

PERBANDINGAN REGRESI NONPARAMETRIK *SPLINE TRUNCATED* DAN *KERNEL GAUSSIAN* DALAM MENGANALISIS FAKTOR-FAKTOR PENENTU INDEKS PEMBANGUNAN MANUSIA (IPM) DI INDONESIA

Uci Nopita Safitri¹, Idhia Sriliana^{2*}, Regina Adelisa³, Muhammad Hafiz⁴, Pepi Novianti⁵

^{1,2,3,4,5}Program Studi S-1 Statistika, Universitas Bengkulu, Indonesia

*e-mail: idhiasriliana@unib.ac.id

DOI: 10.14710/j.gauss.14.2.554-564

Article Info:

Received: 2025-06-01

Accepted: 2025-12-05

Available Online: 2025-12-10

Keywords:

IPM; Regresi; Nonparametrik;
Spline truncated; *Kernel Gaussian*

Abstract: The Human Development Index (HDI) is an important indicator for measuring the quality of development in a region. This study compares two nonparametric regression approaches, namely truncated spline regression and Gaussian kernel regression, in analyzing the factors influencing HDI in Indonesia in 2024. The independent variables used include Expected Years of Schooling (HLS), Mean Years of Schooling (RRLS), and the percentage of the poor population (PPM). Nonparametric regression is chosen for its ability to capture complex relationships between variables without strict linearity assumptions. The results show that both methods effectively model the relationship between the variables and HDI. Truncated spline regression performs better in detecting structural changes, while kernel regression is more flexible in capturing smooth relationships. Model evaluation using the coefficient of determination (R^2) and mean squared error (MSE) indicates that truncated spline yields an R^2 of 92.79% and an MSE of 1.8617, while Gaussian kernel regression results in an R^2 of 82.25% and an MSE of 3.6837. Therefore, truncated spline regression proves to be more accurate in modeling the relationship between determining factors and HDI, and it can serve as a more suitable alternative for analyzing complex and nonlinear patterns in human development policy research.

1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) merupakan indikator penting yang digunakan untuk menilai kualitas pembangunan manusia melalui tiga dimensi utama, yaitu kesehatan, pendidikan, dan standar hidup. Badan Pusat Statistik (2024) menghitung IPM berdasarkan tiga komponen, yaitu umur harapan hidup sebagai dimensi kesehatan, harapan lama sekolah dan rata-rata lama sekolah sebagai dimensi pendidikan, serta pengeluaran riil per kapita yang disesuaikan sebagai dimensi standar hidup. IPM menjadi ukuran strategis dalam menilai keberhasilan pembangunan baik di tingkat nasional maupun daerah, sehingga pemantauannya sangat penting untuk melihat sejauh mana kualitas hidup masyarakat mengalami peningkatan.

Setiap komponen pembentuk IPM memiliki peran signifikan dalam menggambarkan kualitas hidup. Harapan lama sekolah mencerminkan ekspektasi lama pendidikan generasi muda, sedangkan rata-rata lama sekolah menunjukkan tingkat pendidikan penduduk dewasa yang menggambarkan kualitas sumber daya manusia saat ini. Selain itu, variabel sosial ekonomi seperti kemiskinan juga sering digunakan dalam analisis pembangunan manusia karena memiliki korelasi erat dengan kemampuan pemenuhan kebutuhan dasar dan peningkatan kesejahteraan.

IPM memiliki keterkaitan dengan pencapaian *Sustainable Development Goals* (SDGs), atau Tujuan Pembangunan Berkelanjutan. Dimensi kesehatan, pendidikan, dan kesejahteraan ekonomi yang diukur dalam IPM secara langsung berhubungan dengan SDG

1 (pengentasan kemiskinan), SDG 3 (kesehatan dan kesejahteraan), SDG 4 (pendidikan berkualitas), SDG 8 (pertumbuhan ekonomi inklusif), dan SDG 10 (pengurangan ketimpangan). SDGs sendiri merupakan agenda pembangunan global yang ditetapkan oleh Perserikatan Bangsa-Bangsa untuk mencapai pembangunan berkelanjutan hingga tahun 2030 (United Nations, 2015). Oleh karena itu, IPM tidak hanya berfungsi sebagai alat evaluasi pembangunan, tetapi juga sebagai indikator komprehensif dalam menilai kemajuan pencapaian agenda pembangunan global.

Dalam menganalisis faktor-faktor yang memengaruhi IPM, metode statistik diperlukan untuk memperoleh pemahaman yang lebih akurat mengenai pola hubungan antarvariabel. Regresi nonparametrik menjadi salah satu pendekatan yang banyak digunakan karena bersifat fleksibel dan tidak mensyaratkan bentuk fungsi tertentu antara variabel respon dan variabel prediktor. Dua metode yang umum digunakan adalah regresi *spline truncated* dan regresi kernel. Regresi *spline truncated* membagi domain data ke dalam segmen melalui titik simpul (*knots*) sehingga mampu menangkap perubahan pola hubungan secara lokal. Sementara itu, regresi kernel menghasilkan kurva regresi yang halus dan kontinu melalui mekanisme pembobotan menggunakan fungsi kernel.

Berbagai penelitian sebelumnya telah menunjukkan efektivitas regresi nonparametrik dalam pemodelan IPM. Anwar, Rassiyanti dan Pitri (2025) menemukan bahwa regresi kernel Gaussian mampu menangkap pola nonlinier antara variabel sosial-ekonomi dan IPM dengan akurasi tinggi. Dewi dan Ahadiyah (2023) juga membuktikan bahwa konfigurasi tiga *knot* memberikan hasil terbaik dalam pemodelan IPM Jawa Timur. Selain itu, Sudiasa dkk. (2023) menerapkan *multivariable spline truncated* pada pemodelan IPM dan menemukan bahwa metode ini mampu menangani hubungan prediktor yang kompleks. Secara keseluruhan, penelitian-penelitian tersebut menegaskan bahwa regresi nonparametrik merupakan pendekatan yang fleksibel dan akurat untuk memodelkan IPM dan indikator pembangunan terkait.

Meskipun beberapa penelitian telah membahas penggunaan regresi *spline truncated* dan regresi kernel dalam pemodelan pembangunan manusia, kajian yang secara langsung membandingkan performa kedua metode tersebut dalam konteks IPM Indonesia masih terbatas. Minimnya penelitian komparatif ini membentuk celah penelitian (*research gap*) yang perlu diisi melalui analisis yang lebih komprehensif. Oleh karena itu, penelitian ini bertujuan untuk menganalisis performa regresi nonparametrik *spline truncated* dalam memodelkan faktor-faktor yang memengaruhi IPM, menganalisis performa regresi nonparametrik kernel dalam memodelkan faktor-faktor yang sama, serta membandingkan keakuratan dan efektivitas kedua metode tersebut dalam memodelkan faktor-faktor penentu IPM di Indonesia.

2. TINJAUAN PUSTAKA

Analisis regresi dalam statistika adalah salah satu metode untuk menentukan hubungan sebab akibat antara satu variabel dengan variabel yang lain. Secara umum model regresi didefinisikan pada persamaan 1 berikut:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (1)$$

dimana Y dinamakan variabel respon, variabel X dinamakan variabel prediktor, β_0 dan β_1 adalah parameter-parameter model yang nilainya tidak diketahui, serta ε (epsilon) adalah galat acak (*random error*) (Alwi & Musfirah, 2021).

Regresi nonparametrik merupakan metode analisis yang tidak mengasumsikan bentuk hubungan tertentu antara variabel bebas dan variabel terikat (Härdle, 1990). Pendekatan ini bersifat fleksibel karena pola hubungan dibentuk langsung berdasarkan data, sehingga mampu menggambarkan relasi yang bersifat nonlinear maupun kompleks tanpa harus

menentukan bentuk fungsi secara eksplisit (Ruppert *et al.*, 2003). Beberapa tipe regresi nonparametrik yang umum digunakan antara lain Kernel, Spline dan Nearest Neighbor (Wahba, 1990).

Diberikan data berpasangan dengan empat variabel prediktor sehingga dapat dituliskan $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, y_i)$. Selanjutnya diasumsikan pola antara masing-masing variabel prediktor $(x_{1i}, x_{2i}, x_{3i}, x_{4i})$ terhadap variabel respon (y_i) mengikuti model regresi *nonparametrik*. Kurva regresi yang dibentuk dari variabel prediktor diasumsikan bersifat aditif, sehingga (Asti & Darwis, 2023):

$$f(x_{1i}, x_{2i}, x_{3i}, x_{4i}) = f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_4(x_{4i}) = \sum_{p=1}^4 f_p(x_{pi}) \quad (2)$$

Spline truncated merupakan salah satu bentuk regresi nonparametrik yang menggunakan fungsi potongan (*piecewise polynomial*) untuk memodelkan hubungan antara variabel respon dan variabel prediktor (Dai & Zhang, 2013). Metode ini membagi rentang data menjadi beberapa segmen dengan titik batas yang disebut knot, kemudian membentuk fungsi polinomial pada setiap segmen.

Regresi nonparametrik dikenal memiliki tingkat fleksibilitas yang tinggi karena memungkinkan data untuk menentukan sendiri bentuk kurva estimasi regresinya tanpa adanya pengaruh subjektivitas dari peneliti. Salah satu metode estimasi yang digunakan dalam regresi nonparametrik adalah metode spline (De Boor, 2001). Spline merupakan potongan polinomial (*piecewise polynomial*) tersegmentasi yang memiliki sifat fleksibilitas. Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda disebut titik knot (Draper & Smith, 1998). Adapun model dari regresi spline adalah sebagaimana persamaan (1).

$$f(x_i) = \sum_{j=0}^m \beta_j x_i^j + \beta_{(m+1)} \sum_{l=1}^p (x_i - k_l)^m \quad (3)$$

$$\text{dengan } (x_i - k_l)^m = \begin{cases} (x_i - k_l)^m, & x_i \geq k_l \\ 0, & x_i < k_l \end{cases}$$

Estimasi pada model regresi nonparametrik *spline truncated* dilakukan dengan membangun fungsi pemodelan berbentuk polinomial tersegmentasi yang dilengkapi dengan titik knot sebagai lokasi perubahan struktur fungsi. Model ini menggunakan basis polinomial hingga derajat tertentu dan komponen truncated berupa $(x - k_j)_+^m$, di mana k_j merupakan knot. Jika nilai x_i sudah melewati titik knot k_l di hitung $(x_i - k_l)^m$. Jika x_i belum melewati knot nilainya 0. Parameter model diestimasi melalui metode *Ordinary Least Squares (OLS)* yang bekerja dengan meminimalkan jumlah kuadrat selisih antara nilai observasi dan estimasi model sehingga diperoleh fitting kurva yang optimal berdasarkan karakteristik data

Dengan nilai dapat dihitung dengan persamaan (Eubank, 1999):

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4)$$

Titik knot itu sendiri merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola pada interval yang berlainan. Memberikan suatu metode yang baik untuk memilih titik knot optimal yaitu metode GCV. Secara teoritis bahwa metode GCV mempunyai sifat optimal asimtotik [4]. Disamping metode GCV, juga bisa menggunakan metode CV. Metode CV dilakukan dengan menghapus setiap titik data secara bergantian kemudian dijumlahkan dan dikuadratkan dari perbedaan dari setiap titik data yang dihilangkan dari data yang cocok ke semua titik data lainnya (Dani & Adrianingsih, 2021). Rumus dari Metode GCV didefinisikan pada Persamaan (4)

$$GCV(M) = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{C(M)^2}{n}\right]} \quad (5)$$

Kriteria sederhana yang digunakan sebagai ukuran kinerja atas penaksir yang baik adalah *Mean squared error (MSE)* adalah

$$MSE = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2 \quad (6)$$

Regresi spline *nonparametrik* berusaha menguji parameter ini untuk menentukan apakah variabel respon dipengaruhi oleh variabel prediktor. Ada dua sesi pengujian pada parameter ini yaitu pengujian secara serentak atau simultan serta pengujian secara individu.

Kernel Gaussian merupakan salah satu metode dalam regresi nonparametrik yang digunakan untuk mengestimasi hubungan antara variabel respon dan variabel prediktor tanpa mengasumsikan bentuk fungsi tertentu (Farrell & Correa, 2007). Pendekatan ini bekerja dengan memberikan bobot yang lebih besar pada data pengamatan yang berada lebih dekat dengan titik yang sedang diestimasi, sedangkan pengamatan yang lebih jauh diberi bobot lebih kecil. Fungsi Kernel Gaussian memiliki bentuk kurva lonceng yang halus dan simetris, sehingga mampu menghasilkan estimasi yang kontinu dan stabil terhadap perubahan data (Härdle *et al.*, 2004). Metode ini sering digunakan karena sifatnya yang fleksibel dan kemampuannya dalam menangkap pola hubungan nonlinear secara efektif (Hastie *et al.*, 2009).

Gaussian Process (GP) atau proses *Gaussian* adalah teknik populer dalam machine learning dan banyak digunakan dalam analisis time series (Fan & Gijbels, 1996). Proses *Gaussian* adalah generalisasi dari distribusi *Gaussian*, ini mewakili distribusi probabilitas atas fungsi yang seluruhnya ditentukan oleh fungsi mean dan kovarian. Definisi proses *Gaussian* sebagai berikut (Purnaraga *et al.*, 2020).

Proses *Gaussian* sepenuhnya ditentukan oleh fungsi mean dan kovarian. Fungsi mean $m(x)$ dan fungsi kovarian $k(x, x')$ didefinisikan sebagai berikut

$$\begin{aligned} m(x) &= \mathbb{E}[f(x)] \\ k(x, x') &= \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))] \\ k(x, x') &= \exp\left(-\frac{(x, x')^2}{2h^2}\right) \end{aligned} \quad (7)$$

dan proses *Gaussian* ditulis dalam bentuk sebagai berikut:

$$f(x) \sim GP(m(x), k(x, x')) \quad (8)$$

Dalam pengaturan regresi, diasumsikan bahwa diberikan $D = \{(x_i, y_i) | i = 1, \dots, N\}$ dimana $x_i \in \mathbb{R}^D$ dan $y_i \in \mathbb{R}$ kumpulan data D dari N pengamatan dengan menggunakan $f(x)$ berikut: $y_i = f(x_i) + \delta_i$ dimana dengan tujuan untuk memprediksi y^* baru yang diberikan x^* , δ_i adalah *noise Gaussian* dengan mean nol dan varians (Rosadi *et al.*, 2022). Distribusi target pengamatan y sebelumnya diberikan oleh:

$$y \sim N(0, K(X, X)) \quad (9)$$

dimana $K(X, X)$ adalah matriks yang dibentuk dengan pelatihan dan X adalah matriks input ($n \times m$).

Selanjutnya, (X, X) disebut sebagai *Covarian Function* atau dikenal mengevaluasi fungsi kovarian antara semua pasangan titik sebagai *kernel* (matriks Gram). Penggunaan beberapa *kernel* dalam penelitian ini digunakan untuk melihat dari input model mana yang menghasilkan data pelatihan yang optimal (Sanusi *et al.*, 2019).

Titik knot merupakan perpaduan yang menunjukkan perubahan pola perilaku data. Pemilihan titik knot optimal menggunakan *Generalized Cross Validation* (GCV) paling minimum (Suparti *et al.*, 2018). Metode *Generalized Cross Validation* (GCV) dalam regresi *kernel* adalah satu metode untuk memilih *bandwidth* optimal dengan meminimalkan fungsi GCV. Optimasi GCV adalah memilih h optimal yang meminimalkan nilai GCV (Widiyantoro & Ratna, 2023):

$$GCV(h) = \frac{n^{-1} \sum_{i=1}^n [y_i - \hat{y}_i]^2}{\{1 - \text{tr}(H(h)/n)\}^2} \quad (10)$$

Dengan

$GCV(h)$: nilai GCV pada *bandwidth* h

n : GCV pada *bandwidth* h

y_i : data aktual subjek ke- i

\hat{y}_i : hasil estimasi subjek ke- i

$tr(H)$: jumlah dari elemen diagonal utama matriks penghalus $n \times n$

Salah satu kriteria untuk mengukur kebaikan model adalah dengan menggunakan R^2 . Dimana jika nilai $R^2 > 70\%$ maka dapat dikatakan model telah dianggap baik (Wand & Jones, 1995). Berikut adalah persamaan untuk mendapatkan nilai R^2 .

$$R^2 = \frac{SS_{regresi}}{SS_{total}} \quad (11)$$

3. METODE PENELITIAN

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2024. Data mencakup 38 provinsi di Indonesia dengan total 152 observasi. *Software* yang dimanfaatkan untuk menganalisis dalam penelitian ini adalah RStudio.

Tahap-tahap dalam penelitian untuk menganalisis hubungan variabel respon dan faktor pengaruhnya menggunakan pemodelan *spline truncated* dan *Kernel Gaussian* pada data longitudinal adalah sebagai berikut:

1. Persiapan dan Eksplorasi Data

- a. Mengumpulkan data sekunder dari Badan Pusat Statistik (BPS) tahun 2024 yang mencakup 38 provinsi (152 observasi).
- b. Melakukan eksplorasi data melalui statistik deskriptif untuk masing-masing variabel:
 - Y = Indeks Pembangunan Manusia (IPM)
 - X_1 = Harapan Lama Sekolah (HLS)
 - X_2 = Rata-rata Lama Sekolah (RRLS)
 - X_3 = Persentase Penduduk Miskin (PPM)
- c. Membuat *scatter plot* antara IPM dan setiap variabel prediktor untuk melihat pola hubungan awal.

2. Pemodelan Regresi Nonparametrik

a. *Spline truncated*

Adapun tahapan-tahapan yang harus dilakukan dalam pemodelan regresi *nonparametrik* dengan pendekatan *spline* yaitu:

1. Memodelkan indeks pembangunan manusia (IPM) dengan menggunakan model regresi *nonparametrik* *spline* dengan diujicobakan satu, dua, dan tiga titik knot untuk masing-masing variabel prediktor mengacu pada persamaan (3).
2. Menentukan titik knot optimal terbaik berdasarkan nilai GCV yang paling minimum dengan mengacu pada persamaan (4).
3. Melakukan pengujian signifikan parameter secara simultan dengan mengevaluasi performa model menggunakan kriteria seperti koefisien determinasi (R^2) dan *Mean squared error* (MSE) guna memastikan kualitas model yang dihasilkan.
4. Menginterpretasikan hasilnya dan menarik kesimpulan.

b. *Kernel Gaussian*

Adapun tahapan-tahapan yang harus dilakukan dalam pemodelan regresi *nonparametrik* dengan pendekatan *Kernel Gaussian* yaitu:

1. Memodelkan variabel respon (y) menggunakan metode regresi *nonparametrik* *Kernel Gaussian* mengacu pada persamaan (7).

2. Menentukan *bandwidth* optimal berdasarkan metode seperti Generalized Cross-Validation (GCV) mengacu pada persamaan (9).
 3. Melakukan pengujian signifikan parameter secara simultan dengan mengevaluasi performa model menggunakan kriteria seperti koefisien determinasi (R^2) dan *Mean squared error* (MSE) guna memastikan kualitas model yang dihasilkan.
 4. Menginterpretasikan hasilnya dan menarik kesimpulan.
3. Evaluasi dan Interpretasi Model

Kinerja model diukur menggunakan *Koefisien Determinasi* (R^2) untuk melihat proporsi variasi data yang dapat dijelaskan oleh model mengacu pada persamaan (10), serta *Mean Squared Error* (MSE) mengacu pada persamaan (5) untuk menilai tingkat kesalahan prediksi. Selanjutnya, dilakukan perbandingan performa antara metode *spline truncated* dan *kernel Gaussian*. Model dengan nilai R^2 lebih tinggi dan MSE lebih rendah dianggap lebih akurat.

4. HASIL DAN PEMBAHASAN

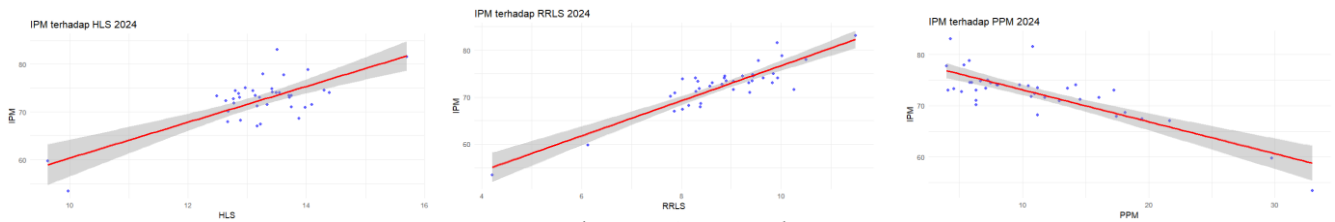
Eksplorasi data dilakukan untuk memahami karakteristik dasar data, distribusi variabel, serta hubungan awal antara variabel respon dan variabel prediktor sebelum dilakukan pemodelan regresi nonparametrik *spline* maupun *kernel*.

Tabel 1. Statistik Deskriptif

Statistik	Y (IPM)	X1 (HLS)	X2 (RRLS)	X3 (PPM)
Min	53.42	9.63	4.21	4.00
1st Qu.	71.08	12.86	8.30	6.32
Median	73.18	13.25	8.89	10.13
Mean	72.39	13.21	8.84	11.15
3rd Qu.	74.34	13.72	9.52	14.06
Max	83.08	15.70	11.49	32.97

Tabel 1. ini menyajikan distribusi variabel dalam penelitian. Statistik deskriptif dari data yang digunakan dalam penelitian ini menunjukkan variasi yang signifikan dalam Indeks Pembangunan Manusia (IPM) dan faktor-faktor yang mempengaruhinya. Nilai IPM memiliki rentang antara 53.42 hingga 83.08, dengan rata-rata 72.39, yang mencerminkan perbedaan tingkat pembangunan manusia antar wilayah. Harapan Lama Sekolah (HLS) memiliki nilai minimum 9.63 tahun dan maksimum 15.70 tahun, dengan median 13.25 tahun, menunjukkan bahwa sebagian besar daerah memiliki angka harapan lama sekolah di atas 12 tahun. Rata-rata Lama Sekolah (RRLS) bervariasi dari 4.21 hingga 11.49 tahun, dengan nilai tengah 8.89 tahun, yang berarti bahwa sebagian besar penduduk hanya menyelesaikan pendidikan dasar hingga menengah pertama.

Sementara itu, Persentase Penduduk Miskin (PPM) menunjukkan variasi yang cukup besar, dengan nilai minimum 4.00 dan maksimum 32.97, serta rata-rata 11.15, yang mengindikasikan adanya kesenjangan ekonomi antar wilayah. Nilai kuartil pertama dan ketiga untuk variabel ini masing-masing sebesar 6.32 dan 14.06, yang menunjukkan bahwa separuh daerah memiliki pengeluaran per kapita dalam rentang tersebut. Secara keseluruhan, distribusi variabel dalam penelitian ini mencerminkan hubungan antara faktor pendidikan dan ekonomi dengan tingkat pembangunan manusia, di mana wilayah dengan tingkat pendidikan dan pengeluaran per kapita yang lebih tinggi cenderung memiliki IPM yang lebih baik.



Gambar 1. *Scatter plot*

Berdasarkan *scatter plot* yang ditampilkan, terdapat hubungan yang jelas antara Indeks Pembangunan Manusia (IPM) dengan tiga variabel: Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RRLS), dan Pengeluaran per Kapita (PPM). Grafik pertama menunjukkan hubungan positif antara IPM dan HLS, di mana peningkatan HLS cenderung diikuti oleh peningkatan IPM. Hal serupa juga terlihat pada grafik kedua, yang menunjukkan korelasi positif antara IPM dan RRLS. Kedua grafik ini mengindikasikan bahwa semakin tinggi tingkat pendidikan (baik harapan maupun rata-rata lama sekolah), semakin tinggi pula IPM. Namun, grafik ketiga menunjukkan hubungan negatif antara IPM dan PPM, di mana peningkatan PPM justru berbanding terbalik dengan IPM. Hal ini mungkin menunjukkan bahwa daerah dengan presentase penduduk miskin lebih tinggi cenderung memiliki IPM lebih rendah, yang bisa disebabkan oleh faktor-faktor lain seperti ketimpangan ekonomi atau distribusi sumber daya yang tidak merata.

Selanjutnya yaitu hasil pemodelan Indeks Pembangunan Manusia (IPM) menggunakan regresi nonparametrik *spline truncated*. Diperoleh nilai GCV minimum yaitu sebesar 2,797 dengan lokasi titik knot yang optimal untuk masing-masing variabel prediktor adalah:

Variabel X_1 (HLS)

$$K_1 = 14,688$$

Variabel X_2 (RRLS)

$$K_1 = 10,276$$

Variabel X_3 (PPM)

$$K_1 = 28,141$$

Dimana dapat dituliskan model regresi *nonparametrik spline truncated* dibatasi dengan 1 titik knot yang merupakan titik knot terbaik adalah:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 (x_{i1} - 14,688)_+ + \hat{\beta}_3 x_{i2} + \hat{\beta}_4 (x_{i2} - 10,276)_+ + \hat{\beta}_5 x_{i3} + \hat{\beta}_6 (x_{i3} - 28,141)_+$$

Ditampilkan terlebih dahulu hasil estimasi parameter model regresi *nonparametrik spline truncated* yang terbaik, yaitu model regresi *nonparametrik spline truncated* dengan 1 titik knot.

Tabel 2. Hasil Model Regresi Spline

Variabel	Hasil Estimasi Parameter
Konstanta	$\beta_0 = 54,615$
X1 = Harapan Lama Sekolah	$\beta_{11} = 0,545$ $\beta_{12} = 1,575$
X2 = Rata-Rata Lama Sekolah	$\beta_{21} = -0,321$ $\beta_{22} = 6,145$
X3 = Persentase Penduduk Miskin	$\beta_{31} = 3,673$ $\beta_{32} = -0,509$

Berdasarkan hasil estimasi parameter pada tabel di atas, maka dapat dituliskan model regresi *nonparametrik spline truncated* yang terbaik dengan 3 titik knot sebagai berikut:

$$\hat{y}_i = 54,615 + 0,545x_{i1} + 1,575(x_{i1} - 14,688)_+ - 0,321x_{i2} + 6,145(x_{i2} - 10,276)_+ + 3,673x_{i3} - 0,509(x_{i3} - 28,141)_+$$

Evaluasi kinerja model regresi *spline truncated* dilakukan dengan menggunakan dua indikator utama, yaitu *Mean squared error (MSE)* dan *Koefisien Determinasi (R²)*. Hasil evaluasi menunjukkan bahwa nilai MSE yang diperoleh sebesar 1,8617, yang mencerminkan tingkat galat prediksi yang relatif rendah. Nilai MSE yang kecil ini mengindikasikan bahwa model memiliki kemampuan estimasi yang baik dalam merepresentasikan hubungan antara variabel-variabel yang dianalisis. Selanjutnya, nilai koefisien determinasi (*R²*) yang diperoleh sebesar 92,79% menunjukkan bahwa proporsi variabilitas data yang dapat dijelaskan oleh model mencapai 92,79%, sementara sisanya sebesar 7,21% kemungkinan disebabkan oleh faktor-faktor lain yang tidak dimodelkan atau oleh keberadaan *noise* dalam data. Dengan demikian, dapat disimpulkan bahwa model regresi *spline truncated* menunjukkan performa yang tinggi dan mampu memberikan representasi yang akurat terhadap struktur hubungan antar variabel dalam konteks analisis Indeks Pembangunan Manusia (IPM).

Selanjutnya berikut ini adalah hasil pemodelan Indeks Pembangunan Manusia (IPM) menggunakan regresi nonparametrik Kernel Gaussian. Pemilihan parameter optimal sangat penting dalam *kernel gaussian* untuk memastikan model yang akurat dan efisien. Metode *Generalized Cross Validation (GCV)* digunakan untuk menentukan parameter *bandwidth* optimal dalam estimasi *kernel*.

Tabel 3. Nilai *Bandwidth* Optimum

Indeks	<i>Bandwidth</i> 1	<i>Bandwidth</i> 2	<i>Bandwidth</i> 3	MSE	GCV Value
1	0,247	0,297	1,182	4,584	6,949
2	0,372	0,446	1,376	5,146	7,037
3	0,124	0,148	0,591	3,684	7,380
4	0,495	0,594	2,347	5,782	7,458
5	0,619	0,743	2,956	6,806	8,014

Tabel 3. merupakan hasil GCV, ditampilkan lima nilai pertama dari parameter GCV yang dihitung. Dari tabel di atas, nilai GCV minimum ditemukan sebesar 6,949. *Bandwidth* optimal dipilih berdasarkan nilai GCV minimum yang ditemukan. Dari hasil perhitungan:

$$\text{Bandwidth Optimal} = [0,247; 0,297; 1,182]$$

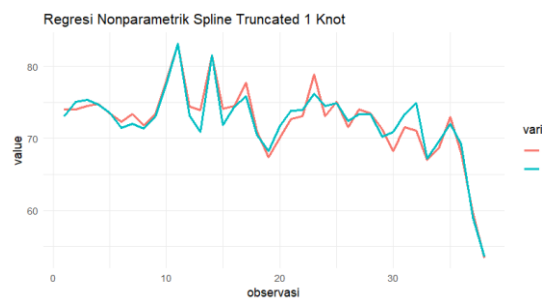
Bentuk umum dari estimator *Kernel Gaussian* kontinu adalah sebagai berikut:

$$k(x, x') = \exp\left(-\frac{(x, x')^2}{2 \times 0,247^2}\right) \times \exp\left(-\frac{(x, x')^2}{2 \times 0,297^2}\right) \times \exp\left(-\frac{(x, x')^2}{2 \times 1,182^2}\right)$$

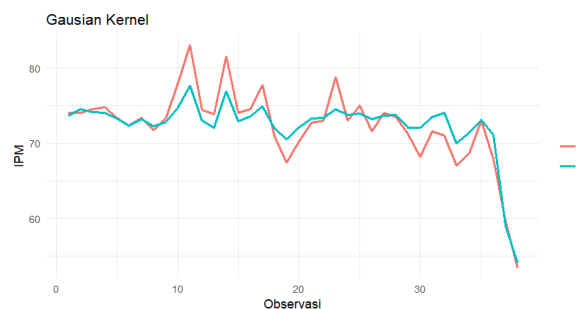
Gaussian Kernel digunakan untuk estimasi kepadatan probabilitas dalam ruang multi-variabel. Pemilihan *bandwidth* sangat penting dalam menentukan bentuk estimasi fungsi distribusi. Dengan menggunakan pendekatan GCV, kita mendapatkan parameter *bandwidth* yang optimal untuk model, yang akan menghasilkan estimasi dengan bias dan variansi yang seimbang.

Evaluasi model regresi *kernel* menggunakan *Mean squared error (MSE)* dan *Koefisien Determinasi (R²)* menunjukkan hasil yang baik. Nilai MSE sebesar 4,58 mengindikasikan galat prediksi yang rendah, sedangkan *R²* sebesar 82,25% menunjukkan model mampu menjelaskan sebagian besar variabilitas data. Dengan demikian, regresi *kernel* efektif dalam memodelkan hubungan antar variabel.

Ditampilkan visualisasi nilai \hat{y} (prediksi) dengan y (nilai aktual) pada Gambar berikut:



Gambar 1. Plot Regresi Spline



Gambar 2. Plot Gaussian *Kernel*

Berdasarkan nilai MSE dan R^2 , regresi *spline truncated* menunjukkan kinerja yang lebih baik dibandingkan regresi kernel. Model *spline truncated* memiliki MSE yang lebih kecil 1,86 dan R^2 yang lebih tinggi 92,79%, sehingga lebih akurat dalam memprediksi dan menjelaskan variabilitas data. Sebaliknya, regresi kernel memiliki MSE lebih besar 4,58 dan R^2 lebih rendah 82,25%, meskipun tetap memberikan performa yang cukup baik. Dengan demikian, regresi *spline truncated* merupakan metode yang lebih unggul dalam memodelkan hubungan antar variabel pada konteks analisis IPM.

Kemudian berdasarkan perbandingan visual antara nilai aktual dan nilai prediksi pada kedua metode, Regresi Spline menunjukkan kecocokan pola yang lebih stabil dan mendekati data aktual, terutama pada bagian tengah dan awal deret. Sebaliknya, Gaussian Kernel cenderung menghasilkan fluktuasi yang lebih tajam, sehingga beberapa bagian tampak kurang mengikuti pola data sebenarnya. Dengan demikian, Regresi Spline dapat dikatakan lebih baik dalam menggambarkan hubungan dan mengikuti pola data pada kasus ini.

5. KESIMPULAN

Berdasarkan hasil analisis regresi *nonparametrik* terhadap Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2024, dapat disimpulkan bahwa metode *Spline truncated* memiliki kinerja yang lebih baik dibandingkan metode *Kernel Gaussian*. Metode *Spline truncated* memberikan performa pemodelan yang lebih baik dibandingkan Kernel Gaussian dalam menganalisis faktor-faktor yang memengaruhi Indeks Pembangunan Manusia (IPM). *Spline truncated* lebih mampu menangkap pola hubungan yang kompleks dan memiliki perubahan struktural, sehingga menghasilkan prediksi yang lebih akurat dan representatif.

Sementara itu, Kernel Gaussian tetap bermanfaat untuk pola hubungan yang lebih halus, namun kurang optimal pada karakteristik data seperti dalam penelitian ini. Temuan ini menunjukkan bahwa *Spline truncated* lebih direkomendasikan sebagai pendekatan analisis, terutama untuk mendukung pengambilan keputusan berbasis data terkait pembangunan manusia.

Hasil penelitian ini diharapkan dapat menjadi referensi bagi pengambil kebijakan dalam perumusan strategi pembangunan manusia yang lebih tepat sasaran dan berbasis data, sehingga dapat meningkatkan kualitas hidup masyarakat di Indonesia.

Penelitian ini memiliki keterbatasan, yaitu penggunaan data hanya dari satu periode (2024) dan variabel yang terbatas, sehingga belum menggambarkan dinamika IPM secara menyeluruh. Penelitian lanjutan disarankan menggunakan data panel atau time series serta menambahkan variabel lain dan metode regresi *nonparametrik* lain untuk memperkuat hasil.

DAFTAR PUSTAKA

- Alwi, W. & Musfirah, M. 2021. Penerapan regresi nonparametrik spline dalam memodelkan faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2018. *Jurnal MSA*, 9(2): 112–119.
- Anwar, R., Rassiyanti, L. & Pitri, R. 2025. Analisis Regresi Kernel Gaussian untuk Memprediksi Indeks Pembangunan Manusia (IPM) Berdasarkan Faktor Sosial-Ekonomi Provinsi di Indonesia. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam*, 4(3): 44–53.
- Asti, A.P. & Darwis, S. 2023. Deteksi kerusakan bearing menggunakan komponen utama kernel. *Jurnal Riset Statistika*, 19–26.
- Badan Pusat Statistik. 2024. *Indeks Pembangunan Manusia 2024*. Jakarta: BPS.
- Dai, Y. & Zhang, Y. 2013. Machine learning in stock price trend forecasting. *Journal of Stanford University*, 21: 2021.
- Dani, A.T.R. & Adrianingsih, N.Y. 2021. Pemodelan regresi nonparametrik dengan estimator *spline truncated* vs deret Fourier. *Jambura Journal of Mathematics*, 3(1): 26–36.
- De Boor, C. 2001. *A practical guide to splines*. Springer, New York.
- Dewi, A.F. & Ahadiyah, K. 2023. Analisis Regresi *Spline truncated* pada Indeks Pembangunan Manusia Jawa Timur tahun 2021. *INFERENSI*, 6(1).
- Draper, N.R. & Smith, H. 1998. *Applied Regression Analysis*. Wiley, New York.
- Eubank, R.L. 1999. *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, New York.
- Fan, J. & Gijbels, I. 1996. *Local Polynomial Modelling and its Applications*. Chapman & Hall/CRC, London.
- Farrell, M.T. & Correa, A. 2007. Gaussian process regression models for predicting stock trends. *Relation*, 10: 3414.
- Härdle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Härdle, W., Müller, M., Sperlich, S. & Werwatz, A. 2004. *Nonparametric and Semiparametric Models*. Springer, Berlin.
(4 penulis → tidak perlu dkk.)
- Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The Elements of Statistical Learning*, 2nd ed. Springer.
- Purnaraga, T., Sifriyani, S. & Prangga, S. 2020. Regresi nonparametrik spline pada data laju pertumbuhan ekonomi di Kalimantan. *BAREKENG*, 14(3): 343–356.
- Rosadi, S., Rinaldi, A., Gunawan, W. & dkk. 2022. Implementasi metode regresi nonparametrik spline untuk menganalisis keuntungan produksi batu-bata. *Jurnal Ilmiah Matematika dan Terapan*, 19(2): 215–226.
- Ruppert, D., Wand, M.P. & Carroll, R.J. 2003. *Semiparametric Regression*. Cambridge University Press.
- Sanusi, W., Syam, R. & Adawiyah, R. 2019. Model regresi nonparametrik dengan pendekatan spline. *Journal of Mathematics, Computations, and Statistics*, 2(1): 70–81.

- Sudiasa, I.N., Yuliana, N., Pradnyana, I.W. & Ramadhan, B. dkk. 2023. Multivariable truncated spline nonparametric regression in modeling HDI of Southeast Sulawesi. *Proceedings of ICSMTR 2023*.
- Suparti, Santoso, R., Prahutama, A. & Devi, A.R. 2018. Regresi *Nonparametrik*. Wade Group.
- United Nations. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: United Nations.
- Wahba, G. 1990. *Spline Models for Observational Data*. SIAM, Philadelphia.
- Wand, M.P. & Jones, M.C. 1995. *Kernel Smoothing*. Chapman & Hall, London.
- Widiyantoro, D.N. & Ratna, M. 2023. Pemodelan Indeks Pembangunan Gender di Pulau Kalimantan menggunakan metode *spline truncated*. *Jurnal Sains dan Seni ITS*, 12(1): D45–D52.