

Comparison of Least Square Spline and Penalized Spline for Modeling Human Development Index Determinants

Tyo Anugrah Putra, Slavina, Thareq Alexander Manggala Napitupulu, Layyin Gisvira, Nila Khoirun Naili Salam, dan Dita Amelia*

Departemen Matematika, Universitas Airlangga, Surabaya, Indonesia

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

Submitted : 10 Juli 2026
Revised : 27 Agustus 2026
Accepted : 30 Agustus 2026

ABSTRACT

The Human Development Index (HDI) is a key indicator of national and regional development and supports the achievement of the Sustainable Development Goals (SDGs), particularly Goal 1 (No Poverty) and Goal 4 (Quality Education). However, substantial disparities in HDI remain across Indonesia's regencies and cities due to complex and nonlinear socioeconomic relationships that cannot always be captured by conventional parametric methods. This study analyzes the determinants of HDI in 514 regencies and cities in Indonesia in 2024 using nonparametric spline regression by comparing Penalized Spline and Least Square Spline estimators. The explanatory variables include mean years of schooling, labor force participation rate, percentage of senior-high-school graduates, and poverty rate. Secondary data from BPS were analyzed through spline basis construction, smoothing parameter selection using Generalized Cross Validation (GCV), parameter estimation, significance testing, and residual diagnostics. The results show that all predictors have nonlinear relationships with HDI. Mean years of schooling and the percentage of senior-high-school graduates positively affect HDI, whereas labor force participation and poverty rate have negative effects. The second-order Penalized Spline model with four knot points achieved the best performance, yielding the lowest GCV (5.388838), the lowest MSE (4.948574), and the highest adjusted R^2 (86.94%). Residual diagnostics confirmed normality and zero mean. Overall, Penalized Spline regression provides a flexible and accurate approach for modeling HDI determinants and informing evidence-based regional development policy.

Keywords: Nonparametric Regression, Penalized Spline, Least Square Spline, Human Development Index, Generalized Cross Validation.



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

Indeks Pembangunan Manusia (IPM) merupakan indikator strategis untuk mengevaluasi keberhasilan pembangunan, baik di tingkat nasional maupun daerah, melalui dimensi kesehatan, pendidikan, dan standar hidup layak. Oleh karena itu, analisis terhadap faktor-faktor yang memengaruhi IPM menjadi penting sebagai dasar dalam perumusan kebijakan pembangunan yang efektif dan berkelanjutan (Rumanama *et al.*, 2022). Selain itu, peningkatan IPM juga mendukung pencapaian *Sustainable Development Goals* (SDGs), khususnya Tujuan 1 (*No Poverty*) dan Tujuan 4 (*Quality Education*). Meskipun IPM Indonesia menunjukkan tren yang terus meningkat setiap tahun, masih terdapat kesenjangan pembangunan manusia yang cukup besar antar kabupaten/kota. Kondisi tersebut mengindikasikan bahwa pembangunan belum berlangsung secara merata dan masih dipengaruhi oleh berbagai faktor sosial, ekonomi, dan pendidikan yang berbeda pada setiap wilayah (Rumanama *et al.*, 2022).

Berbagai penelitian menunjukkan bahwa rata-rata lama sekolah, Tingkat Partisipasi Angkatan Kerja (TPAK), persentase penduduk tamat SMA, dan persentase penduduk miskin merupakan faktor-faktor yang berpengaruh signifikan terhadap IPM (Fitriyah *et al.*, 2021). Selain itu, penelitian Rumanama *et al.* (2022) juga menunjukkan karakteristik demografi dan sosial ekonomi turut menjelaskan variasi IPM antarwilayah. Hasil-hasil penelitian tersebut mengindikasikan bahwa hubungan antara IPM dengan faktor penentunya bersifat kompleks dan tidak selalu linier.

Sebagian besar penelitian terdahulu masih menggunakan pendekatan regresi parametrik, seperti regresi linier berganda, yang mengasumsikan hubungan linier antarvariabel (Fitriyah *et al.*, 2021; Rumanama *et al.*, 2022). Padahal, hubungan antara faktor-faktor sosial ekonomi dengan IPM sering kali bersifat nonlinier, sehingga pendekatan parametrik berpotensi menghasilkan model yang kurang mampu menggambarkan pola hubungan yang sebenarnya.

Regresi nonparametrik *spline* merupakan salah satu pendekatan yang lebih fleksibel karena mampu mengakomodasi perubahan pola hubungan pada titik-titik knot tanpa mengharuskan bentuk fungsi tertentu. Di antara berbagai metode *spline*, *penalized spline* memiliki keunggulan dalam mengendalikan tingkat kemulusan kurva melalui parameter penalti (λ) yang ditentukan menggunakan metode GCV, sehingga mampu menghasilkan keseimbangan antara tingkat kemulusan model dan goodness of fit (Arman, 2025).

Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis pengaruh rata-rata lama sekolah, Tingkat Partisipasi Angkatan Kerja (TPAK), persentase penduduk tamat SMA, dan persentase penduduk miskin terhadap Indeks Pembangunan Manusia pada 514 kabupaten/kota di Indonesia tahun 2024 menggunakan regresi nonparametrik *penalized spline* dan *least square spline*. Selanjutnya, kedua metode dibandingkan berdasarkan nilai *Generalized Cross Validation* (GCV), *Mean Squared Error* (MSE), dan koefisien determinasi (R^2) untuk memperoleh model terbaik serta menginterpretasikan hubungan antara faktor-faktor penentu dengan IPM.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif yang berlandaskan pada filsafat positivisme untuk menguji hipotesis melalui analisis data kuantitatif atau statistik (Risnita, 2024). Metode ini digunakan untuk menganalisis hubungan antarvariabel melalui identifikasi variabel, pengukuran data, serta analisis statistik guna menjelaskan fenomena yang diteliti (Waruwu *et al.*, 2025). Data yang digunakan merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) mengenai Indeks Pembangunan Manusia (IPM) tahun 2024 dengan unit analisis sebanyak 514 kabupaten/kota di Indonesia. Variabel yang diteliti disajikan pada Tabel 1.

Tabel 1. Variabel penelitian

Variabel	Keterangan	Skala Pengukuran
Y	Indeks Pembangunan Manusia (IPM)	Rasio
X_1	Rata-rata Lama Sekolah (RLS)	Rasio
X_2	Tingkat Partisipasi Angkatan Kerja (TPAK)	Rasio
X_3	Persentase Penduduk Tamat SMA	Rasio
X_4	Persentase Penduduk Miskin	Rasio

Analisis data dalam penelitian ini dilakukan secara sistematis untuk memodelkan hubungan antara faktor-faktor sosial-ekonomi terhadap IPM di Indonesia. Adapun langkah-langkah analisis yang dilakukan sebagai berikut.

1. Pengumpulan dan Persiapan Data

Melakukan tabulasi dan pembersihan (*data cleaning*) terhadap dataset IPM dan keempat variabel prediktor X_1, X_2, X_3 , dan X_4 dari 514 kabupaten/kota di Indonesia tahun 2024. Data disusun dalam bentuk matriks agar siap dianalisis menggunakan perangkat lunak komputasi statistik.

2. Eksplorasi Data dengan Analisis Statistika Deskriptif

Menghitung ukuran pemusatan dan penyebaran data, seperti rata-rata, nilai minimum, maksimum, dan simpangan baku untuk mengidentifikasi karakteristik, distribusi data, serta keberadaan *outlier* pada tiap variabel (Silvia, 2020).

3. Identifikasi Pola Hubungan (Pemeriksaan Asumsi Linieritas)

Membuat visualisasi *scatter plot* antara variabel respon dengan masing-masing variabel prediktor. Eksplorasi grafis ini dilakukan untuk melihat indikasi bentuk pola data sekaligus memberikan justifikasi visual bahwa data tidak mengikuti pola linier yang kaku, sehingga tepat jika dianalisis menggunakan regresi nonparametrik.

4. Deteksi Multikolinieritas

Melakukan pengujian multikolinieritas antarvariabel prediktor menggunakan nilai *Variance Inflation Factor* (VIF) untuk memastikan tidak ada hubungan linier yang kuat (Salma, 2020). Nilai VIF dihitung dengan rumus berikut.

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

5. Analisis Regresi Nonparametrik Model *Least Square Spline*

Pembentukan model regresi nonparametrik *least square spline* dilakukan dengan membangun model menggunakan berbagai kombinasi titik knot. Estimasi parameter model ini diperoleh melalui pendekatan (*Ordinary Least Squares*/OLS) tanpa melibatkan parameter penalti ($\lambda = 0$). Penulisan model regresi nonparametrik secara umum mengacu pada Chamidah & Lestari (2022) sebagai berikut.

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

Menurut Eubank (1999), rumus *estimator* koefisien regresi ($\hat{\beta}_{LS}$) adalah sebagai berikut.

$$\hat{\beta}_{LS} = (X_{\lambda}^T X_{\lambda})^{-1} X_{\lambda}^T y \quad (3)$$

6. Analisis Regresi Nonparametrik Model *Penalized Spline*

Pembentukan model regresi nonparametrik *penalized spline* dengan membangun matriks basis *spline* pada setiap variabel prediktor. Proses estimasi ini menggunakan nilai λ optimal serta mengakomodasi matriks penalti agar tidak terjadi *overfitting* sehingga diperoleh rumus estimator parameter model menurut Wahba (1990) sebagai berikut.

$$\hat{\beta}_{PS} = (X^T X + \lambda D)^{-1} X^T y \quad (4)$$

7. Pengujian Signifikansi dan Asumsi *Residual Model Least Square Spline dan Penalized Spline*

Pengujian dilakukan dengan melakukan uji signifikansi model dan parameter pada model *least square spline* dan *penalized spline* untuk mengetahui pengaruh variabel prediktor terhadap variabel respons. Selanjutnya, dilakukan pengujian asumsi *residual* guna memastikan model yang diperoleh telah memenuhi asumsi statistik yang diperlukan.

a. Uji Serentak/Uji F

Uji F dilakukan menggunakan pendekatan ANOVA untuk mengetahui signifikansi variabel prediktor secara simultan terhadap variabel respon. Statistik uji yang digunakan menurut Fadlirhohim *et al.* (2024) sebagai berikut.

$$F = \frac{MSR}{MSE} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - k - 1)} \quad (5)$$

dengan k menyatakan banyaknya parameter regresi (tidak termasuk intersep) yang digunakan dalam model.

b. Uji Parsial/Uji t

Uji t dilakukan untuk mengetahui signifikansi secara pasial setiap parameter dalam model regresi nonparametrik *spline* terhadap variabel respons. Menurut James *et al.* (2021) statistik uji yang digunakan sebagai berikut.

$$t = \frac{\hat{\beta}_i}{S E_{\hat{\beta}_i}} \quad (6)$$

c. Uji Asumsi Normalitas *Residual*

Uji normalitas *residual* dilakukan untuk mengetahui apakah *residual* model regresi mengikuti distribusi normal (Al Fatih *et al.*, 2026). Pengujian ini menggunakan *Kolmogorov-Smirnov* dengan statistik uji sebagai berikut.

$$D_{hitung} = \sup_x |F_n(x) - F_0(x)| \quad (7)$$

d. Uji Asumsi Rata-rata *Residual*

Asumsi rata-rata *residual* bertujuan untuk memastikan bahwa model yang dikonstruksi bersifat *unbiased*. Salah satu syarat yang harus dipenuhi oleh *residual* model regresi adalah nilai harapan atau rata-rata dari *residual* sama dengan nol dan statistik uji dari uji-t menurut Montgomery *et al.* (2021) adalah sebagai berikut.

$$t = \frac{\bar{\varepsilon}}{\frac{S_{\varepsilon}}{\sqrt{n}}} \quad (8)$$

e. Uji Asumsi *Residual* Identik (Homoskedastisitas)

Asumsi *residual* identik digunakan untuk melihat homogenitas dari variansi *residual*, di mana variansi *residual* disyaratkan bernilai konstan. Untuk mendeteksi adanya heteroskedastisitas digunakan uji *Glejser* dan statistik uji dari *Glejser* menurut Hasanah (2008) adalah sebagai berikut.

$$F = \frac{\sum_{i=1}^n (|\hat{e}_i| - |\bar{\hat{e}}|)^2 / (k - 1)}{\sum_{i=1}^n (|e_i| - |\bar{\hat{e}}|)^2 / (n - k)} \quad (9)$$

8. Perbandingan Model *Penalized Spline* dan *Least Square Spline*

Menghitung ukuran kebaikan model (*goodness of fit*) untuk menilai performa dan tingkat akurasi dari model *penalized spline* yang telah terbentuk. Menurut Nurhuda *et al.* (2022), evaluasi ini dilakukan dengan melihat nilai GCV minimum, koefisien determinasi (R^2) maksimum, serta menghitung nilai *Mean Squared Error* (MSE) minimum sebagai indikator tingkat kesalahan prediksi. Rumus GCV sebagai pembanding utama ditetapkan sebagai berikut.

$$GCV(\lambda) = \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{f}_{\lambda}(x_i))^2}{[n^{-1} \text{tr}(I - A(\lambda))]^2} \quad (10)$$

9. Interpretasi Model dan Penarikan Kesimpulan

Menginterpretasikan bentuk kurva tersebut secara ekonomi dan demografi, menganalisis pada interval variabel prediktor yang memberikan dampak paling signifikan, serta menarik kesimpulan akhir.

III. HASIL DAN PEMBAHASAN

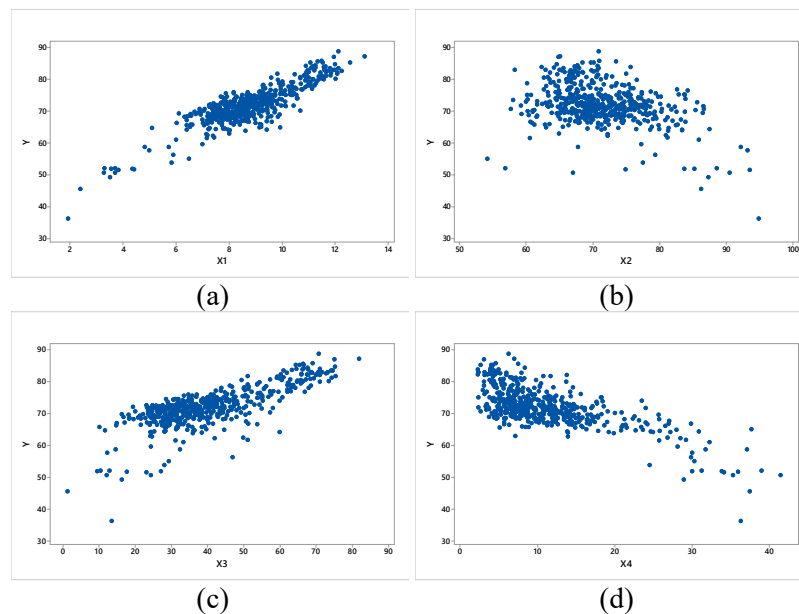
A. Statistika Deskriptif

Hasil analisis statistika deskriptif untuk variabel prediktor (X_1 , X_2 , X_3 , dan X_4) serta variabel respons (Y) menurut Kabupaten/Kota di Indonesia Tahun 2024 disajikan pada **Tabel 2** sebagai berikut.

Tabel 2. Statistika deskriptif variabel

Variabel	Jumlah	Rata-Rata	Standar Deviasi	Minimal	Median	Maksimal
Indeks Pembangunan Manusia (Y)	514	71,9189	6,28770	36,3	71,685	88,77
Rata-rata Lama Sekolah (X_1)	514	8,74086	1,61360	1,92	8,615	13,1
Tingkat Partisipasi Angkatan Kerja (X_2)	514	71,2836	6,22478	54,18	70,65	94,83
Persentase Penduduk Tamat SMA (X_3)	514	40,0316	14,5438	1,29	36,9575	81,85
Persentase Penduduk Miskin (X_4)	514	11,9189	7,07277	2,23	9,32	41,42

Kemudian dilakukan analisis statistika deskriptif menggunakan *scatter plot* pada **Gambar 1** sebagai berikut.



Gambar 1. *Scatter plot* variabel prediktor terhadap Indeks Pembangunan Manusia (IPM)

Scatter plot pada Gambar 2 menunjukkan pola hubungan yang cenderung tidak sepenuhnya linier. X_1 dan X_3 berhubungan positif dengan Y namun variasinya besar pada nilai rendah, X_2 berhubungan negatif landai, sedangkan X_4 berhubungan negatif dengan variasi yang membesar pada nilai tinggi. Pola-pola tersebut mengindikasikan regresi nonparametrik lebih sesuai digunakan untuk mengakomodasi hubungan fleksibel antarvariabel.

B. Deteksi Multikolinieritas

Hipotesis yang digunakan untuk deteksi multikolinieritas adalah sebagai berikut.

H_0 : Tidak terdapat gejala multikolinieritas antar variabel; H_1 : Terdapat gejala multikolinieritas antar variabel
Dengan daerah kritis tolak H_0 jika $VIF > 10$. Hasil uji multikolinieritas disajikan pada tabel berikut.

Tabel 3. Nilai VIF variabel prediktor

Variabel Prediktor	Nilai VIF
Rata-rata Lama Sekolah (X_1)	7,829
Tingkat Partisipasi Angkatan Kerja (X_2)	1,241
Persentase Penduduk Tamat SMA (X_3)	6,132
Persentase Penduduk Miskin (X_4)	1,594

Berdasarkan **Tabel 3**, seluruh variabel prediktor memiliki nilai $VIF < 10$, maka keputusan yang diambil adalah

terima H_0 . Sehingga dapat disimpulkan bahwa tidak terdapat ketergantungan pada variabel prediktor.

C. Analisis Regresi Nonparametrik *Least Square Spline*

Analisis dilakukan menggunakan metode *Least Square Spline* (*LS-Spline*). Pemilihan model optimal dilakukan dengan membandingkan kombinasi 3 orde polinomial (1, 2, dan 3) dengan 4 jumlah titik knot pada masing-masing variabel prediktor, sehingga diperoleh sebanyak 64 kombinasi model. Model terbaik dipilih berdasarkan GCV minimum di antara kombinasi yang memenuhi seluruh asumsi residual. Hasil pengujian menunjukkan orde polinomial 2 sebagai model optimal. Perbandingan GCV untuk 4 kombinasi jumlah titik knot pada orde 2 disajikan pada **Tabel 4**.

Tabel 4. Perbandingan nilai GCV model *LS-Spline* orde 2 pada berbagai jumlah titik knot

Orde	Knot	Titik Knot Optimal	GCV Minimum
2	1	X_1 (8,265); X_2 (71,84); X_3 (33,145); X_4 (8,98)	5,489668
2	2	X_1 (7,8379; 9,2671); X_2 (69,1381; 73,8953); X_3 (27,5545; 43,9056); X_4 (6,9546; 10,5078)	5,484132
2	3	X_1 (7,5575; 8,2650; 9,9575); X_2 (67,9625; 71,8400; 74,8250); X_3 (25,7750; 33,1450; 52,9325); X_4 (6,380; 8,980; 11,425)	5,449987
2	4*	X_1 (7,468; 7,917; 8,886; 10,774); X_2 (67,686; 70,714; 73,14; 75,508); X_3 (23,796; 29,48; 38,695; 58,517); X_4 (6,03; 7,92; 9,622; 11,906)	5,432140*

Berdasarkan **Tabel 4**, model *least square spline* orde 2 dengan 4 titik knot dipilih sebagai model terbaik karena menghasilkan GCV minimum sebesar 5,432140, lebih kecil dibandingkan kombinasi 1, 2, dan 3 titik knot (Tabel 4). Titik-titik knot optimal yang diperoleh untuk X_1 , X_2 , X_3 , dan X_4 berturut-turut adalah (7,468; 7,917; 8,886; 10,774), (67,686; 70,714; 73,140; 75,508), (23,796; 29,480; 38,695; 58,517), dan (6,030; 7,920; 9,622; 11,906), yang menandai lokasi perubahan pola hubungan nonparametrik antara masing-masing prediktor dengan IPM.

D. Pengujian Signifikansi dan Asumsi *Residual Least Square Spline*

Berdasarkan hasil pemilihan model optimal pada sub bab sebelumnya, model *Least Square Spline* orde 2 dengan 4 titik knot ($GCV = 5,432140$) terpilih sebagai kandidat model terbaik. Tahap selanjutnya adalah pengujian signifikansi model dan parameter, serta verifikasi pemenuhan asumsi *residual*.

Model *LS-Spline* orde 2 dengan 4 titik knot selanjutnya diuji signifikansinya secara simultan menggunakan uji F , menghasilkan nilai $F = 143,1457$ yang lebih besar dari $F_{0,05;(24,489)} = 1,5397$ dengan $p - value = 1,019 \times 10^{-2} < \alpha = 0,05$. Model dinyatakan signifikan secara simultan. Pengujian signifikansi parameter secara parsial menunjukkan bahwa parameter β_{11} , β_{12} , δ_{22} , δ_{23} , dan δ_{24} signifikan pada $\alpha = 0,05$. Komponen polinomial X_1 (RLS) menjadi parameter yang paling dominan signifikan, diikuti sebagian komponen *truncated spline* variabel X_2 (TPAK).

Tabel 5. Ringkasan pengujian asumsi *residual* model *LS-Spline* orde 2, 4 titik knot

Uji Asumsi	Statistik Uji	Daerah Kritis	P-Value	Keterangan
Normalitas (Kolmogorov-Smirnov)	0,04199	0,0596	0,3252	Normal
Rata-rata <i>Residual</i> (Uji-t)	0,000	[-1,964; 1,964]	1,000	Mean = 0
Homogenitas (Glejser)	1,8114	2,232	0,1253	Homogen

Berdasarkan **Tabel 5**, model *Least Square Spline* orde 2 dengan 4 titik knot memenuhi ketiga asumsi *residual*. Uji Kolmogorov-Smirnov menunjukkan *residual* berdistribusi normal ($p - value = 0,3252$), uji-t menunjukkan rata-rata *residual* tidak berbeda signifikan dari nol ($p - value = 1,000$), dan uji Glejser menunjukkan tidak terdapat gejala heteroskedastisitas sehingga *residual* bersifat homogen ($p - value = 0,1253$). Dengan demikian, model ini juga memenuhi asumsi dasar *residual* dan layak digunakan untuk analisis lebih lanjut.

E. Analisis Regresi Nonparametrik *Penalized Spline*

Analisis regresi nonparametrik juga dilakukan menggunakan metode *Penalized Spline* sebagai pembanding, dengan prosedur pemilihan model yang sama: membandingkan kombinasi 3 orde polinomial (1, 2, dan 3) dengan 4 jumlah titik knot pada masing-masing variabel prediktor, sehingga diperoleh 64 kombinasi model, di mana model terbaik dipilih berdasarkan GCV minimum di antara kombinasi yang memenuhi seluruh asumsi residual. Hasil

pengujian kembali menunjukkan orde polinomial dua sebagai model optimal. Perbandingan GCV untuk keempat kombinasi jumlah titik knot pada orde tersebut disajikan pada **Tabel 6**.

Tabel 6. Perbandingan nilai GCV model *P-Spline* orde 2 pada berbagai jumlah titik knot

Orde	Knot	Titik Knot Optimal	GCV Minimum	Lambda (λ) Optimal
2	1	X_1 (8,265); X_2 (71,84); X_3 (33,145); X_4 (8,98)	5,488926	0,031
2	2	X_1 (7,8379; 9,2671); X_2 (69,1381; 73,8953); X_3 (27,5545; 43,9056); X_4 (6,9546; 10,5078)	5,480823	0,011
2	3	X_1 (7,5575; 8,2650; 9,9575); X_2 (67,9625; 71,8400; 74,8250); X_3 (25,7750; 33,1450; 52,9325); X_4 (6,380; 8,980; 11,425)	5,431808	0,031
2	4*	X_1 (7,468; 7,917; 8,886; 10,774); X_2 (67,686; 70,714; 73,14; 75,508); X_3 (23,796; 29,48; 38,695; 58,517); X_4 (6,03; 7,92; 9,622; 11,906)	5,388838*	0,071*

Berdasarkan **Tabel 6**, model *P-Spline* orde 2 dengan 4 titik knot menghasilkan GCV minimum sebesar 5,388838, lebih kecil dibandingkan model dengan 1, 2, dan 3 titik knot. Titik knot optimal yang diperoleh adalah X_1 (7,468; 7,917; 8,886; 10,774), X_2 (67,686; 70,714; 73,140; 75,508), X_3 (23,796; 29,480; 38,695; 58,517), dan X_4 (6,030; 7,920; 9,622; 11,906). Titik-titik knot ini mencerminkan perubahan pola hubungan nonparametrik antara masing-masing prediktor terhadap IPM pada segmen nilai tertentu.

F. Pengujian Signifikansi dan Asumsi *Residual Penalized Spline*

Model *Penalized Spline* orde 2 dengan 4 titik knot ($GCV = 5,3888$; $\lambda = 0,071$) terpilih sebagai kandidat model terbaik. Tahap selanjutnya adalah pengujian signifikansi model dan parameter, serta pemenuhan asumsi *residual*.

Uji F simultan menghasilkan $F = 142,0883 > F_{0,05; (24,489)} = 1,5397$ dengan $p - value = 4,928 \times 10^{-203} < \alpha = 0,05$, sehingga model signifikan secara simultan. Pengujian signifikansi parameter secara parsial menunjukkan bahwa parameter β_{11} , β_{12} , δ_{23} , dan δ_{24} memiliki $p - value < 0,05$ sehingga signifikan, sementara parameter lainnya tidak signifikan secara parsial. Hal ini mengindikasikan bahwa komponen polinomial X_1 (RLS) dan sebagian komponen *truncated spline* X_2 (TPAK) memberikan kontribusi paling nyata terhadap model.

Tabel 7. Ringkasan pengujian asumsi *residual* model *Penalized Spline* orde 2, 4 titik knot

Uji Asumsi	Statistik Uji	Daerah Kritis	P-Value	Keterangan
Normalitas (Kolmogorov-Smirnov)	0,03696	0,0596	0,4839	Normal
Rata-rata <i>Residual</i> (Uji-t)	-0,000	[-1,964; 1,964]	1,000	Mean = 0
Homogenitas (Glejser)	2,0242	2,232	0,0898	Homogen

Berdasarkan **Tabel 7**, model *penalized spline* orde 2 dengan 4 titik knot memenuhi ketiga asumsi *residual*. uji Kolmogorov–Smirnov menunjukkan *residual* berdistribusi normal, uji-t menunjukkan rata-rata *residual* tidak berbeda signifikan dari nol, dan uji Glejser menunjukkan tidak terdapat gejala heteroskedastisitas sehingga *residual* bersifat homogen. Dengan demikian, model layak digunakan untuk analisis lebih lanjut.

G. Perbandingan Model dan Pemilihan Model Terbaik

Setelah kedua metode dianalisis, perbandingan dilakukan antara model terbaik *least square spline* dan *penalized spline* berdasarkan GCV, MSE, R^2 , R^2 *Adjusted*, dan uji F. Hasil perbandingan disajikan pada **Tabel 8**.

Tabel 8. Perbandingan kriteria kebaikan model terbaik *Penalized Spline* dan *Least Square Spline*

Metode	Knot	GCV	MSE	R^2 (%)	R^2 adj (%)	F
<i>Penalized Spline*</i>	4*	5,3888*	4,9486*	87,4587	86,9382*	142,09
<i>Least Square Spline</i>	4	5,4321	5,1679	87,5398*	86,9283	143,15

Berdasarkan **Tabel 8**, model *penalized spline* orde 2 dengan 4 titik knot unggul pada kriteria GCV, MSE, dan R^2 *Adjusted*, meskipun R^2 *least square spline* sedikit lebih tinggi dengan selisih yang tidak substansial secara praktis. Mengingat GCV adalah kriteria utama pemilihan model pada regresi nonparametrik *spline* karena kemampuan menyeimbangkan bias dan variansi estimasi, model *penalized spline* orde 2 dengan 4 titik knot ditetapkan sebagai model terbaik untuk memodelkan faktor-faktor yang memengaruhi IPM menurut kabupaten/kota di Indonesia 2024.

Keunggulan *penalized spline* bersumber dari mekanisme penalti dengan parameter λ yang mengendalikan kompleksitas kurva tanpa mengorbankan ketepatan *fitting*. Berbeda dengan *least square spline* yang murni meminimumkan jumlah kuadrat galat sehingga lebih rentan *overfitting* pada pola hubungan kompleks.

Persamaan model terbaik yang terbentuk dapat dituliskan secara umum sebagai berikut.

$$\hat{y} = 0,2031 + f_1(X_1) + f_2(X_2) + f_3(X_3) + f_4(X_4) \quad (11)$$

di mana $f_k(X_k)$ merupakan komponen *spline* untuk variabel prediktor ke- k yang terbentuk dari suku polinomial dan suku *truncated spline* dengan 4 titik knot. Setiap komponen menangkap pola hubungan antara masing-masing prediktor dan IPM, di mana variabel Rata-rata Lama Sekolah menunjukkan hubungan positif yang melambat, variabel TPAK menunjukkan pola optimal pada kisaran 72–74%, variabel Persentase Tamat SMA memperlihatkan efek U-terbalik, dan variabel Persentase Penduduk Miskin menunjukkan dampak negatif yang semakin melemah pada nilai tinggi.

Dari hasil yang diperoleh, didapatkan fungsi untuk setiap variabel yaitu

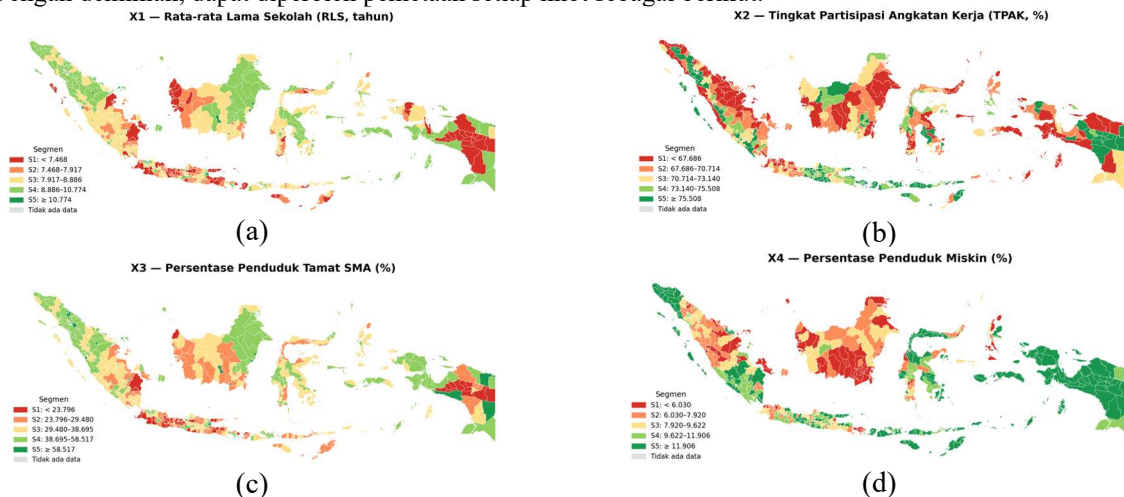
$$f_1(X_1) = \begin{cases} -4,2534X_1^2 + 5,6979X_1 & X_1 < 7,468 \\ -2,9394X_1^2 - 13,9359X_1 + 73,3832 & 7,468 \leq X_1 < 7,917 \\ -1,8287X_1^2 - 31,5266X_1 + 143,0970 & 7,917 \leq X_1 < 8,886 \\ -1,5762X_1^2 - 36,0140X_1 + 163,0186 & 8,886 \leq X_1 < 10,774 \\ -2,2412X_1^2 - 21,6848X_1 + 85,9373 & X_1 \geq 10,774 \end{cases} \quad (12)$$

$$f_2(X_2) = \begin{cases} -0,3286X_2^2 + 0,3829X_2 & X_2 < 67,686 \\ -0,8042X_2^2 + 64,7833X_2 - 2175,8877 & 67,686 \leq X_2 < 70,714 \\ 0,8370X_2^2 - 167,3359X_2 + 6020,7596 & 70,714 \leq X_2 < 73,14 \\ -1,7271X_2^2 - 207,6411X_2 + 7685,9662 & 73,14 \leq X_2 < 75,508 \\ 0,9660X_2^2 - 199,1164X_2 + 7640,2859 & X_2 \geq 75,508 \end{cases} \quad (13)$$

$$f_3(X_3) = \begin{cases} -0,6707X_3^2 + 0,2705X_3 & X_3 < 23,796 \\ 0,9423X_3^2 - 76,5287X_3 + 912,2840 & 23,796 \leq X_3 < 29,48 \\ -0,7482X_3^2 + 23,1324X_3 - 556,8734 & 29,48 \leq X_3 < 38,369 \\ 0,0691X_3^2 - 40,1566X_3 + 667,4273 & 38,369 \leq X_3 < 58,517 \\ -0,7816X_3^2 + 59,3637X_3 + 2273,3295 & X_3 \geq 58,517 \end{cases} \quad (14)$$

$$f_4(X_4) = \begin{cases} -0,1964X_4^2 - 0,1302X_4 & X_4 < 6,03 \\ 0,4297X_4^2 - 7,6785X_4 + 22,7529 & 6,03 \leq X_4 < 7,92 \\ -0,1344X_4^2 + 1,2567X_4 - 12,6070 & 7,92 \leq X_4 < 9,622 \\ -0,5387X_4^2 + 9,0385X_4 - 50,0727 & 9,622 \leq X_4 < 11,906 \\ 1,1139X_4^2 - 30,3295X_4 + 184,5289 & X_4 \geq 11,906 \end{cases} \quad (15)$$

Dengan demikian, dapat diperoleh pemetaan setiap knot sebagai berikut.



Gambar 2. Pemetaan wilayah setiap variabel prediktor berdasarkan segmen.

Pemetaan menunjukkan RLS tinggi terkonsentrasi di Jawa dan beberapa kota besar Sumatra, sementara RLS rendah banyak ditemukan di Papua, NTT, dan pedalaman Kalimantan. TPAK tinggi justru lebih banyak di Indonesia timur, mengindikasikan tingginya partisipasi kerja tidak selalu mencerminkan kondisi ekonomi yang baik karena masih didominasi sektor informal. Persentase tamat SMA tidak menunjukkan pola regional yang jelas dan lebih dipengaruhi karakteristik serta kebijakan lokal. Kemiskinan menunjukkan pola spasial yang tegas yaitu, rendah di Jawa dan Bali, tinggi di Papua, Papua Barat, Maluku, dan NTT yang sejalan dengan pola RLS, yang mengindikasikan keterkaitan antara rendahnya pendidikan dan tingginya kemiskinan serta kesenjangan pembangunan Indonesia barat–timur, sebagaimana ditunjukkan oleh Kurniasari dan Oktavilia (2023) yang menemukan perbedaan determinan kemiskinan yang signifikan antara wilayah Indonesia barat dan timur. Temuan ini juga sejalan dengan hasil penelitian Firmansyah dan Khairunnisa (2023) yang menunjukkan bahwa rata-rata lama sekolah (RLS) dan tingkat partisipasi angkatan kerja (TPAK) berpengaruh signifikan terhadap penurunan persentase penduduk miskin di Indonesia.

IV. KESIMPULAN

Hasil analisis menunjukkan hubungan nonlinier dan nonparametrik antara rata-rata lama sekolah, TPAK, persentase tamat SMA, dan kemiskinan terhadap IPM kabupaten/kota di Indonesia tahun 2024, sehingga regresi nonparametrik lebih tepat digunakan. Model *penalized spline* orde dua dengan empat titik knot memberikan performa terbaik (GCV 5,3888; MSE 4,9486; *Adjusted R*² 86,94%). Rata-rata lama sekolah berpengaruh positif, TPAK memiliki rentang optimal, pengaruh tamat SMA menguat pada nilai tinggi, dan dampak negatif kemiskinan melemah pada wilayah dengan kemiskinan sangat tinggi. Pemetaan spasial mengonfirmasi kesenjangan pembangunan Indonesia barat–timur yang terkait rendahnya pendidikan dan tingginya kemiskinan, dan diharapkan menjadi dasar pengembangan regresi nonparametrik spasial serta referensi kebijakan peningkatan IPM.

UCAPAN TERIMA KASIH

Penulis mengucapkan rasa terima kasih, apresiasi, serta penghargaan setinggi-tingginya kepada Badan Pusat Statistik (BPS) atas penyediaan data publikasi resmi yang menjadi basis utama analisis dalam pemodelan ini. serta kepada rekan-rekan sejawat atas diskusi dan dukungan dalam penyempurnaan artikel ini.

DAFTAR PUSTAKA

- Amalina, R. N., & Sulistiowati, D. (2026). Perbandingan Regresi Nonparametrik B-Splines dan P-Splines dalam Memodelkan Hubungan Indeks Pembangunan Manusia dan Persentase Penduduk Miskin di Sumatera Utara. *Imajiner: Jurnal Matematika Dan Pendidikan Matematika*, 8(2), 555–564.
- Arman, M. R. (2025). PEMODELAN *GEOGRAPHICALLY WEIGHTED NONPARAMETRIC REGRESSION* DENGAN SPLINE PENALIZED (Studi Kasus: Persentase Penduduk Miskin di Provinsi Sulawesi Selatan Tahun 2023). Universitas Hasanuddin.
- Chamidah, N., & Lestari, S. P. (2022). *Analisis Regresi Nonparametrik dengan Perangkat Lunak R*. Airlangga University Press
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.
- Firmansyah, A., & Khairunnisa, D. (2023). Dampak pendidikan terhadap pengentasan kemiskinan di Indonesia dengan menggunakan analisis data panel. *Emerging Statistics and Data Science Journal*. <https://journal.uin.ac.id/esds/article/view/27956>
- Fitriyah, Z., Irsalina, S., Rizq Herlandy K, A., & Widodo, E. (2021). ANALISIS FAKTOR YANG BERPENGARUH TERHADAP IPM MENGGUNAKAN REGRESI LINEAR BERGANDA. 2(3). <https://doi.org/10.46306/lb.v2i3>
- Fadlirhohim, R. D., Sifriyani, S., & Dani, A. T. R. (2024). MODELING STUNTING PREVALENCE IN INDONESIA USING SPLINE TRUNCATED SEMIPARAMETRIC REGRESSION. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 18(3), 2015-2028.
- Hasanah, N. N. (2008). *Pengujian heterokedastisitas pada regresi non linear dengan menggunakan uji Glejser* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim). 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.

- Kurniasari, A., & Oktavilia, S. (2023). Determinants of poverty in Western Indonesia and Eastern Indonesia. *Economics Development Analysis Journal*, 12(1), 95.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Nurhuda, G. N., Wasono, W., & Nohe, D. A. (2022). Nonparametric Regression Modeling Based on Spline Truncated Estimator on Simulation Data. *Jurnal Matematika, Statistika dan Komputasi*, 19(1), 172-182.
- Risnita, R. (2024). Pendekatan penelitian kuantitatif dan kualitatif serta tahapan penelitian. *Jurnal Genta Mulia*, 15(1), 82–92.
- Rumanama, S. M., Persulesy, E. R., & Leleury, Z. A. (2022). PEMILIHAN MODEL REGRESI TERBAIK DENGAN MENGGUNAKAN METODE STEPWISE (Studi Kasus: Data IPM Indonesia Tahun 2020). *PARAMETER: Jurnal Matematika, Statistika Dan Terapannya*, 1(1), 69–78. <https://doi.org/10.30598/parameterv1i1pp69-78>
- Sa'adah, N., Andiansyah, T., Hidayat, T., Amni, C., & Sunyanti, S. (2025). Analisis Distribusi dan Tren Nilai Mahasiswa dengan Pendekatan Statistik Deskriptif dan Visualisasi Data. *Sudo Jurnal Teknik Informatika*, 4(3), 203-211.
- Salma, 2020. Uji Asumsi Klasik Adalah: Yuk Pelajari Serta Ketahui Jenisnya. <https://tambahpinter.com/uji-asumsi-klasik-serta-jenisnya/>
- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics (SIAM).
- Waruwu, M., Pu'at, S. N., Utami, P. R., Yanti, E., & Rusydiana, M. (2025). Metode Penelitian Kuantitatif: Konsep, Jenis, Tahapan dan Kelebihan. *Jurnal Ilmiah Profesi Pendidikan*, 10(1), 917–932. <https://doi.org/10.29303/jipp.v10i1.3057>