stunting* dibandingkan faktor persalinan oleh tenaga kesehatan.

Selain pengujian signifikansi model, dilakukan evaluasi terhadap asumsi residual yang meliputi normalitas, rata-rata residual, homogenitas varians, dan independensi residual. Hasil pengujian menunjukkan bahwa residual memenuhi asumsi normalitas ($p\text{-value} = 0,926$), memiliki rata-rata yang tidak berbeda dari nol ($p\text{-value} = 1,000$), serta tidak mengalami autokorelasi berdasarkan statistik Durbin–Watson sebesar 2,628. Sementara itu, uji Glejser ($p\text{-value} = 0,021$) mengindikasikan adanya heteroskedastisitas, yang berarti variasi dari residual model tidak konstan atau tidak sama di sepanjang rentang nilai variabel prediktor. Meskipun demikian, kondisi tersebut tidak menjadi pertimbangan utama dalam regresi nonparametrik *Penalized Spline* selama model mampu merepresentasikan pola data dengan baik. Hal ini didukung oleh nilai R^2 sebesar 92,31% serta nilai MSE dan GCV yang relatif kecil, sehingga model tetap layak digunakan dalam analisis. Meskipun nilai R^2 yang diperoleh tinggi, tetapi tidak mengindikasikan adanya *overfitting* karena orde, jumlah *knot*, serta parameter penghalus (λ) dipilih melalui metode *Generalized Cross Validation* (GCV), yang langsung mengendalikan kompleksitas model. Selain itu, nilai GCV yang rendah (14,9389) menunjukkan bahwa model sudah optimal, bukan hasil *overfitting*.

D. Perbandingan Model Regresi Linear Berganda dengan *Penalized Spline*

Tahap komparasi model dilakukan untuk mengevaluasi kinerja serta memilih model terbaik antara pendekatan parametrik (Regresi Linear Berganda) dan pendekatan nonparametrik (*Penalized Spline*) dalam menjelaskan faktor-faktor yang memengaruhi prevalensi gizi buruk balita di Indonesia tahun 2024. Kriteria pemilihan model terbaik didasarkan pada nilai *Mean Squared Error* (MSE) yang paling minimum serta koefisien determinasi (R^2) yang paling tinggi. Ringkasan perbandingan kebaikan suai kedua model disajikan pada Tabel 9.

Tabel 9. Perbandingan Kinerja Model Regresi Linear Berganda dan *Penalized Spline*

Pendekatan Model	MSE	R^2
Regresi Linear Berganda (Parametrik)	18,8118	53,03%
<i>Penalized Spline</i> Orde 2, 3 <i>knot</i> (Nonparametrik)	2,652538	92,31%

Berdasarkan hasil evaluasi, model nonparametrik *Penalized Spline* orde 2 dengan 3 *knot* terpilih sebagai model terbaik karena performanya jauh lebih unggul dibandingkan Regresi Linear Berganda. Model ini menghasilkan nilai MSE yang jauh lebih minimum sebesar 2,652538 dan nilai koefisien determinasi (R^2) mencapai 92,31%. Sementara model parametrik hanya menghasilkan MSE sebesar 18,8118 dengan R^2 sebesar 53,03%. Keunggulan estimasi ini dicapai melalui kendali parameter penghalus (λ) optimal sebesar 0,0004, yang menghasilkan kriteria *Generalized Cross Validation* (GCV) minimum senilai 14,9389. Selain itu, signifikansi parameter *spline* pada X_2 , X_3 , dan X_4 membuktikan adanya fluktuasi tajam angka gizi buruk pada ambang batas interval wilayah tertentu. Hal ini menegaskan bahwa pendekatan *Penalized Spline* sanggup menangkap aspek heterogenitas data riil di lapangan secara akurat tanpa gejala *overfitting*.

IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa meskipun regresi linier berganda menyatakan seluruh variabel prediktor berpengaruh signifikan terhadap prevalensi gizi buruk balita di Indonesia tahun 2024, pola hubungan data cenderung tidak linier sehingga kurang optimal direpresentasikan oleh model linier. Pendekatan regresi nonparametrik *Penalized*

Spline dengan orde polinomial 2, tiga titik knot, dan parameter penghalus $\lambda = 0,0004$ memberikan kinerja yang lebih baik, ditunjukkan oleh penurunan nilai MSE dari 18,8118 menjadi 2,6525 serta peningkatan R^2 dari 53,03% menjadi 92,31%. Pengujian model menunjukkan bahwa secara simultan model signifikan, sedangkan secara parsial hanya variabel sanitasi akses aman, imunisasi dasar lengkap, dan keluarga berisiko *stunting* yang berpengaruh signifikan. Dengan demikian, regresi nonparametrik *Penalized Spline* lebih sesuai digunakan untuk memodelkan prevalensi gizi buruk balita karena mampu menangkap hubungan yang lebih kompleks. Penelitian selanjutnya disarankan menambahkan variabel lain dan mempertimbangkan metode pemodelan yang lebih beragam agar hasil yang diperoleh semakin komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas bimbingan, arahan, dan masukan yang konstruktif selama proses penelitian. Ucapan terima kasih juga disampaikan kepada Badan Pusat Statistik serta instansi terkait yang telah menyediakan data publik sehingga penelitian ini dapat dilaksanakan dan diselesaikan dengan baik.

DAFTAR PUSTAKA

- Afdhal, F. dan Arsi, R. (2023). "Faktor-Faktor yang Mempengaruhi Status Gizi Balita di Wilayah Kerja Puskesmas Prabumulih Timur Tahun 2022". *JURNAL KESEHATAN TERAPAN*, Vol. 10, No. 1, hal. 82–95.
- Budge, S., Ambelu, A., Bartram, J., Brown, J., dan Hutchings, P. (2022). "Environmental Sanitation and the Evolution of Water, Sanitation and Hygiene". *Bulletin of the World Health Organization*, Vol. 100, No. 4, hal. 286–288.
- Carolyn Wongkar, D., dan Fahmuddin, M.S. (2023), "Analisis Regresi Nonparametrik Spline Truncated untuk Menganalisis Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Sulawesi Selatan", *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, Vol. 5, No. 2, hal. 55–63.
- Dani, A.T.R., Ni'matuzzahroh, L., Ratnasari, V., dan Budiantara, I.N. (2021). "Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal". *Inferensi*, Vol. 4, No. 1, hal. 47.
- Desmita, R., Siregar, R.S., Ivanda, V., Hanoselina, Y., dan Helmi, R.F. (2025). "Peran Tenaga Kesehatan dalam Pemberdayaan Keluarga untuk Pencegahan Stunting". *Jurnal Kesehatan*.
- Ega Tanaty, R. (2023), "Regresi Nonparametrik Spline Truncated untuk Memodelkan Tingkat Pengangguran Terbuka di Pulau Kalimantan", *Jurnal Statistika dan Aplikasinya*, Vol. 7, No. 2.
- Hadjarati, H., Kadir, S., dan Bait, Y. (2022). "Penyuluhan Pencegahan *Stunting* pada Anak dalam Mencapai Tujuan *Sustainable Development Goals* (SDGs)". *Jurnal Pengabdian Kesehatan Masyarakat*, Vol. 3, No. 1, hal. 1–14.
- James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R*. Edisi ke-2. Springer, New York.
- Maryani, S., Kadar, S., Alif, R., Rosyidah, N., Trisilawati, R., Zulaika, C., Apriyani, S., Hastuti, P., Kistimbar, S., Hendra, A., dan Rahmad, A. (2024). *Penanganan Stunting untuk Desa*. Yayasan Kita Menulis, Medan.
- Mardiatmoko, G. (2020), "The Importance of the Classical Assumption Test in Multiple Linear Regression Analysis (A Case Study of the Preparation of the Allometric Equation of Young Walnuts)", *Journal Barekeng*, Vol. 14, No. 3, hal. 333–342.
- Octavia, D.H., Auliarani, A., Siswanto, S., dan Kalondeng, A. (2024), "Modeling Determinants of Composite Stock Price Index Based on Multivariable Nonparametric Penalized Spline Regression Model", *Jurnal Matematika, Statistika dan Komputasi*, Vol. 20, No. 3, hal. 497–512.
- Oktavia, A., Punuh, M.I., dan Sanggelorang, Y. (2025). "Hubungan Riwayat Imunisasi Dasar dengan Status Gizi pada Balita di Wilayah Kerja Puskesmas Remboken". *Jurnal Kesehatan Masyarakat*.
- Pangaribuan, A.C., Dwi, M., dan Mustika, S. (2026). "Determinan Prevalensi *Stunting* Menurut Provinsi di Indonesia". *JEBI: Jurnal Ekonomi dan Bisnis*, Vol. 3, No. 9, hal. 211–224.
- Rani, A., Idrus, K., Asriana, R., dan Hasmah Sage, N. (2025). "Dampak Kesehatan dan Gizi yang Baik pada Anak Usia Dini". *Indonesian Journal of Health and Psychology*, Vol. 1, No. 1.

- Sarmilah dan Amalita, N. (2026). "Perbandingan Regresi Linear dan *Penalized Spline* dalam Analisis Indikator Pendidikan di Indonesia Tahun 2024". *Jurnal Statistika*, Vol. 8, No. 2, hal. 511–517.
- Silalahi, R.A., Hafsari, A.A., Situmorang, D., Emaninta, N., Ginting, B., Girsang, A.B., Martin, M., Febriyansi, E., dan Ompusunggu, P. (2024), "Hasil Perhitungan Asumsi Klasik: Tentang Uji Autokorelasi, Normalitas, Dan Heterokedatisitas", *Jurnal Ilmiah Multidisipliner (JIM)*, Vol. 8, No. 12.
- Wati, I., Suhartati, S., Devin, D., dan Arbiyah, A. (2025). "Peran Bidan dalam Mendukung Persalinan Aman dan Nyaman". *RIGGS: Journal of Artificial Intelligence and Digital Business*, Vol. 4, No. 3, hal. 8538–8544
- Wooldridge, J.M. (2020). *Introductory Econometrics: A Modern Approach*, Ed. 7, Cengage Learning India.