

Perbandingan Regresi Nonparametrik *B-Splines* dan *P-Splines* dalam Memodelkan Hubungan Indeks Pembangunan Manusia dan Persentase Penduduk Miskin di Sumatera Utara

Razita Nur Amalina¹, Dwi Sulistiowati²

^{1,2}Universitas Negeri Padang

¹razitanr@gmail.com

ABSTRAK

Kemiskinan merupakan permasalahan multidimensional yang masih menjadi tantangan utama di Indonesia karena berdampak pada kondisi ekonomi, kesehatan, pendidikan, dan kualitas hidup masyarakat. Meskipun Indeks Pembangunan Manusia (IPM) di Provinsi Sumatera Utara terus mengalami peningkatan, penurunan persentase penduduk miskin tidak selalu menunjukkan pola yang sebanding, sehingga diperlukan pendekatan analisis yang fleksibel untuk memahami hubungan keduanya. Penelitian ini bertujuan membandingkan kinerja regresi nonparametrik *B-Splines* dan *P-Splines* dalam memodelkan hubungan persentase penduduk miskin terhadap IPM pada kabupaten/kota di Sumatera Utara tahun 2024. IPM digunakan sebagai variabel respon, sedangkan persentase penduduk miskin sebagai variabel prediktor. Hasil analisis menunjukkan bahwa model *B-Splines* optimal diperoleh pada derajat 1 dengan satu titik knot dan menghasilkan nilai *Generalized Cross Validation* (GCV) sebesar 7,668706 serta *Adjusted R²* sebesar 0,5842. Sementara itu, model *P-Splines* optimal diperoleh pada orde 1 dengan jumlah knot 7 menghasilkan nilai GCV sebesar 8,005025 dan *Mean Squared Error* (MSE) sebesar 4,771814. Nilai penalti yang besar menunjukkan bahwa hubungan antara persentase penduduk miskin dan IPM cenderung linear menurun. Berdasarkan kriteria kinerja model, regresi *P-Splines* memberikan hasil yang lebih baik dan lebih stabil dibandingkan regresi *B-Splines*. Oleh karena itu, regresi *P-Splines* dapat digunakan sebagai pendekatan yang efektif dalam memodelkan hubungan indikator Pembangunan manusia dan kemiskinan sebagai dasar perumusan kebijakan berbasis data.

Kata Kunci: regresi nonparametrik; B-Spline; P-Spline; Indeks Pembangunan Manusia; kemiskinan.

ABSTRACT

Poverty remains a multidimensional challenge in Indonesia, affecting economic conditions, health, education, and overall quality of life. Although the Human Development Index (HDI) in North Sumatera continues to improve, reductions in poverty rates do not always follow proportionally, indicating the need for flexible modeling approaches to examine their relationship. This study aims to compare the performance of nonparametric B-Splines and P-Splines regression methods in modeling the effect of poverty percentage on HDI across regencies and cities in North Sumatera based on 2024 data. HDI is treated as the response variable, while the percentage of poor population serves as the predictor. The results show that the optimal B-Splines model is obtained with degree 1 and one knot, producing a Generalized Cross Validation (GCV) value of 7.668706 and an Adjusted R² of 0.5842. The optimal P-Splines model is obtained at order 1 with 7 knot and a large smoothing parameter, yielding a lower GCV value of 8.005025 and a Mean Squared Error (MSE) of 4.771814, indicating better predictive performance and model stability. The large smoothing parameter suggests that the relationship between poverty percentage and HDI is predominantly linear and negative. These findings demonstrate that P-Splines regression provides a more accurate and stable model than B-Splines regression. Therefore, penalized nonparametric regression can serve as an effective approach for modeling socio-economic relationships and supporting evidence-based policy formulation.

Keywords: nonparametric regression; B-Splines; P-Spline; Human Development Index; poverty.

PENDAHULUAN

Kemiskinan merupakan permasalahan multidimensional yang masih menjadi tantangan utama dalam pembangunan di berbagai negara berkembang, termasuk Indonesia. Kemiskinan tidak hanya berkaitan dengan keterbatasan pendapatan, tetapi juga berhubungan dengan akses terhadap pendidikan, kesehatan, dan standar hidup layak (Zulfalinda & Damayanti, 2025). Oleh karena itu, upaya pengentasan kemiskinan menjadi bagian penting dalam strategi pembangunan nasional yang berkelanjutan. Salah satu indikator yang digunakan untuk mengukur keberhasilan pembangunan manusia adalah Indeks Pembangunan Manusia (IPM) yang mencerminkan capaian dalam dimensi kesehatan, pendidikan, dan standar hidup layak (Gaol et al., 2024).

Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa IPM di Provinsi Sumatera Utara meningkat dari 71,74 pada tahun 2019 menjadi 75,13 pada tahun 2023. Meskipun demikian, tingkat kemiskinan di beberapa kabupaten/ kota masih relatif tinggi dan tidak selalu mengalami penurunan yang sejalan dengan peningkatan IPM. Kondisi ini mengindikasikan bahwa hubungan antara IPM dan persentase penduduk miskin bersifat kompleks serta berpotensi tidak mengikuti pola linear sederhana.

Pendekatan regresi parametrik memiliki keterbatasan, terutama ketika bentuk hubungan antara variabel tidak diketahui secara eksplisit atau data mengandung outlier yang signifikan. Untuk mengatasi keterbatasan tersebut, regresi nonparametrik menjadi alternatif yang relevan. Regresi nonparametrik tidak mengharuskan spesifikasi bentuk fungsi tertentu, sehingga lebih fleksibel dalam menangkap pola hubungan yang kompleks. Dalam konteks ini, metode *B-Splines* dan *P-Splines* menjadi pilihan yang banyak digunakan karena mampu membentuk kurva regresi yang halus dengan mempertimbangkan kompleksitas data. *B-Splines* dikenal sebagai *piecewise polynomial* yang efektif dalam membangun kurva lokal dengan hubungan terbatas (*local support*), sedangkan *P-Splines* menggabungkan pendekatan *B-Splines* dengan penalti terhadap kelengkungan kurva guna menghindari *overfitting*.

Berbagai penelitian terdahulu telah memanfaatkan regresi nonparametrik dalam memodelkan hubungan antara variabel sosial ekonomi. Misalnya, Syahrul et al., (2024) menemukan bahwa metode regresi spline truncated dapat memberikan kecocokan yang lebih baik dibandingkan regresi kernel dalam menggambarkan hubungan antara rata-rata lama sekolah dan pengeluaran per kapita di Indonesia tahun 2023 berdasarkan kriteria GCV dan MSE, menunjukkan kemampuan pendekatan spline dalam menangkap pola hubungan nonlinear pada data sosial ekonomi kontemporer. Selain itu, Siregar & Jatmiko (2019) yang membandingkan metode regresi nonparametrik *B-Splines* dan *P-Splines* dalam memodelkan hubungan antara IPM dan persentase penduduk miskin di kabupaten/ kota di Indonesia menggunakan data tahun 2017. Hasil penelitian ini menunjukkan bahwa kedua metode menghasilkan kurva regresi yang halus dan dapat menangkap pola hubungan dalam data dengan baik. Namun, *B-Splines* memiliki nilai *Mean Squared Error* (MSE) yang lebih kecil, yaitu sebesar 0,1928378; sehingga dianggap lebih unggul dalam memodelkan hubungan kedua variabel tersebut. Temuan ini memperkuat pentingnya penggunaan metode regresi nonparametrik dalam menganalisis hubungan sosial-ekonomi yang tidak selalu linier dan rawan terhadap pencilan (outlier).

Penelitian ini dilakukan untuk membandingkan performa metode regresi nonparametrik *B-Splines* dan *P-Splines* dalam memodelkan hubungan antara IPM dan persentase penduduk miskin di kabupaten/ kota Provinsi Sumatera Utara tahun 2024. Melalui analisis ini, diharapkan diperoleh informasi yang lebih akurat mengenai bentuk hubungan kedua variabel serta metode pemodelan yang paling sesuai berdasarkan kriteria perbandingan nilai MSE dan *Adjusted R²*. Hasil dari penelitian ini diharapkan dapat

memberikan kontribusi metodologis dalam penerapan regresi nonparametrik pada data sosial-ekonomi serta memberikan masukan kebijakan berbasis data dalam upaya penanggulangan kemiskinan daerah.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Penelitian terapan ini menggunakan data sekunder dari publikasi BPS tahun 2024 untuk menganalisis hubungan antara Persentase Penduduk Miskin (variabel prediktor) dan Indeks Pembangunan Manusia (variabel respon). Penjelasan mengenai variabel penelitian dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala Pengukuran
Y	Indeks Pembangunan Manusia (IPM)	Rasio
X	Persentase Penduduk Miskin	Rasio

Regresi Nonparametrik

Metode analisis yang digunakan dalam penelitian ini adalah regresi nonparametrik *B-Splines* dan *P-Splines*. Regresi nonparametrik merupakan pendekatan statistika yang berfungsi untuk memodelkan hubungan antara variabel prediktor dan respon tanpa bergantung pada asumsi bentuk fungsi tertentu. Secara umum model regresi nonparametrik dapat ditulis sebagai berikut.

$$y_i = f(x_i) + e_i, \quad i = 1, 2, \dots, n$$

Regresi nonparametrik digunakan jika pola hubungan antara variabel respons dan variabel prediktor tidak diketahui dalam bentuk kurva regresinya. Dengan kata lain, bentuk fungsi $m(x_i)$ dan jumlah parameter yang harus diestimasi dalam model nonparametrik tidak diketahui sejak awal.

Regresi B-Spline

Regresi *B-Splines* merupakan pendekatan dalam regresi nonparametrik yang digunakan untuk mengestimasi bentuk kurva tanpa asumsi spesifik terhadap struktur fungsional hubungan antar variabel. Menurut Eubank, (1999), *B-spline* merupakan fungsi polinomial potongan (*piecewise polynomial*) dengan dukungan lokal (*local support*) untuk derajat polinomial tertentu. *B-spline* ke- j dengan derajat v , yang dibangun berdasarkan simpul-simpul (*knots*) sebanyak u titik simpul, memerlukan tambahan simpul sebanyak $2v$, sehingga diperoleh barisan simpul $T(t_1, \dots, t_v, t_{v+1}, \dots, t_{u+v}, t_{u+v+1}, \dots, t_{u+2v})$ dimana $t_1 = \dots = t_v < t_{v+1} < \dots < t_{u+v} < t_{u+v+1} = \dots = t_{u+2v}$. Sehingga *B-spline* ke- j untuk $j = 1, \dots, v + u$ didefinisikan secara rekursif dengan rumus:

$$B_j(x; v) = \frac{x - t_j}{t_{j+v-1} - t_j} B_j(x; v - 1) + \left(1 - \frac{x - t_{j+1}}{t_{j+v} - t_{j+1}} B_{j+1}(x; v - 1) \right)$$

dimana

$$B_j(x; v) = \begin{cases} 1, & \text{jika } t_j \leq x \leq t_{j+1} \\ 0, & \text{lainnya} \end{cases}$$

B-spline yang dinormalisasikan $\forall x: \sum_{j=1}^{v+u} B_j(x; v) = 1$.

Dalam regresi, rata-rata $y_i = f(x_i) + e_i$ untuk $i = 1, \dots, n$, diasumsikan bahwa fungsi $f(x_i)$ adalah fungsi yang halus, dan dapat didekati dengan kombinasi linear dari basis *B-spline*:

$$f(x) \approx \sum_{j=1}^m \alpha_j B_j(x; v)$$

Dimana $\alpha = (\alpha_1, \dots, \alpha_j)$ adalah vektor koefisien dari basis *B-spline* $B_j(\cdot; v)$ dengan derajat bebas v dan $u + \text{lequidistant}$ knot untuk $j = 1, \dots, v + u = m$. Fungsi objektif dari regresi *B-spline* ditulis sebagai

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \left\{ \sum_{i=1}^n \left(Y_i - \sum_{j=1}^m \alpha_j B_j(x_i; v) \right)^2 \right\}$$

Sehingga model regresi menjadi

$$y_i = \sum_{j=1}^m \alpha_j B_j(x_i; v), \quad i = 1, 2, \dots, n$$

Kualitas model regresi B-Spline sangat ditentukan oleh ketepatan pemilihan titik knot optimal. Dalam membangun model ini, dua parameter krusial yang harus ditetapkan adalah orde regresi dan jumlah knot. Sebagai titik transisi tempat data berubah, penentuan jumlah knot menuntut keseimbangan, dimana jumlah yang berlebihan akan mengakibatkan model menjadi kaku dan *overfit*, sementara jumlah yang terlalu minim akan membuat kurva terlalu landai (*oversmooth*) sehingga gagal menangkap pola distribusi data yang sebenarnya.

Menurut Goepf et al., (2025), pemodelan *B-spline* dilakukan dengan mencoba berbagai konfigurasi knot, lalu memilih model yang meminimalkan nilai RSS. Nilai RSS mencerminkan seberapa baik kurva spline mendekati data observasi. Namun, memilih jumlah knot yang terlalu banyak dapat menurunkan RSS secara drastis tetapi menyebabkan *overfitting*. Sebaliknya, terlalu sedikit knot dapat menyebabkan *underfitting*.

Regresi P-Spline

P-Splines atau *Penalized Splines* merupakan salah satu pendekatan dalam regresi spline yang memuat fungsi penalti guna mengontrol tingkat kemulusan kurva. Dalam regresi P-Splines digunakan *Penalized Least Square* (PLS) sebagai metode optimasi, yaitu suatu pendekatan yang memodifikasi metode *least square* dengan menambahkan komponen penalti. PLS berfungsi sebagai kriteria optimasi yang menyeimbangkan antara kecocokan terhadap data dan kemulusan kurva. Tujuan utamanya adalah menghasilkan estimasi fungsi yang tidak hanya sesuai dengan data, tetapi juga memiliki varian galat yang rendah. Fungsi objektif dari *P-Spline* dapat dirumuskan sebagaimana ditunjukkan dalam persamaan berikut (Siregar & Jatmiko, 2019).

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^m \alpha_j \beta_j(x_i; v) \right)^2 + \lambda \sum_{j=d+1}^m (\Delta^d \alpha_j)^2$$

dimana, $\lambda > 0$ merupakan parameter *smoothing* dan Δ^d merupakan *difference operator*,

$$\Delta^d \alpha_1 = \sum_{t=0}^d (-1)^t \alpha_{(1-t)} \quad d \in \mathbb{N}$$

Teknik Analisis Data

Estimasi kurva regresi dalam penelitian ini diolah menggunakan perangkat lunak *RStudio*. Adapun Langkah-langkah analisis yang dilakukan adalah sebagai berikut.

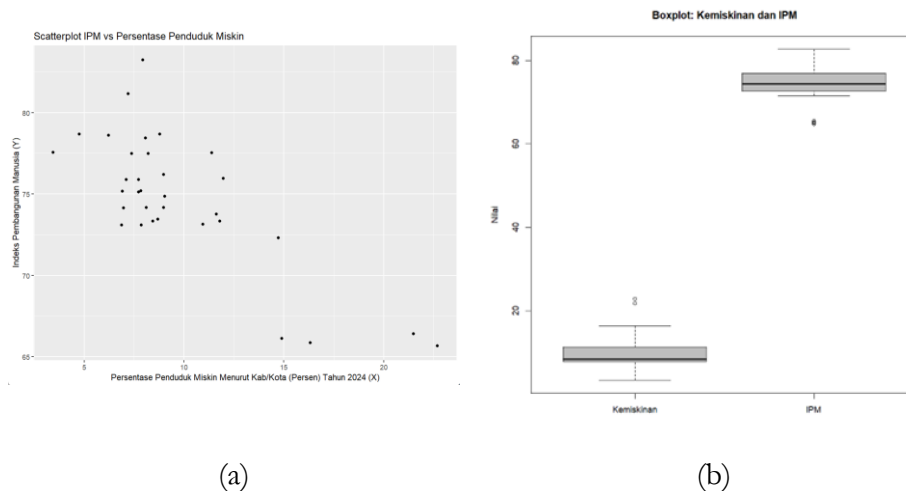
1. Melakukan eksplorasi data melalui visualisasi *scatter plot* dan *boxplot* antara variabel prediktor (IPM) dan variabel respon (Persentase Penduduk Miskin) untuk mengidentifikasi pola dan *outlier* dalam data.
2. Melakukan pemodelan dan membuat kurva regresi B-Splines sesuai dengan persamaan regresi. Parameter α diestimasi dengan metode *least square*, dengan kriteria pemilihan model terbaik berdasarkan kriteria nilai RSS dan GCV minimum.
3. Melakukan pemodelan regresi menggunakan basis spline dengan penalti atau p-spline. Estimasi dilakukan dengan mempertimbangkan nilai parameter *smoothing* λ untuk menghindari *overfitting* dan *underfitting*.

4. Melakukan perbandingan performa antara model regresi *B-Splines* dan *P-Splines* dengan membandingkan kurva hasil estimasi serta nilai evaluasi model seperti RSS, GCV, dan MSE.
5. Menetapkan model terbaik untuk mengestimasi fungsi regresi antara variabel IPM dan persentase penduduk miskin berdasarkan kriteria nilai MSE terkecil.

HASIL DAN PEMBAHASAN

Eksplorasi Awal

Eksplorasi data ini dilakukan untuk memahami keterkaitan antara tingkat kemiskinan dan IPM di tingkat kabupaten/ kota se-Provinsi Sumatera Utara. Fokus utama eksplorasi ini mencakup pemetaan pola sebaran, karakteristik distribusi, dan pendeteksi data pencilan (*outlier*). Hasil eksplorasi ini digunakan untuk menjadi acuan yang mendukung analisis lanjutan yang lebih presisi serta memperkuat landasan perumusan kebijakan berbasis data.



Gambar 1. (a) Scatter Plot dan, (b) Boxplot Persentase Penduduk Miskin terhadap IPM Kabupaten/Kota Sumatera Utara

Gambar 1 (a) menunjukkan hubungan negatif antara Persentase Penduduk Miskin dan IPM di kabupaten/kota Provinsi Sumatera Utara. Titik-titik data cenderung membentuk pola menurun, yang berarti semakin tinggi Persentase Penduduk Miskin, cenderung semakin rendah nilai IPM. Mayoritas kabupaten/kota berada dalam rentang kemiskinan 5–15% dan IPM 73–78, sementara beberapa daerah tampak sebagai *outlier* dengan Persentase Penduduk Miskin di atas 20% dan IPM di bawah 70. Hal ini mengindikasikan bahwa Persentase Penduduk Miskin yang tinggi masih menjadi penghambat utama bagi peningkatan kualitas pembangunan manusia di daerah tersebut.

Gambar 1 (b) memperlihatkan perbandingan distribusi Persentase Penduduk Miskin dan IPM di kabupaten/kota Provinsi Sumatera Utara. Nilai Persentase Penduduk Miskin memiliki rentang yang lebih lebar dengan beberapa outlier, menunjukkan variasi yang cukup besar antar kabupaten/kota. Sebaliknya, nilai IPM cenderung terkonsentrasi pada kisaran tinggi (sekitar 75) dengan sedikit penyimpangan, menandakan bahwa sebagian besar daerah memiliki capaian pembangunan manusia yang relatif baik. Namun demikian, keberadaan outlier pada kedua variabel menunjukkan bahwa masih terdapat kabupaten/kota yang secara signifikan tertinggal dan perlu mendapatkan perhatian khusus dalam perumusan kebijakan Pembangunan.

Hasil Pemodelan Regresi *B-Spline*

Ringkasan nilai GCV yang diperoleh dari pemodelan regresi *B-Splines* dengan variasi orde serta jumlah titik knot dipaparkan dalam Tabel 2.

Tabel 2. Nilai GCV dan RSS Berdasarkan Variasi Orde dan Titik Knot

Degree/Orde	Jumlah Knot	Knot Optimal	GCV	RSS
1	1	8,21	7,668706	209,1465
	2	7,80333; 9	8,213325	209,3153
	3	7,37; 8,21; 11,39	8,767222	208,2879
2	1	8,21	8,238697	209,9619
	2	7,80333	8,683312	206,2944
	3	7,37; 8,21; 11,39	8,800959	194,4212
3	1	8,21	8,149612	193,6150
	2	7,80333; 9	8,462799	186,9509
	3	7,37; 8,21; 11,39	9,047581	185,3383

Tabel tersebut menyajikan kombinasi nilai GCV dan RSS untuk penggunaan tiga titik knot pada model orde 1, 2, dan 3. Berdasarkan hasil perbandingan, model dengan orde 1 menunjukkan nilai GCV yang paling minimum. Berdasarkan hal tersebut, estimasi parameter dilanjutkan pada model orde 1, yang hasilnya disajikan pada Tabel 3.

Tabel 3. Hasil Estimasi Parameter Model *B-Spline*

Orde Model	Parameter	Koefisien	t-Statistik	<i>P-Value</i>	F-Statistik	Adj R ²
1	α_0	77,960	40,536	<2e-16	23,48	0,5842
	α_1	-2,765	-0,801	0,43		
	α_2	-13,521	-5,608	4,19e-06		

Berdasarkan hasil estimasi parameter model *B-Spline* orde 1, diketahui bahwa model tersebut signifikan secara simultan pada tingkat signifikansi 5% dengan nilai F-statistik sebesar 23,48. Secara parsial, parameter α_0 dan α_2 terbukti signifikan dengan p-value masing-masing kurang dari 0,05, sedangkan parameter α_1 tidak signifikan (p-value = 0,43). Model ini memiliki nilai *Adjusted R²* yakni 0,5842, yang berarti sekitar 58,42% variabilitas data dapat dijelaskan oleh model, dengan demikian dapat disimpulkan bahwa model tersebut memiliki kesesuaian yang cukup baik.

Hasil estimasi menunjukkan bahwa model *B-Spline* orde 1 mampu menjelaskan sekitar 58,42% variasi IPM. Nilai koefisien basis spline yang signifikan mengindikasikan bahwa perubahan persentase penduduk miskin memiliki pengaruh nyata terhadap variasi IPM pada tingkat kabupaten/ kota. Namun demikian, terdapat satu parameter yang tidak signifikan secara parsial, yang menunjukkan bahwa bentuk hubungan yang ditangkap oleh model cenderung sederhana dan tidak menunjukkan perubahan pola yang kompleks pada setiap interval knot.

Temuan ini sejalan dengan karakteristik *B-Spline* yang sangat bergantung pada jumlah dan posisi titik knot. Dengan hanya satu titik knot optimal, model menunjukkan bahwa variasi lokal hubungan antara kemiskinan dan IPM relatif terbatas. Hal ini mengindikasikan bahwa pada data tahun 2024, pola hubungan antara kedua variabel tidak menunjukkan perubahan struktur yang tajam pada rentang nilai tertentu. Secara substantif, kondisi ini dapat mencerminkan bahwa peningkatan IPM di sebagian besar kabupaten/ kota terjadi secara relatif konstan tanpa perbedaan dinamika pembangunan yang ekstrem antar wilayah.

Hasil Pemodelan Regresi *P-Spline*

Hasil perhitungan nilai GCV untuk berbagai kombinasi orde dan jumlah knot pada model *P-Spline* disajikan dalam Tabel 4 berikut ini.

Tabel 4. Nilai GCV Berbagai Orde dan Jumlah Knot

Degree/Orde	Jumlah Knot	GCV
1	5	8,841579
	6	8,222094
	7	8,005025
	8	8,106750
	9	8,217535
2	10	8,296846
	5	8,493983
	6	8,414920
	7	8,338945
	8	8,457128
3	9	8,450753
	10	8,277577
	5	8,282212
	6	8,341979
	7	8,402829
	8	8,474542
	9	8,494035
	10	8,439731

Tabel 4 menunjukkan bahwa kombinasi orde 1 dengan knot 7 memberikan hasil terbaik dengan nilai GCV terkecil yakni 8,005025. Selanjutnya dilakukan estimasi parameter untuk model orde 1 knot 7. Hasil estimasi parameter orde 1 disajikan pada Tabel 5 berikut ini.

Tabel 5. Hasil Estimasi Parameter Model *P-Spline*

Orde Model	Parameter	Koefisien	t-Statistik	<i>P-Value</i>	F-Statistik	Adj R ²
1	α_0	9,7303	24,32	<2e-16	26,54	0,703

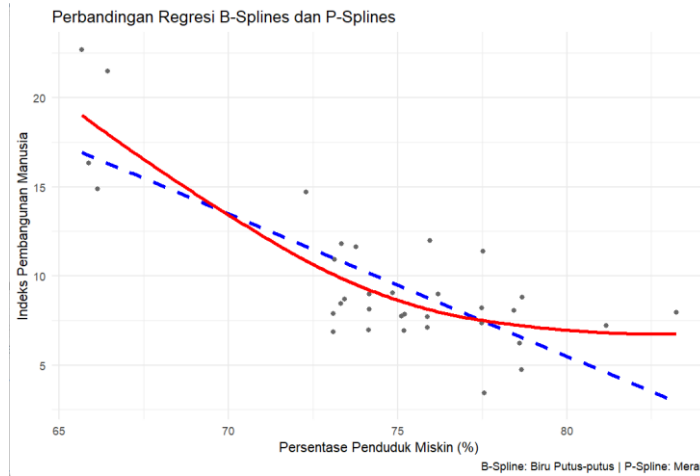
Berdasarkan Tabel 5, model *P-Spline* orde 1 menunjukkan parameter α_0 memiliki koefisien senilai 9,7303 dengan nilai t-statistik 24,32 dan p-value <2e-16, yang menunjukkan signifikansi statistik yang sangat kuat. Nilai *Adjusted R*² sebesar 0,703 mengindikasikan bahwa model dapat menjelaskan sekitar 70,3% variasi dalam data, menunjukkan kecocokan model yang cukup baik.

Model *P-Spline* menghasilkan nilai *Adjusted R*² yang lebih tinggi dibandingkan *B-Spline*, yang menunjukkan kemampuan model dalam menangkap variasi data yang lebih besar. Penggunaan komponen penalti dalam *P-Spline* memungkinkan kurva regresi menjadi lebih halus dan stabil, sehingga mampu meminimalkan fluktuasi lokal yang tidak esensial.

Nilai *smoothing* parameter yang besar mengindikasikan bahwa hubungan antara persentase penduduk miskin dan IPM cenderung mendekati pola linear menurun. Dengan kata lain, peningkatan kemiskinan secara konsisten diikuti oleh penurunan kualitas pembangunan manusia tanpa adanya perubahan pola hubungan yang drastis pada rentang tertentu. Temuan ini memperkuat indikasi bahwa kemiskinan merupakan faktor struktural yang berpengaruh langsung terhadap capaian manusia di daerah.

Perbandingan Model Regresi *B-Spline* dan *P-Spline*

Gambar 2 menyajikan perbandingan visual antara kurva hasil estimasi model regresi *B-Spline* dan *P-Spline*.



Gambar 2. Perbandingan Plot Regresi *B-Spline* dan *P-Spline*

Gambar tersebut menunjukkan perbandingan antara regresi *B-Spline* (garis berwarna biru) dan *P-Spline* (garis berwarna merah putus-putus) dalam memodelkan hubungan antara Persentase Penduduk Miskin dan IPM. Kedua model menunjukkan tren negatif, yaitu semakin tinggi persentase penduduk miskin, semakin rendah IPM. Namun, *P-Spline* terlihat lebih halus dan stabil dibandingkan *B-Spline*, yang cenderung memiliki fluktuasi lokal.

Hubungan negatif yang stabil antara persentase penduduk miskin dan IPM mencerminkan bahwa disparitas Pembangunan antar kabupaten/ kota di Provinsi Sumatera Utara masih dipengaruhi faktor struktural seperti akses Pendidikan, kualitas layanan kesehatan, dan kesempatan kerja formal. Kabupaten/ kota dengan tingkat kemiskinan tinggi cenderung memiliki keterbatasan akses terhadap layanan dasar, sehingga peningkatan kualitas hidup Masyarakat berlangsung lebih lambat. Kondisi ini menunjukkan bahwa peningkatan indikator Pembangunan manusia belum merata secara spasial, sehingga kebijakan pengentasan kemiskinan perlu difokuskan pada wilayah dengan tingkat kerentanan sosial ekonomi yang lebih tinggi.

Evaluasi model lebih lanjut dengan menggunakan perbandingan nilai MSE yang dapat membantu menentukan model yang paling sesuai secara statistik.

Tabel 6. Perbandingan Nilai MSE

MSE B-Spline	MSE P-Spline
6,880807	4,771814

Berdasarkan Tabel 6, model regresi *B-Spline* memiliki nilai MSE sebesar 6,880807 sedangkan *P-Spline* memiliki nilai MSE sebesar 4,771814. Karena MSE yang lebih rendah menunjukkan kesalahan prediksi yang lebih kecil, maka secara kuantitatif *P-Spline* dianggap lebih baik dalam hal akurasi prediksi pada data ini.

Perbedaan performa antara *B-Spline* dan *P-Spline* menunjukkan bahwa pendekatan dengan penalti memberikan estimasi yang lebih stabil pada data sosial ekonomi regional. Keunggulan *P-Spline* dalam penelitian ini konsisten dengan temuan penelitian kontemporer yang menunjukkan bahwa *penalized smoothing* mampu mengurangi risiko *overfitting* dan meningkatkan akurasi prediksi pada data dengan variasi antar wilayah yang tinggi (McTeer et al., 2024).

Secara substansif, hasil ini menunjukkan bahwa hubungan antara kemiskinan dan pembangunan manusia di Sumatera Utara lebih tepat dipahami sebagai hubungan struktural

yang stabil, bukan hubungan lokal yang berubah-ubah pada interval tertentu. Oleh karena itu, pendekatan pemodelan yang menekankan kelulusan kurva lebih sesuai dalam menggambarkan dinamika sosial ekonomi daerah.

Implikasi kebijakan dari temuan ini adalah bahwa penurunan kemiskinan berpotensi memberikan dampak langsung dan konsisten terhadap peningkatan IPM. Dengan demikian, strategi pembangunan yang berfokus pada pengurangan kemiskinan struktural, seperti peningkatan akses pendidikan, layanan kesehatan, dan kesempatan kerja, memiliki relevansi empiris yang kuat berdasarkan hasil pemodelan ini.

Kelebihan dan Kekurangan Metode

Metode regresi *B-Spline* merupakan pendekatan nonparametrik yang fleksibel dalam memodelkan hubungan nonlinier melalui pengaturan jumlah titik knot, dengan keunggulan pada proses estimasi yang relatif sederhana. Namun, metode ini rentan terhadap *overfitting* apabila jumlah knot terlalu banyak serta memerlukan pemilihan posisi knot yang cenderung subjektif dan dapat memengaruhi kestabilan model, terutama pada bagian ujung data.

Untuk mengatasi keterbatasan tersebut, *P-Spline* menambahkan komponen penalti terhadap kelengkungan kurva sehingga kompleksitas model dapat dikontrol melalui parameter *smoothing*, menghasilkan estimasi yang lebih halus, stabil, dan umumnya memiliki kesalahan prediksi yang lebih kecil, sebagaimana ditunjukkan oleh nilai MSE yang lebih rendah dalam penelitian ini.

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam interpretasi hasil. Model yang dibangun hanya menggunakan satu variabel prediktor, yaitu persentase penduduk miskin, sehingga variasi IPM yang dipengaruhi oleh faktor lain seperti tingkat pendidikan, pengeluaran per kapita, dan akses layanan kesehatan belum sepenuhnya terakomodasi. Selain itu, analisis menggunakan data potong lintang tahun 2024 sehingga belum menggambarkan dinamika hubungan dalam jangka waktu yang lebih panjang. Oleh karena itu, penelitian selanjutnya disarankan menggunakan pendekatan multivariat dan data panel untuk memperoleh gambaran hubungan yang lebih komprehensif.

PENUTUP

Berdasarkan hasil analisis regresi nonparametrik terhadap data IPM dan persentase penduduk miskin kab/ kota di Provinsi Sumatera Utara, diperoleh kesimpulan sebagai berikut.

1. Model regresi *B-Spline* optimal diperoleh pada orde 1 dengan satu titik knot dan menghasilkan nilai GCV sebesar 7,668706 serta Adjusted R^2 sebesar 0,5842. Model ini signifikan secara simultan dan mampu menjelaskan sekitar 58,42% variasi data, sehingga memiliki tingkat kecocokan yang cukup baik.
2. Model regresi *P-Spline* optimal diperoleh pada orde 1 dengan jumlah knot 7 dan menghasilkan nilai *Adjusted R²* sebesar 0,703 serta nilai MSE sebesar 4,771814, yang lebih kecil dibandingkan model *B-Spline* dengan MSE sebesar 6,880807. Ini membuktikan bahwa model *P-Spline* lebih akurat dan stabil dibandingkan model lainnya.
3. Berdasarkan perbandingan kriteria evaluasi model, regresi *P-Spline* memberikan performa yang lebih unggul dibandingkan regresi *B-Spline* dalam memodelkan hubungan antara IPM dan persentase penduduk miskin di Sumatera Utara tahun 2024. Hasil ini menunjukkan bahwa pendekatan *penalized smoothing* mampu menyeimbangkan fleksibilitas model dan stabilitas estimasi, sehingga kurva regresi tidak hanya mengikuti pola data tetapi juga menghindari fluktuasi lokal yang tidak merepresentasikan struktur hubungan sebenarnya.

Untuk memperoleh pemahaman yang lebih mendalam, disarankan bagi peneliti selanjutnya untuk mengembangkan model regresi multivariat. Dengan menambahkan variabel prediktor lain seperti tingkat pengangguran, pertumbuhan ekonomi, maupun rata-rata lama sekolah. Diharapkan nantinya model yang dihasilkan dapat memberikan gambaran yang lebih komprehensif mengenai berbagai determinan yang memengaruhi IPM. Dengan demikian, temuan utama penelitian ini adalah bahwa hubungan antara kemiskinan dan pembangunan manusia di Provinsi Sumatera Utara bersifat negatif dan relatif stabil secara struktural, serta lebih tepat dimodelkan menggunakan pendekatan smoothing dengan penalti dibandingkan pendekatan spline bebas.

REFERENSI

- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing* (2nd ed.). Marcel Dekker.
- Gaol, R. I. L., Tampubolon, A., Nasution, M., Nasution, M. D., & Sianturi, A. F. (2024). Analisis Pengaruh Pengangguran dan Indeks Pembangunan Manusia (IPM) Terhadap Jumlah Penduduk Miskin di Sumatera Utara Tahun 2012-2023. *MANTAP: Journal of Management Accounting, Tax and Production*, 2(2), 349–358. <https://doi.org/10.57235/mantap.v2i2.2956>
- Goepp, V., Bouaziz, O., & Nuel, G. (2025). Spline regression with automatic knot selection. *Computational Statistics and Data Analysis*, 202. <https://doi.org/10.1016/j.csda.2024.108043>
- McTeer, M., Henderson, R., Anstee, Q. M., & Missier, P. (2024). Handling Overlapping Asymmetric Data Sets—A Twice Penalized P-Spline Approach. *Mathematics*, 12(5), 1–33. <https://doi.org/10.3390/math12050777>
- Siregar, R. S. K., & Jatmiko, Y. A. (2019). Perbandingan Regresi B-Splines dan P-Splines pada Hubungan Indeks Pembangunan Manusia dan Persentase Penduduk Miskin Kabupaten / Kota di Indonesia. *Emasains: Jurnal Edukasi ...*, VIII, 101–111. <https://ojs.mahadewa.ac.id/index.php/emasains/article/view/277>
- Syahrlul, M., Amanda, H. S., Fahrani, I. R., Yasmin, Y., Purnamasari, N. A., & Baskara, Z. W. (2024). Perbandingan Regresi Nonparametrik Kernel dan Spline pada Pemodelan Hubungan antara Rata-Rata Lama Sekolah dan Pengeluaran per Kapita di Indonesia. *Indonesian Journal of Applied Statistics and Data Science*, 1(1), 11–19. <https://doi.org/10.29303/ijasds.v1i1.5725>
- Zulfalinda, A., & Damayanti, A. (2025). Analysis of the Effect of Open Unemployment Rate, Poverty Rate, and Economic Growth on the Human Development Index (HDI). *Journal of Business Management and Economic Development*, 3(03), 1078–1091. <https://doi.org/10.59653/jbmed.v3i03.1987>