

# Penalized Spline Regression Modeling on the Human and Cultural Development Index (IPMK) for 2022

Sarmilah, Fadhilah Fitri\*, Musthafa Imran

Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: [fadhilahfitri@fmipa.unp.ac.id](mailto:fadhilahfitri@fmipa.unp.ac.id)

Submitted : 05 Oktober 2025

Revised : 04 November 2025

Accepted : 01 Desember 2025

## ABSTRACT

*Human and cultural development is a multidimensional phenomenon whose relationship with socioeconomic factors is often complex and nonlinear, making it challenging to model with conventional parametric approaches. This study aims to model the influence of socioeconomic variables on the Human and Cultural Development Index (IPMK) across 34 provinces in Indonesia in 2022 using the nonparametric Penalized Spline (P-spline) regression method within a Generalized Additive Model (GAM) framework. Secondary data from the Central Statistics Agency (BPS) were used, with predictor variables including School Participation Rate (APS), percentage of access to safe drinking water, Gini Ratio, per capita expenditure, average years of schooling (RLS), and open unemployment rate (TPT). Initial data exploration via scatterplots confirmed nonlinear relationship patterns between the predictor variables and IPMK. The best model was obtained using a first-order cubic spline with 10 knot points, selected based on the minimum Generalized Cross Validation (GCV) criterion. The modeling results demonstrated excellent performance, with an Adjusted R<sup>2</sup> value of 0.842 and a Deviance Explained of 92.3%. Significance analysis indicated that access to safe drinking water, per capita expenditure, average years of schooling, and the open unemployment rate significantly influence IPMK. Visual interpretation of the significant spline curves revealed informative relationship patterns, such as the diminishing returns effect of per capita expenditure. This study concludes that the P-spline approach is effective and interpretable for modeling complex nonlinear relationships in development data, providing a richer evidence base for policy formulation.*

**Keywords:** Penalized Spline, Nonparametric Regression, Generalized Cross Validation, Human and Cultural Development Index, Generalized Additive Model.



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

Pembangunan manusia adalah aspek yang fundamental dalam peningkatan kualitas hidup dan kesejahteraan suatu negara. Kerangka awal pembangunan manusia diformulasikan oleh UNDP melalui *Human Development Report* 1990, yang menekankan bahwa pembangunan harus berorientasi pada perluasan pilihan manusia di bidang kesehatan, pendidikan, serta standar hidup yang layak (UNDP, 1990). Pemikiran ini sejalan dengan konsep kapabilitas Amartya Sen dalam *Development as Freedom*, yang menegaskan bahwa pembangunan tidak hanya diukur melalui akumulasi sumber daya ekonomi, tetapi melalui kemampuan (*capabilities*) individu untuk menjalani kehidupan yang dinilai bermakna (Sen, 1999). Dalam konteks pendidikan, pendekatan kapabilitas menekankan pemberian keterampilan dan pengetahuan yang dibutuhkan individu untuk mencapai tujuan hidup, sekaligus memperluas kebebasan, martabat, dan kesejahteraan manusia, bukan sekedar memenuhi kebutuhan pasar tenaga kerja (Iswahyudi dan Asnawi, 2024). Di Indonesia, selain Indeks Pembangunan Manusia (IPM) yang terstandar global, dikembangkan pula Indeks Pembangunan Manusia dan Kebudayaan (Indeks PMK) sebagai alat ukur komprehensif yang menilai capaian pembangunan manusia secara multidimensional.

Indeks PMK adalah alat ukur kuantitatif berbasis data yang dikembangkan oleh Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan (Kemenko PMK) bersama dengan Badan Pusat Statistik (BPS) untuk mengukur capaian pembangunan manusia dan kebudayaan secara menyeluruh dan terukur (Kemenko PMK, 2023). Menurut Satu Data (Kemenko PMK, 2025), Indeks PMK berfungsi sebagai pelengkap IPM dengan memandang kesejahteraan sebagai konsep multidimensional (*multidimensional aspect*), yang mencakup dimensi karakter,

kapabilitas dan budaya manusia Indonesia. Indeks PMK disusun untuk membantu pemantauan dan pengendalian program pembangunan, sehingga dapat menjadi acuan dalam perumusan kebijakan yang lebih tepat sasaran.

Untuk melakukan peningkatan terhadap capaian Indeks PMK, maka diperlukan pemahaman secara mendalam terkait faktor-faktor yang memengaruhinya. Berbagai studi mengenai determinan IPM di Indonesia dan kawasan ASEAN secara konsisten menyoroti pentingnya variabel-variabel sosial ekonomi. Variabel seperti Angka Partisipasi Sekolah (APS), Akses terhadap Air Minum Layak, *Gini Ratio* (IGR), Pengeluaran per Kapita (IPK), Rata-Rata Lama Sekolah (RLS), dan Tingkat Pengangguran Terbuka (TPT) merepresentasikan dimensi penting dalam pembangunan manusia, yaitu akses terhadap pendidikan, kondisi kesehatan lingkungan, distribusi pendapatan, tingkat kesejahteraan ekonomi, dan kondisi pasar tenaga kerja (Andriani dan Alfian, 2020). Namun, hubungan antara variabel-variabel sosial ekonomi tersebut dengan indeks pembangunan seperti Indeks PMK seringkali tidak sederhana. Kondisi sosial ekonomi yang sangat beragam antarprovinsi di Indonesia mengindikasikan bahwa asumsi hubungan linier—dimana peningkatan satu unit pada variabel prediktor akan memberikan dampak yang sama di semua tingkat—kemungkinan besar tidak valid. Sebagai contoh, dampak peningkatan Pengeluaran per Kapita terhadap Indeks PMK mungkin sangat besar pada tingkat pendapatan rendah, namun efeknya bisa jadi menurun setelah mencapai tingkat kesejahteraan tertentu (*diminishing returns*). Hubungan yang kompleks dan nonlinier semacam ini tidak dapat ditangkap secara akurat oleh model regresi linier konvensional.

Oleh karena itu, diperlukan pendekatan statistik dan metode yang fleksibel, salah satunya yaitu regresi nonparametrik, khususnya *Penalized Spline (P-spline)*. *P-spline*, yang sering diimplementasikan dalam kerangka *Generalized Additive Models* (GAMs), memungkinkan estimasi kurva regresi yang fleksibel mengikuti pola data tanpa memaksakan bentuk fungsional tertentu, sekaligus menjaga kehalusan kurva melalui mekanisme penalti agar terhindar dari *overfitting*.

Berdasarkan hal tersebut, tujuan penelitian ini ialah untuk menganalisis pengaruh variabel-variabel sosial ekonomi terhadap Indeks PMK di Indonesia pada tahun 2022 menggunakan metode regresi *P-spline*. Pendekatan ini diharapkan tidak hanya mengidentifikasi faktor-faktor signifikan, tetapi juga mengungkap bentuk hubungan nonlinier yang mendasarinya. Hasil analisis ini diharapkan dapat memberikan gambaran yang lebih akurat mengenai determinan pembangunan manusia dan kebudayaan, serta menjadi landasan ilmiah yang kuat untuk perumusan kebijakan publik yang berbasis bukti empiris.

## II. METODE PENELITIAN

### A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Objek pengamatan dalam penelitian ini adalah data mengenai 34 provinsi di Indonesia tahun 2022, serta variabel yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Variabel penelitian

Variabel	Keterangan
Variabel Respon (y)	Indeks Pembangunan Manusia dan Kebudayaan (IPMK)
Variabel Prediktor (x)	
$X_1$	Angka Partisipasi Sekolah (APS) (%)
$X_2$	Persentase rumah tangga yang memiliki akses terhadap sumber air minum layak (%)
$X_3$	<i>Gini ratio</i> (IGR)
$X_4$	Pengeluaran per kapita (Rupiah)
$X_5$	Rata-rata Lama Sekolah (RLS) (Tahun)
$X_6$	Tingkat Pengangguran Terbuka (TPT) (%)

(Sumber: Data BPS, 2022).

Secara sistematis, tahapan analisis data Indeks PMK di provinsi Indonesia menggunakan estimator *p-spline* adalah sebagai berikut:

1. Eksplorasi data
2. Pemodelan Statistik
  - a. Pemilihan Kerangka Kerja Model
  - b. Spesifikasi Fungsi Penghalus dengan *Penalized Spline (P-spline)*
  - c. Pemilihan Parameter Model Optimal
3. Evaluasi dan Validasi Model
  - a. *Goodness-of-Fit*

- b. Uji Signifikansi
- c. Analisis Diagnostik Residual
- 4. Interpretasi Hasil

### B. Regresi Nonparametrik

Regresi nonparametrik diartikan sebagai suatu pendekatan dalam analisis regresi di mana tidak ada penentuan pasti terhadap bentuk kurva dari fungsi regresinya. Pada metode ini, kurva regresi hanya dianggap memiliki pola yang halus (smooth), yang berarti termasuk dalam suatu ruang fungsi tertentu, sehingga memiliki fleksibilitas tinggi (Winart, 2020). Model regresi nonparametrik dapat dirumuskan seperti pada Persamaan (1).

$$y_i = f(x_i) + \varepsilon_i, \quad (1)$$

dengan,

- $y_i$  : variabel respon ke  $i$
- $f(x_i)$  : fungsi *smooth* yang tidak diketahui ke  $-i$
- $x_i$  : variabel prediktor
- $\varepsilon_i$  : galat acak yang diasumsikan independen, identik, dan berdistribusi normal dengan *mean* nol dan varians  $\sigma^2$  ( $N(0, \sigma^2)$ ).

Tujuan utama dari regresi nonparametrik adalah untuk mengestimasi fungsi  $f$  yang tidak diketahui bentuknya tersebut langsung dari data. Pendekatan ini berguna ketika eksplorasi data awal, seperti melalui *scatterplot*, menunjukkan adanya pola yang kompleks atau nonlinier yang tidak ditangkap secara akurat oleh model linier. Terdapat berbagai metode untuk mengestimasi fungsi  $f$ , dan salah satu yang paling populer dan fleksibel adalah *Penalized Spline* (*P-spline*) yang menjadi fokus dalam penelitian ini.

### C. Generalized Additive Model (GAM)

*Generalized Additive Model* (GAM) merupakan perluasan dari model linier tergeneralisasi (GLM) yang memungkinkan hubungan antara variabel respon dan prediktor dimodelkan secara nonparametrik melalui fungsi-fungsi penghalus (*smooth functions*) (Nguyen, 2020).

GAM dapat dirumuskan sebagai berikut:

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_j(x_{ij}) + \varepsilon_i, \quad i = 1, 2, 3, \dots, n \quad (2)$$

dengan,

- $y_i$  : variabel respon ke- $i$
- $\beta_0$  : intersep
- $f_j(x_{ij})$  : fungsi penghalus nonparametrik untuk variabel prediktor ke- $j$
- $\varepsilon_i$  : galat acak yang diasumsikan independen, identik, dan berdistribusi normal dengan *mean* nol dan varians  $\sigma^2$  ( $N(0, \sigma^2)$ ).

### D. Model Regresi *Penalized Spline*

Regresi *Penalized Spline* (*P-spline*) merupakan salah satu metode *smoothing* yang populer karena kesederhanaannya. Fungsi penghalus  $f_i$  dalam model GAM diestimasi menggunakan basis *P-spline*, dengan mengombinasikan basis fungsi *B-spline* dengan penalti pada koefisien untuk mengontrol tingkat kehalusan kurva. Fungsi  $f(x)$  untuk satu prediktor dapat direpresentasikan sebagai kombinasi linier dari fungsi basis *B-spline*:

$$f(x) = \sum_{i=1}^{m+k} \beta_i B_{i-m,m}(x) \quad (3)$$

dengan  $B_{i-m,m}$  adalah basis *B-spline* yang berorde  $m$  pada titik knot ke- $i$  dan  $\beta_i$  merupakan parameter regresi *B-spline* pada titik knot ke- $i$ . Untuk menghindari *overfitting*, *P-spline* meminimalkan jumlah kuadrat sisa (*residual sum of squares*) yang diberi penalti, sehingga diperoleh fungsi *penalized least square* (PLS):

$$PLS(\lambda) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int [f''(t)]^2 dt \quad (4)$$

(Widyowati, 2016)

Parameter  $\lambda \geq 0$  adalah parameter pernghalus (smoothing parameter) yang mengontrol keseimbangan antara *goodness-of-fit* model terhadap data dan kehalusan fungsi. Penambahan nilai  $\lambda$  akan menghasilkan kurva yang lebih bergelombang dan lebih dekat dengan data.

Pemilihan parameter smoothing  $\lambda$  dalam metode ini dilakukan dengan menggunakan *Generalized Cross Validation* (GCV). Menurut Lee dalam Wahyuningsih, et al, 2019, parameter penghalus yang optimal ditentukan berdasarkan nilai GCV terkecil. Fungsi GCV dirumuskan sebagai berikut:

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left(\frac{1}{n} \text{tr}[I - H(\lambda)]\right)^2} \quad (5)$$

(Eubank, 1999)

dengan

$$MSE(\lambda) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}_{\lambda}(x_i))^2 \quad (6)$$

(Eubank, 1999)

dimana,

$GCV(\lambda)$  : nilai GCV untuk parameter *smoothing*  $\lambda$

$n$  : jumlah observasi

$y_i$  : nilai aktual

$\hat{f}_{\lambda}(x_i)$  : nilai estimasi

$\text{tr}(A_{\lambda})$  : jejak (*trace*) dari matriks proyeksi  $A_{\lambda}$  (*effective degrees of freedom/edf*)

### III. HASIL DAN PEMBAHASAN

#### A. Analisis Deskriptif dan Eksplorasi Data

Analisis deskriptif dilakukan untuk melihat gambaran secara umum mengenai karakteristik data setiap variabel yang digunakan dalam penelitian. Tabel 2 menyajikan ringkasan statistik untuk Indeks PMK dan enam variabel prediktor di 34 provinsi di Indonesia pada tahun 2022.

Tabel 2. Statistika deskriptif variabel penelitian

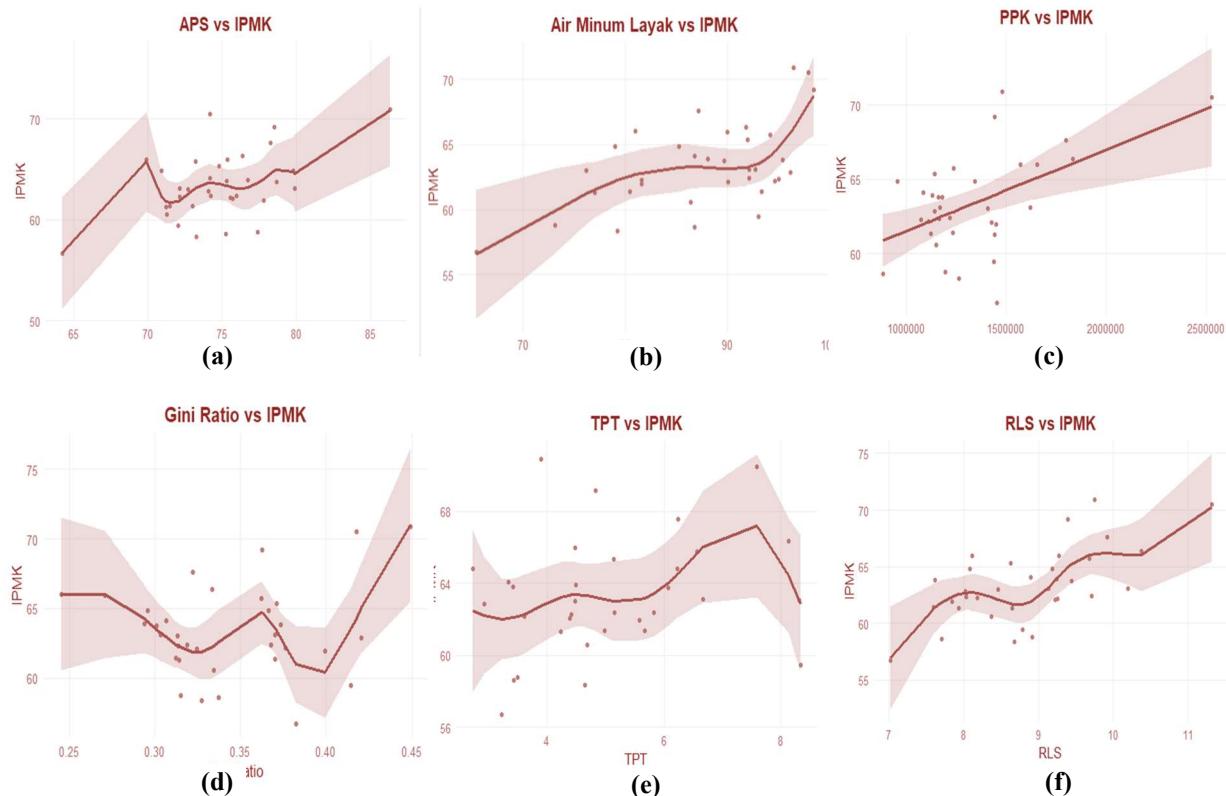
Ukuran	Variabel						
	Indeks PMK (Y)	Angka Partisipasi Sekolah (X <sub>1</sub> )	Air Minum Layak (X <sub>2</sub> )	Gini Ratio (X <sub>3</sub> )	Pengeluaran Per Kapita (X <sub>4</sub> )	Rata Rata Lama Sekolah (X <sub>5</sub> )	Tingkat Pengangguran Terbuka (X <sub>6</sub> )
N	34	34	34	34	34	34	34
Rata-Rata	63.37	74.78	87.64	0.3437	1338086	8.839	5.045
Median	63.05	74.53	89.83	0.3340	1234662	8.835	4.665
Standar Deviasi	3.28	3.78	7.86	0.046	306707.9	0.92	1.54
Nilai Minimum	56.71	64.27	65.39	0.2455	884102	7.020	2.725
Nilai Maksimum	70.91	86.31	98.42	0.4490	2525347	11.310	8.330

(Sumber: Data BPS 2022, diolah).

Tabel 2 menyajikan statistik deskriptif dari tujuh variabel yang diamati pada 34 observasi. Secara umum, rata-rata Indeks Pembangunan Manusia dan Kebudayaan (Indeks PMK) adalah 63.37 dengan standar deviasi 3.28, menunjukkan variasi sedang antar daerah. Angka Partisipasi Sekolah dan akses terhadap Air Minum Layak cukup

tinggi, masing-masing dengan rata-rata 74.78 dan 87.64. *Gini Ratio* rata-rata sebesar 0.3437 mencerminkan ketimpangan pendapatan sedang. Pengeluaran per Kapita memiliki variasi yang cukup besar, terlihat dari standar deviasi yang tinggi (306707.9). Rata-rata lama sekolah adalah 8.839 tahun, sedangkan tingkat pengangguran terbuka rata-rata sebesar 5.045%. Nilai minimum dan maksimum pada tiap variabel menunjukkan adanya disparitas yang cukup besar antar wilayah yang diamati.

Sebelum dilakukannya pemodelan dengan metode Spline, perlu dilihat pola hubungan variabel respon (Indeks PMK) dan masing-masing variabel prediktor dengan menggunakan *scatterplot* dan dapat dilihat pada Gambar 1.



**Gambar 1.** (a) *Scatterplot* APS dengan IPMK, (b) *Scatterplot* Air Minum Layak dengan IPMK, (c) *Scatterplot* PPK dengan IPMK, (d) *Scatterplot* Gini Ratio dengan IPMK, (e) *Scatterplot* TPT dengan IPMK, dan (c) *Scatterplot* RLS dengan IPMK (Sumber: data diolah )

Pada Gambar 1 dapat dilihat bahwa ke-6 *scatterplot* menunjukkan bahwa seluruh variabel memiliki hubungan non-linier terhadap Indeks PMK, dengan karakteristik kelengkungan yang bervariasi. Selain itu, titik-titik data juga tidak menunjukkan bentuk pola tertentu, yang berarti tidak adanya kejelasan fungsi regresi. Karena informasi tentang bentuk fungsi terbatas dan polanya tidak jelas, maka regresi nonparametrik dapat digunakan untuk memodelkan data, salah satunya dengan regresi spline dengan metode *penalized spline (P-spline)*.

## B. Pemodelan Indeks PMK Regresi *P-spline*

Pada regresi *P-spline*, GCV minimum digunakan sebagai penentu titik knot yang optimal. Dalam penelitian ini, kombinasi orde 1–3 serta jumlah titik knot 5,10 dan 15 dipilih karena rentang tersebut mewakili tingkat kelengkungan dan kompleksitas spline yang umum direkomendasikan untuk data sosial-ekonomi yang cenderung halus. Orde 1–3 mencakup perubahan fleksibilitas basis spline tanpa menimbulkan fluktuasi kurva yang tidak stabil, sedangkan jumlah knot 5, 10, dan 15 memungkinkan pengujian kepadatan knot dari rendah hingga menengah untuk menilai risiko *underfitting* maupun *overfitting*. Hasil pemilihan konfigurasi optimal disajikan pada Tabel 3.

**Tabel 3.** Pemilihan titik knot optimal

Orde	Banyak Titik Knot	GCV	MSE
1	5	4.779561	2.156756
	10	<b>0.126215</b>	<b>0.000369</b>
	15	1.856335	0.044590
2	5	5.006528	2.540122
	10	3.199675	0.709235
	15	0.986712	0.016268
3	5	5.024014	2.545103
	10	4.082185	1.161112
	15	3.400321	0.632318

(Sumber: Data BPS 2022, diolah).

Tabel 3 menunjukkan bahwa model dengan orde 1 dan 10 titik knot merupakan model *P-spline* dengan titik knot yang optimal karena memiliki nilai GCV yang paling kecil, yaitu 0.126215. Hasil ini menunjukkan bahwa penelitian terdahulu, yaitu Putri & Dwidayati dalam Kusunartutik & Dwidayati, 2022 yang menyatakan bahwa hasil dari nilai GCV yang semakin kecil tidak selalu dari penambahan titik knot. Pada Tabel 3, dapat dilihat bahwa nilai GCV terkecil didapatkan pada 10 titik knot, bukan 15 knot.

Setelah didapatkan model regresi *P-spline* dengan titik knot optimal, model dievaluasi untuk melihat kinerjanya secara keseluruhan yang dapat dilihat pada Tabel 4.

**Tabel 4.** Ringkasan kinerja model regresi

Ukuran Evaluasi	Nilai
R-squared (Adjusted)	0.842
Deviance Explained	92.3%

(Sumber: Data BPS 2022, diolah).

Hasil pada tabel 4 menunjukkan bahwa model memiliki daya jelaskan yang sangat tinggi. Nilai *Adjusted R<sup>2</sup>* sebesar 0.842 mengindikasikan bahwa 84.2% variasi pada Indeks PMK mampu dijelaskan oleh variabel-variabel prediktor dalam model. Nilai *Deviance Explained* 92.3% juga mengonfirmasi bahwa model berhasil menangkap sebagian besar variabilitas dalam data.

### C. Uji Signifikansi

Setelah diperoleh informasi bahwa model cukup baik dalam menjelaskan data, maka dilakukan uji signifikansi untuk mengetahui variabel prediktor mana yang berpengaruh secara signifikan terhadap Indeks PMK. Sebuah variabel dianggap berpengaruh signifikan apabila memiliki nilai p kurang dari atau sama dengan 0.05. Uji signifikansi dari masing-masing variabel dapat dilihat pada Tabel 5.

**Tabel 5.** Uji Signifikansi

Variabel	edf	Ref.df	F	P-value
<i>s(X<sub>1</sub>)</i>	0.04886	7	0.011	0.110239
<i>s(X<sub>2</sub>)</i>	2.86929	9	1.418	0.003915
<i>s(X<sub>3</sub>)</i>	6.32949	9	3.632	0.000940
<i>s(X<sub>4</sub>)</i>	4.45687	7	4.029	0.000363
<i>s(X<sub>5</sub>)</i>	1.39718	9	0.393	0.045607
<i>s(X<sub>6</sub>)</i>	1.86351	9	0.518	0.032301

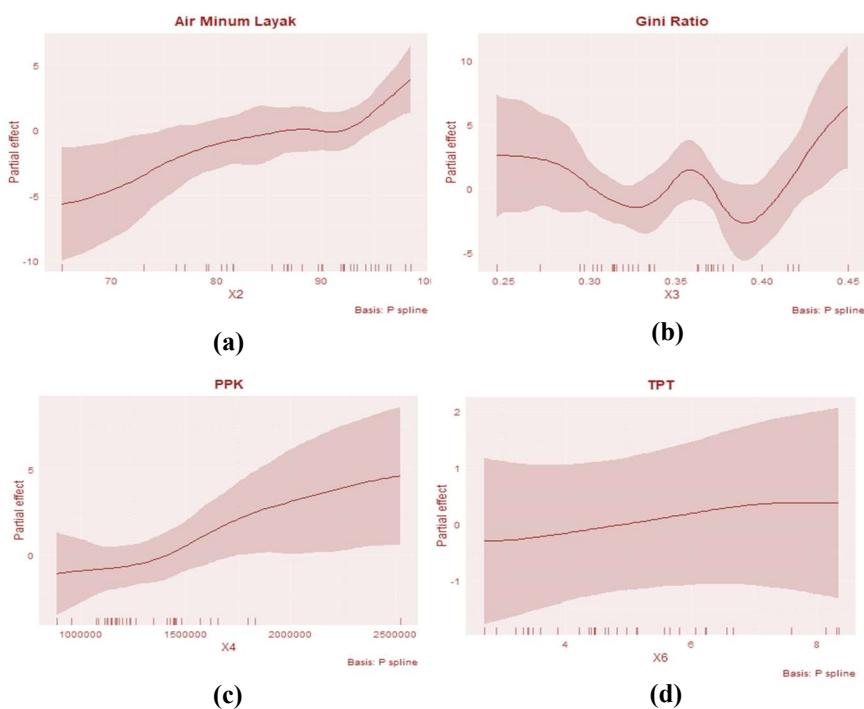
(Sumber: Data BPS 2022, diolah).

Tabel 5 menunjukkan hasil uji signifikansi untuk setiap fungsi penghalus. Variabel  $X_2, X_3, X_4$  dan  $X_6$  memiliki nilai p  $\leq 0.05$ , dengan demikian variabel Akses Air Minum Layak, *Gini Ratio* (IGR), Pengeluaran per Kapita, dan Tingkat Pengangguran Terbuka (TPT) memiliki pengaruh yang signifikan secara statistik terhadap Indeks PMK. Artinya, aspek layanan dasar, ketimpangan ekonomi, daya beli masyarakat, serta ketersediaan lapangan pekerjaan memiliki kontribusi nyata dalam membentuk kualitas pembangunan manusia dan kebudayaan di suatu wilayah.

Sementara itu, untuk Angka Partisipasi Sekolah (APS) ( $X_1$ ) dan Rata-Rata Lama Sekolah (RLS) ( $X_5$ ) tidak menunjukkan pengaruh yang signifikan dalam model ini. Temuan ini menarik, karena indikator pendidikan menjadi komponen penting dalam Indeks Pembangunan Manusia (IPM) (BPS, 2024), tetapi ternyata tidak berpengaruh terhadap

Indeks PMK, yang tidak hanya mencakup dimensi pendidikan, kesehatan dan ekonomi, tetapi juga nilai-nilai kebudayaan, partisipasi sosial dan identitas masyarakat.

Nilai edf yang lebih besar dari 1 untuk variabel-variabel signifikan mengonfirmasi bahwa hubungan antara variabel-variabel tersebut dengan Indeks PMK bersifat nonlinier. Untuk memahami bagaimana variabel-variabel signifikan memengaruhi Indeks PMK, kurva spline parsial dari model GAM diplot dan ditunjukkan pada Gambar 2.

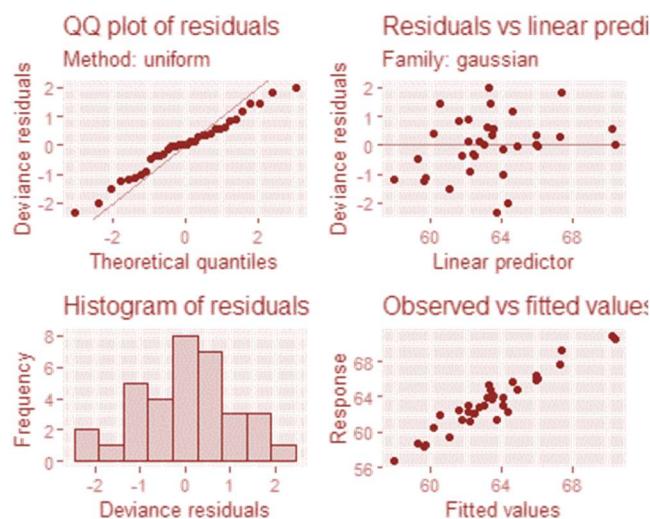


**Gambar 2.** (a) Efek parsial akses air minum layak terhadap Indeks PMK, (b) Efek parsial *gini ratio* terhadap Indeks PMK, (c) Efek parsial PPK terhadap Indeks PMK, dan (d) Efek parsial TPT terhadap Indeks PMK (Sumber: data diolah)

Gambar 2 menunjukkan efek marginal dari setiap prediktor terhadap Indeks PMK, dimana kurva spline parsial mengonfirmasi bahwa hubungan antara variabel-variabel prediktor dan Indeks PMK bersifat nonlinier. Akses Air Minum Layak ( $X_2$ ) dan Pengeluaran per Kapita (PPK) ( $X_4$ ) memperlihatkan pola positif yang stabil, sehingga keduanya tampak sebagai faktor yang paling kuat dalam meningkatkan Indeks PMK. Sementara itu, *Gini Ratio* ( $X_3$ ) menunjukkan pola nonlinier yang lebih kompleks, dengan pengaruh negatif pada rentang tertentu. Ini mengindikasikan bahwa penting dilakukan pemerataan distribusi pendapatan dalam pembangunan manusia dan kebudayaan.

Untuk variabel Tingkat Pengangguran Terbuka (TPT) ( $X_6$ ), kurva efek marginal tampak relatif datar dengan sedikit peningkatan pada rentang nilai menengah. Pola ini mengindikasikan bahwa meskipun TPT signifikan secara statistik dalam model, besaran kontribusinya terhadap perubahan Indeks PMK tidak sebesar variabel lain. Hal ini dapat terjadi karena dimensi kebudayaan dalam Indeks PMK dipengaruhi oleh faktor sosial yang tidak sepenuhnya tercermin dalam variasi tingkat pengangguran, sehingga perubahan TPT tidak serta merta menghasilkan fluktuasi besar pada indeks tersebut.

Kemudian untuk memastikan bahwa model yang diperoleh tidak hanya memiliki performa prediksi yang baik tetapi juga memenuhi asumsi-asumsi statistik, dilakukan analisis diagnostik dan disajikan pada Gambar 3.



Gambar 3. Hasil diagnostik model (Sumber: data diolah)

Gambar 3 menampilkan empat plot yang digunakan untuk mengevaluasi kesesuaian model, dimana Q-Q plot dan histogram residual berguna untuk memeriksa normalitas residual, dimana pada gambar menunjukkan bahwa residual memenuhi asumsi normalitas, karena titik-titik residual pada Q-Q plot tersebar di garis diagonal dan pada histogram distribusi residual mendekati bentuk lonceng simetris. Kemudian plot residual terhadap variabel prediktor linier berguna untuk melihat apakah asumsi homoskedastisitas terpenuhi, dimana pada gambar menunjukkan sebaran titik acak di sekitar garis horizontal nol tanpa pola yang jelas. Sehingga hal ini mengindikasikan bahwa varians konstan dan asumsi homoskedastisitas terpenuhi. Terakhir untuk plot variabel respon terhadap *fitted value* untuk menilai akurasi model secara keseluruhan, dimana titik-titik data berkerumun erat di sekitar garis identitas menunjukkan bahwa nilai yang diprediksi oleh model sangat sesuai dengan nilai aktual dari Indeks PMK.

Dengan demikian, secara keseluruhan, hasil analisis diagnostik ini memberikan keyakinan bahwa model *P-spline* yang terbentuk adalah model yang valid secara statistik, andal dan cocok untuk data yang dianalisis.

#### IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode regresi nonparametrik *Penalized Spline (P-spline)* untuk memodelkan hubungan kompleks antara variabel-varaibel sosial ekonomi dan Indeks Pembangunan Manusia dan Kebudayaan (Indeks PMK) di 34 provinsi di Indonesia pada tahun 2022. Pendekatan ini terbukti efektif, menghasilkan model yang andal dengan daya penjelas yang tinggi, yaitu dengan nilai *Adjusted R<sup>2</sup>* sebesar 0.842 dan *Deviance Explained* 92.3%.

Analisis menunjukkan bahwa empat dari enam variabel prediktor—akses air minum layak, Rasio Gini, pengeluaran per kapita, dan tingkat pengangguran terbuka—memiliki pengaruh yang signifikan secara statistik terhadap IPMK. Temuan yang paling menonjol adalah sifat hubungan yang nonlinier. Hubungan pengeluaran per kapita menunjukkan efek pengembalian yang menurun (*diminishing returns*), sementara *Gini Ratio* menunjukkan hubungan yang kompleks dan non-monotonik. Hal ini menyiratkan bahwa strategi kebijakan pembangunan mungkin perlu disesuaikan berdasarkan tingkat kesejahteraan dan ketimpangan awal suatu daerah.

Validasi model melalui analisis diagnostik mengonfirmasi bahwa asumsi-asumsi statistik kunci telah terpenuhi, sehingga memperkuat validitas kesimpulan yang ditarik. Dengan demikian, penelitian ini menunjukkan bahwa metode *P-spline* tidak hanya mampu menangkap kompleksitas data pembangunan yang seringkali nonlinier, tetapi juga menyediakan alat interpretasi visual yang kuat melalui plot kurva parsial, yang dapat menerjemahkan hasil statistik menjadi wawasan kebijakan yang dapat ditindaklanjuti.

## DAFTAR PUSTAKA

- Andriani, H., & Alfian, M. R. (2020). Aplikasi software R pada analisis angka harapan hidup di Jawa Timur menggunakan estimator P-spline. *Explore*, 10(1), 1. <https://doi.org/10.35200/explore.v10i1.361>
- Badan Pusat Statistik. (2022). *Data Indeks PMK*. <https://www.bps.go.id/id>
- Badan Pusat Statistik. (2024). *Indeks Pembangunan Manusia Indonesia tahun 2024* (No. 85).
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing* (2nd ed.). <https://doi.org/10.1201/9781482273144>
- Iswahyudi, & Asnawi, N. (2024). Menuju kebijakan sosial berorientasi kapabilitas: Telaah pemikiran Amartya Sen dan Martha Nussbaum. *Multikultura*, 3(4), 666–679. <https://doi.org/10.7454/multikultura.v3i4.1071>
- Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan. (2023). *Laporan akhir penyusunan Indeks PMK tahun 2023*.
- Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan. (2025, Juli 7). *Indeks Pembangunan Manusia dan Kebudayaan Kemenko PMK*. Satu Data Indonesia. <https://data.go.id/dataset/dataset/indeks-pembangunan-manusia-dan-kebudayaan-kemenko-pmk>
- Kusunartutik, F., & Dwidayati, N. K. (2022). Pemilihan titik knot optimal menggunakan metode GCV dalam regresi nonparametrik spline truncated. *Indonesian Journal of Mathematics and Natural Sciences*, 45(2), 69–76. <https://doi.org/10.15294/ijmns.v45i2.39727>
- Nguyen, M. (2020). *A guide of analysis*. Bookdown. [https://bookdown.org/mike/data\\_analysis/](https://bookdown.org/mike/data_analysis/)
- Sen, A. (1999). *Development as freedom*. Oxford University Press. <https://global.oup.com/academic/product/development-as-freedom-9780198297581>
- United Nations Development Programme. (1990). *Human development report 1990: Concept and measurement of human development*. <https://hdr.undp.org/content/human-development-report-1990>
- Wahyuningsih, T. D., Handajani, S. S., & Indriati, D. (2019). Penerapan generalized cross validation dalam model regresi smoothing spline pada produksi ubi jalar di Jawa Tengah. *Indonesian Journal of Applied Statistics*, 1(2), 117. <https://doi.org/10.13057/ijas.v1i2.26250>
- Widyowati, C. (2016). *Pemodelan persentasi kemiskinan di Jawa Timur dengan pendekatan regresi nonparametrik aditif berdasarkan estimator penalized spline*.
- Winart. (2020). Pendekatan regresi semiparametrik spline (Pada data nilai Ujian Nasional siswa SMKN 1 Nguling Pasuruan). *Jurnal Sains dan Seni Pomits*, 3(1), 70.