

Pemodelan Indeks Kualitas Lingkungan Hidup di Indonesia dengan Spline Truncated dan MARS

Marfa Audilla Fitri¹, Suliyanto², M. Fariz Fadillah Mardianto³, Elly Ana⁴

^{1,2,3,4}Program Studi Statistika, Universitas Airlangga

E-mail: marfa.audilla.fitri-2021@fst.unair.ac.id¹, suliyanto@fst.unair.ac.id²,
m.farizfadillah.m@fst.unair.ac.id³, elly-a@fst.unair.ac.id⁴

Diajukan 1 April 2025 **Diperbaiki** 17 Juni 2025 **Diterima** 24 Juni 2025

Abstrak

Latar Belakang: Indonesia dengan kekayaan alamnya menghadapi tantangan besar dalam menjaga kualitas lingkungan akibat urbanisasi dan pertumbuhan ekonomi pesat. Berdasarkan *Environmental Performance Index* 2022, Indonesia memperoleh skor 28,2 dari 100, menempati peringkat ke-164 dari 180 negara. Indeks Kualitas Lingkungan Hidup (IKLH) 2022 tercatat sebesar 72,42, masih dalam kategori cukup. Untuk meningkatkan IKLH, diperlukan analisis terhadap faktor-faktor yang memengaruhinya.

Tujuan: Penelitian ini bertujuan menganalisis faktor-faktor yang berpengaruh signifikan terhadap IKLH di Indonesia menggunakan pendekatan statistik yang sesuai.

Metode: Penelitian ini menggunakan metode regresi nonparametrik, yaitu *Multivariate Adaptive Regression Spline* (MARS) dan *spline truncated* multiprediktor. Variabel prediktor yang dianalisis meliputi Indeks Pembangunan Manusia (IPM), kepadatan penduduk, sanitasi layak, persentase kemiskinan, dan Produk Domestik Regional Bruto (PDRB). Data diperoleh dari 34 provinsi di Indonesia tahun 2022.

Hasil: Metode *spline truncated* menghasilkan kinerja terbaik dengan MSE 5,63308, GCV 10,42, dan R² sebesar 82,63%, dibandingkan MARS dengan MSE 7,685, GCV 16,014, dan R² 79,3%. Semua variabel prediktor berpengaruh signifikan terhadap IKLH.

Kesimpulan: Faktor sosial dan ekonomi terbukti berpengaruh signifikan terhadap kualitas lingkungan. *Spline truncated* menjadi metode terbaik dalam memodelkan hubungan tersebut untuk mendukung penyusunan kebijakan lingkungan.

Kata kunci: IKLH, *Spline Truncated*, MARS, Regresi Nonparametrik.

Abstract

Background: Indonesia, endowed with abundant natural resources, faces substantial challenges in maintaining environmental quality amid rapid urbanization and economic growth. The 2022 *Environmental Performance Index* ranked Indonesia 164th out of 180 countries with a score of 28.2. Regionally, Indonesia ranked 22nd among 25 Asia-Pacific countries. The *Environmental Quality Index* (EQI), crucial for achieving the Sustainable Development Goals (SDGs), was recorded at 72.42 in 2022, classified as "fair." This condition underscores the need for in-depth analysis of key factors influencing environmental quality.

Objective: This study aims to examine significant factors affecting the *Environmental Quality Index* (EQI) across Indonesian provinces using appropriate nonparametric statistical methods.

Methods: A nonparametric regression approach, specifically the *Multivariate Adaptive Regression Spline* (MARS) and the *truncated spline* multipredictor model, was applied. Predictor variables included the *Human Development Index* (HDI), population density, access to proper sanitation, poverty rate, and *Gross Regional Domestic Product* (GRDP). Secondary data for 34 provinces in 2022 were sourced from the Central Bureau of Statistics and the Ministry of Environment.

Results: The *truncated spline* model demonstrated superior performance, achieving a minimal MSE of 5.63308, minimal GCV of 10.42, and R² of 82.63%, outperforming MARS, which yielded a minimal MSE of 7.685, GCV of 16.014, and R² of 79.3%. All predictor variables significantly influenced EQI.

Conclusion: Social and economic factors were found to significantly affect environmental quality. The *truncated spline* approach offers an effective modeling alternative, providing critical insights to support environmental policy development at the provincial level.

Keywords : IKLH, *Spline Truncated*, MARS, Nonparametric Regression.

PENDAHULUAN

Indonesia dengan kekayaan alamnya menghadapi tantangan besar dalam pemeliharaan dan perlindungan lingkungan. Pertumbuhan Produk Domestik Regional Bruto (PDRB) Indonesia yang pesat dan urbanisasi yang tinggi meningkatkan tekanan terhadap lingkungan, seperti meningkatnya konsumsi sumber daya alam, emisi gas rumah kaca, dan degradasi lingkungan akibat aktivitas ekonomi dan pembangunan infrastruktur. Hal ini sejalan dengan penilaian global yang menunjukkan bahwa pelestarian lingkungan di Indonesia masih tergolong rendah, bahkan dibandingkan dengan negara-negara lain di kawasan Asia Pasifik. Indonesia masih memperoleh nilai 28,2 dari 100 dan menempati peringkat ke-164 dari 180 negara di seluruh dunia. Secara regional, Indonesia berada di peringkat ke-22 dari 25 negara di Asia Pasifik (Yale, 2022). Selain itu, Pada tahun 2022, Indeks Kualitas Lingkungan Hidup (IKLH) di Indonesia sebesar 72,42 dari skala 100 yang masih dalam kategori cukup (Kementerian Lingkungan Hidup dan Kehutanan, 2023).

IKLH memiliki peranan yang penting dalam mencapai *Sustainable Development Goals* (SDGs), pada pilar pembangunan lingkungan terdapat enam poin di antaranya yaitu pada SDGS poin ke-6, 11, 12, 13, 14, dan 15 (Badan Perencanaan Pembangunan Nasional, 2017). Pengelolaan lingkungan yang sudah dilakukan pemerintah antara lain pengendalian pencemaran udara, pengendalian pencemaran air, pengendalian kerusakan lahan, pengendalian kerusakan ekosistem gambut, dan pengendalian pencemaran dan kerusakan pesisir dan laut. Dengan program-program tersebut, terjadi peningkatan IKLH pada tahun 2022 sebesar 0,97 poin dari tahun 2021 (Kementerian Lingkungan Hidup dan

Kehutanan Republik Indonesia, 2022). Namun, masih terdapat 7 provinsi yang belum mencapai target IKLH. Hal tersebut pastinya disebabkan oleh faktor-faktor tertentu yang cocok dianalisis lebih lanjut menggunakan metode statistik untuk menganalisis faktor yang berpengaruh signifikan dan faktor yang saling berinteraksi terhadap IKLH.

Untuk menganalisis indikator yang signifikan terhadap IKLH, salah satu metode yang digunakan adalah analisis regresi, yang memiliki dua tujuan utama yaitu menyelidiki hubungan antara variabel respons dan variabel prediktor, serta melakukan prediksi (Silverman, 1985). Model regresi terdiri dari tiga jenis: parametrik, nonparametrik, dan semiparametrik (Quadratullah, 2013). Jika bentuk kurva diketahui, model regresi parametrik digunakan (Islamiyati, 2022), sedangkan model nonparametrik lebih sesuai jika bentuk kurva tidak diketahui (Lestari et al., 2018). Karena data yang dianalisis tidak membentuk pola atau tren tertentu, penelitian ini memilih metode regresi nonparametrik, yaitu *Multivariate Adaptive Regression Spline* (MARS) dan *spline truncated* multiprediktor. Metode MARS unggul karena mampu mengidentifikasi interaksi antar variabel prediktor dan lebih fleksibel dalam menangani variabel prediktor yang tidak menunjukkan tren linear atau spesifik terhadap variabel respon (Wibowo, 2019), dengan mengombinasikan *spline truncated* dan pemisahan rekursif regresi spline (Friedman, 1991). Metode ini disarankan untuk variabel prediktor yang berjumlah antara tiga hingga dua puluh, agar interaksi yang memengaruhi IKLH dapat diidentifikasi (Steinberg, 1999). Selain itu, penggunaan *spline* membantu membuat model yang lebih fleksibel dan adaptif terhadap perubahan pola atau perilaku data pada interval tertentu, yang sulit dicapai oleh model linier atau polinomial (Chamidah et al., 2022). Keunggulan *spline* terletak pada kemampuannya untuk menafsirkan pola data sesuai perubahan yang terjadi (Dani & Ni'matuzzahroh, 2021), serta menggunakan konsep *truncated* untuk membatasi variabel respons seperti IKLH

hanya pada rentang yang sesuai, sehingga dapat mengurangi bias akibat *outlier* atau data yang kurang relevan (Hardle, 1994).

Berdasarkan penjelasan pada penelitian-penelitian sebelumnya dan realita di atas, terdapat faktor-faktor yang diduga berpengaruh terhadap IKLH seperti Indeks Pembangunan Manusia (IPM), Kepadatan Penduduk, Sanitasi Layak, Persentase Kemiskinan, dan Produk Domestik Regional Bruto (PDRB). Dengan adanya dugaan bahwa terdapat interaksi pada faktor-faktor tersebut serta pada plot IKLH terhadap masing-masing variabel prediktor tidak membentuk pola tren linear atau khusus (Risambessy et al., 2022), maka penelitian ini cocok dianalisis menggunakan regresi nonparametrik. Berdasarkan uraian di atas, peneliti tertarik untuk memodelkan IKLH dan faktor-faktornya menggunakan pendekatan MARS dan *spline truncated* multiprediktor. Hasil penelitian ini diharapkan dapat menjadi acuan dalam perbaikan dan kebijakan lebih lanjut untuk meningkatkan kualitas lingkungan hidup di Indonesia.

METODE

Desain Penelitian

Penelitian ini merupakan penelitian kuantitatif observasional yang bersifat deskriptif analitik. Pendekatan yang digunakan adalah *cross-sectional*, yaitu data dikumpulkan pada satu titik waktu untuk mengevaluasi hubungan antara variabel respon dan variabel prediktor. Metode statistik yang digunakan dalam penelitian ini adalah MARS dan *Spline Truncated* untuk mengetahui pengaruh dan interaksi beberapa variabel prediktor terhadap variabel respon, yaitu Indeks Kualitas Lingkungan Hidup (IKLH).

Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh provinsi di Indonesia yang berjumlah 34 provinsi. Karena

jumlah populasi dapat terjangkau seluruhnya, maka penelitian ini menggunakan sampel jenuh (*total sampling*), di mana semua anggota populasi dijadikan sampel.

Teknik Sampling

Teknik pengambilan sampel yang digunakan adalah *sampling jenuh*, yaitu teknik penentuan sampel ketika semua anggota populasi digunakan sebagai sampel. Teknik ini sesuai karena jumlah populasi hanya 34 provinsi dan seluruhnya tersedia datanya untuk dianalisis.

Subjek Penelitian

Data yang digunakan dalam penelitian ini adalah data tingkat provinsi di Indonesia, mencakup seluruh 34 provinsi. Data diperoleh dari sumber resmi seperti Badan Pusat Statistik (BPS) dan Kementerian Lingkungan Hidup dan Kehutanan (KLHK).

Populasi dalam penelitian ini adalah seluruh provinsi di Indonesia, yang berjumlah 34 provinsi. Karena seluruh populasi dapat dijangkau, maka penelitian ini menggunakan metode sampel jenuh (*total sampling*), di mana seluruh anggota populasi dijadikan sampel penelitian. Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon dan lima variabel prediktor, yaitu:

- 1) Indeks Kualitas Lingkungan Hidup (IKLH) (Y)
- 2) Indeks Pembangunan Manusia (IPM) (X_1)
- 3) Kepadatan Penduduk (X_2)
- 4) Persentase Akses Sanitasi Layak (X_3)
- 5) Persentase Penduduk Miskin (PPM) (X_4)
- 6) Produk Domestik Regional Bruto (PDRB) (X_5)

Semua variabel dalam penelitian ini berbentuk data kontinu. Variabel-variabel tersebut dipilih berdasarkan kontribusinya dalam menggambarkan kondisi lingkungan hidup dan faktor sosial-ekonomi di masing-masing provinsi. Analisis yang dilakukan bertujuan untuk memahami hubungan

antara kualitas lingkungan hidup dengan berbagai faktor sosial-ekonomi, sehingga dapat menjadi dasar dalam perumusan kebijakan pembangunan berkelanjutan di tingkat provinsi.

Teknik Analisis Data

Teknik analisis data dalam penelitian ini terdiri atas beberapa tahapan berikut:

1) Eksplorasi Data

Mendeskripsikan variabel penelitian meliputi nilai minimal, maksimal, rata-rata, dan standar deviasi. Menyajikan diagram batang Indeks Kualitas Lingkungan Hidup (IKLH) pada 34 provinsi serta scatter plot antara IKLH dan masing-masing variabel prediktor (IPM, Kepadatan Penduduk, Sanitasi Layak, Persentase Penduduk Miskin, dan PDRB).

2) Pemodelan Regresi Nonparametrik *Spline Truncated*

Memodelkan hubungan antara IKLH dan faktor-faktor prediktornya menggunakan metode *spline truncated* multiprediktor, dengan pemilihan model berdasarkan nilai koefisien determinasi (R^2) maksimal dan *Generalized Cross Validation* (GCV) minimal. Dilakukan uji signifikansi simultan dan parsial, serta pengujian asumsi residual.

3) Pemodelan MARS

Metode analisis data lainnya yang digunakan dalam penelitian ini adalah regresi nonparametrik dengan

pendekatan MARS. Model regresi nonparametrik untuk n pengamatan dapat dinyatakan sebagai:

$$y_i = f(x_i) + \varepsilon_i ; i = 1, 2, \dots, n$$

Model MARS yang optimal dipilih melalui proses *trial-and-error* dengan menguji berbagai kombinasi parameter, termasuk BF (*Basis Function*), MI (*Maximum Interaction*), dan MO (*Minimal Observation*) (Khoirunnisa et al., 2024). Kriteria utama untuk menentukan model terbaik adalah nilai GCV yang paling rendah. Model MARS secara umum dapat dinyatakan sebagai berikut:

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{km})]_+$$

Proses pemodelan mencakup pemilihan nilai BF, MI, dan MO terbaik berdasarkan nilai GCV minimal, R^2 maksimal, dan *Mean Square Error* (MSE) minimal, serta dilakukan uji signifikansi simultan dan parsial, serta pengujian asumsi residual.

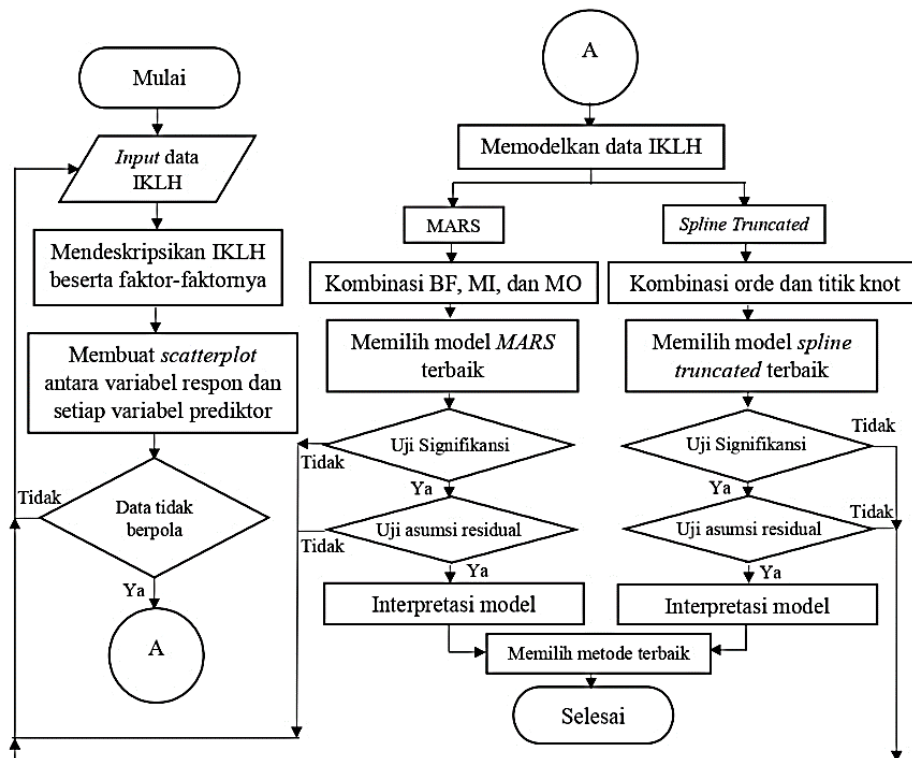
4) Interpretasi Model

Menginterpretasikan hasil model *spline truncated* dan MARS berdasarkan nilai koefisien serta hubungan antara faktor-faktor prediktor terhadap IKLH.

5) Pemilihan Model Terbaik

Menentukan metode terbaik antara *spline truncated* dan MARS berdasarkan kriteria MSE minimal, GCV minimal, dan R^2 maksimal

Sebagai ilustrasi, tahapan analisis data tersebut dapat digambarkan dalam bentuk diagram alir pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

HASIL DAN PEMBAHASAN

Statistika Deskriptif

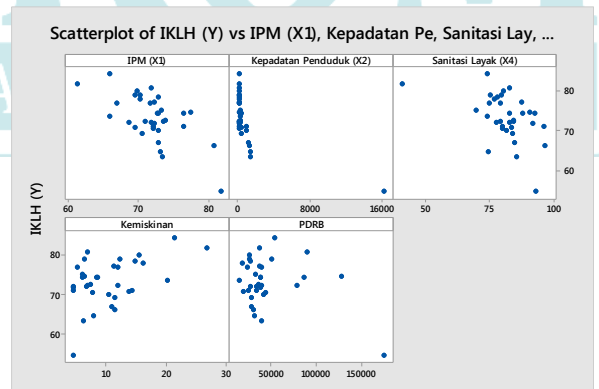
Analisis deskriptif ditunjukkan pada Tabel 1.

Tabel 1. Karakteristik IKLH dan Indikator IKLH

Variabel	Mean	Minimal	Maksimal
Y	73,067	54,650	84,220
X ₁	71,968	61,390	81,650
X ₂	751	10	16158
X ₃	81,00	40,34	96,21
X ₄	10,299	4,530	26,800
X ₅	43557	13093	174942

Statistik deskriptif menunjukkan adanya perbedaan signifikan di wilayah Indonesia. rata-rata IKLH sebesar 73,067 dengan kualitas lingkungan hidup yang cukup baik, sedangkan rata-rata IPM sebesar 71,968 menunjukkan pembangunan manusia yang relatif merata. Kepadatan penduduk bervariasi, dari 10 hingga 16.158 jiwa/km², dengan rata-rata 751 jiwa/km². Rata-rata sanitasi layak sebesar 81%, tetapi ada kesenjangan akses di beberapa wilayah. Tingkat kemiskinan bervariasi dari 4,53% hingga 26,8%, dengan rata-rata

10,3%. PDRB menunjukkan ketimpangan ekonomi yang besar, dengan rata-rata sebesar Rp 43.557 miliar.



Gambar 2. Scatterplot Tiap Variabel Prediktor terhadap Variabel IKLH

Berdasarkan Gambar 2, visualisasi scatterplot antara variabel respon dan masing-masing variabel prediktor, tidak ditemukan pola hubungan tertentu yang jelas. Hal ini menunjukkan bahwa variabel-variabel tersebut cocok digunakan sebagai komponen nonparametrik, sehingga metode regresi nonparametrik seperti *Spline Truncated* dapat diterapkan untuk pemodelan.

Pemodelan dengan Regresi Nonparametrik *Spline Truncated*

Setelah itu, dilakukan pemodelan Indeks Kualitas Lingkungan Hidup di Indonesia tahun 2022 menggunakan regresi nonparametrik *spline truncated* dengan 3 titik knot. Pemilihan model terbaik dari hasil ketiga pemodelan ditentukan berdasarkan kriteria GCV minimum seperti pada Tabel 2.

Tabel 2. Perbandingan Nilai Knot

Banyak Knot	Titik Knot				
	τ_1	τ_2	τ_3	τ_4	τ_5
Satu Knot	63,04	1328,2	44,90	6,347	26304,
	388	04	082	959	9
Dua Knot	63,45	1657,7	46,04	6,802	29608,
	73	551	10		004
	79,99	14839,	91,64	24,98	16172
	61	795	92	20	9,56
Tiga Knot	63,45	1657,7	46,04	6,802	29608,
	73	551	10		004
	79,99	14839,	91,64	24,98	16172
	61	795	92	20	9,56
	81,23	15828,	95,06	26,34	17163
	65	449	98		8,68

Estimasi Model Terbaik

Setelah itu ditentukan kebaikan model pada setiap titik knot seperti pada Tabel 3.

Tabel 3. Perbandingan Kebaikan Model Setiap Titik Knot

Titik Knot	GCV	R^2	MSE
1	10,85632	80,59822	6,34850
2	10,41895	82,63382	5,63308
3	10,42581	82,62462	5,63679

Berdasarkan Tabel 3, dapat diperoleh nilai GCV paling minimal dari masing-masing jumlah titik Knot adalah 10,41895 yaitu pada titik knot 2. Sehingga, dua titik knot dipilih sebagai model terbaik dari regresi nonparametrik *spline truncated*.

Model regresi nonparametrik *spline truncated* yang terbentuk dengan dua titik knot dan lima variabel atau komponen nonparametrik adalah:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - K_1)_+ + \hat{\beta}_3 (x_1 - K_2)_+ + \hat{\beta}_4 x_2 + \hat{\beta}_5 (x_2 - K_3)_+ + \hat{\beta}_6 (x_2 - K_4)_+ + \hat{\beta}_7 x_3 + \hat{\beta}_8 (x_3 - K_5)_+ +$$

$$\hat{\beta}_9 (x_3 - K_6)_+ + \hat{\beta}_{10} x_4 + \hat{\beta}_{11} (x_4 - K_7)_+ + \hat{\beta}_{12} (x_4 - K_8)_+ + \hat{\beta}_{13} x_5 + \hat{\beta}_{14} (x_5 - K_9)_+ + \hat{\beta}_{15} (x_5 - K_{10})_+$$

Tahap berikutnya adalah mengestimasi parameter model regresi nonparametrik *spline truncated* dengan menggunakan metode *Ordinary Least Squares (OLS)*, yang menghasilkan estimasi nilai untuk seluruh parameter dalam model regresi nonparametrik *spline truncated* seperti berikut.

Tabel 4. Estimasi Nilai Parameter

Variabel	Parameter	Nilai Estimasi
X_1	$\hat{\beta}_0$	0,032
	$\hat{\beta}_1$	1,077
	$\hat{\beta}_2$	-0,964
	$\hat{\beta}_3$	-0,038
X_2	$\hat{\beta}_4$	-0,141
	$\hat{\beta}_5$	0,001
	$\hat{\beta}_6$	0,075
X_3	$\hat{\beta}_7$	0,344
	$\hat{\beta}_8$	0,108
	$\hat{\beta}_9$	0,026
X_4	$\hat{\beta}_{10}$	0,260
	$\hat{\beta}_{11}$	0,121
X_5	$\hat{\beta}_{12}$	-0,475
	$\hat{\beta}_{13}$	-0,005
	$\hat{\beta}_{14}$	0,001
	$\hat{\beta}_{15}$	0,001

Pengujian Signifikansi Model Regresi Nonparametrik *Spline Truncated*

Uji simultan digunakan untuk mengetahui apakah parameter model regresi signifikan atau tidak.

Tabel 5. Hasil Uji F (Simultan)

Sumber	Jumlah Kuadrat	Rataan Kuadrat	P -value
Regresi	928,5228	61,901	0,00024
Galat	183,6735	10,204	
Total	1112,196		

Berdasarkan Tabel 5, nilai P -value sebesar 0,00024 yang artinya kurang dari α (0,05). Sehingga, diperoleh keputusannya adalah tolak H_0 . Hal ini menunjukkan bahwa minimal terdapat satu parameter yang tidak sama dengan 0. Setelah diperoleh hasil signifikan pada uji serentak, selanjutnya

adalah melakukan uji individu parameter model.

Pengujian signifikansi parameter model secara individu dilakukan untuk mengetahui parameter mana saja yang berpengaruh signifikan.

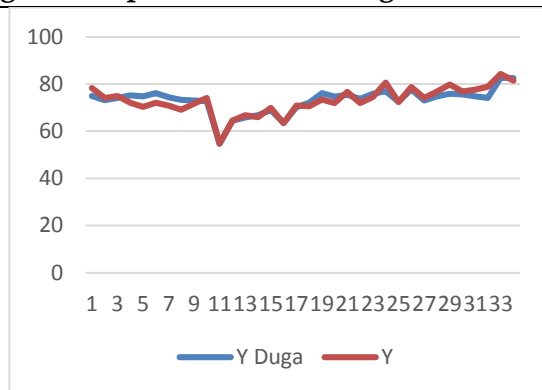
Tabel 6. Hasil Uji Individu Parameter

Parameter	P-value	Keterangan
$\hat{\beta}_0$	0,001	Signifikan
$\hat{\beta}_1$	0,001	Signifikan
$\hat{\beta}_2$	0,003	Signifikan
$\hat{\beta}_3$	0,366	Signifikan
$\hat{\beta}_4$	0,001	Signifikan
$\hat{\beta}_5$	0,002	Signifikan
$\hat{\beta}_6$	0,002	Signifikan
$\hat{\beta}_7$	0,001	Signifikan
$\hat{\beta}_8$	0,001	Signifikan
$\hat{\beta}_9$	0,144	Tidak signifikan
$\hat{\beta}_{10}$	0,896	Tidak signifikan
$\hat{\beta}_{11}$	0,022	Signifikan
$\hat{\beta}_{12}$	0,081	Tidak signifikan
$\hat{\beta}_{13}$	0,159	Tidak signifikan
$\hat{\beta}_{14}$	0,275	Tidak signifikan
$\hat{\beta}_{15}$	0,002	Signifikan

Berdasarkan Tabel 6, diketahui bahwa terdapat 11 parameter yang signifikan terhadap model, sedangkan ada 5 parameter lainnya yang tidak signifikan. Namun, meski terdapat parameter yang tidak signifikan, variabel tersebut masih digunakan karena minimal dalam satu variabel terdapat satu parameter yang signifikan, sehingga variabel prediktor $X_1, X_2, X_3, X_4,$ dan X_5 memberikan pengaruh yang signifikan terhadap model.

Interpretasi Model

Model regresi nonparametrik *spline truncated* terbaik adalah model dengan dua titik knot.



Gambar 3. Plot Y dan Y Duga *Spline Truncated*

Berdasarkan Gambar 3, dapat diketahui bahwa nilai hasil dugaan (\hat{Y}) mendekati nilai dari data faktual, yaitu IKLH (Y). Selain itu, berdasarkan model terbaik yang telah didapatkan, yaitu:

$$\hat{y} = 0,032 + 1,077x_1 - 0,964(x_1 - 63,4574)_+ - 0,038(x_1 - 79,9961)_+ - 0,141x_2 + 0,001(x_2 - 1657,755)_+ + 0,07(x_2 - 14839,8)_+ + 0,344x_3 + 0,108(x_3 - 46,0412)_+ + 0,026(x_3 - 91,6491)_+ + 0,26x_4 + 0,121(x_4 - 6,8024)_+ - 0,475(x_4 - 24,9820)_+ - 0,005x_5 + 0,001(x_5 - 29608)_+ + 0,001(x_5 - 161729,6)_+$$

Interpretasi terhadap model di atas adalah jika variabel selain X_1 dianggap konstan, maka pengaruh IPM (X_1) terhadap IKLH sebagai berikut.

$$\hat{y} = 0,032 + 1,077x_1 - 0,964(x_1 - 63,4574)_+ - 0,038(x_1 - 79,9961)_+$$

Apabila suatu provinsi memiliki IPM kurang dari 63,4574 maka kenaikan satu angka IPM akan menyebabkan IKLH naik sebesar 1,077 satuan. Apabila suatu provinsi memiliki IPM di antara 63,4574 hingga 79,9961 maka kenaikan satu angka IPM akan menyebabkan IKLH naik sebesar 0,133 satuan. Apabila suatu daerah memiliki IPM lebih dari 79,9961 maka kenaikan satu angka IPM akan menyebabkan IKLH turun sebesar 0,075 satuan. Dalam hal ini, pada tingkat IPM rendah hingga menengah, peningkatan IPM yang mencerminkan kualitas pembangunan berdampak positif pada IKLH karena adanya perbaikan nyata. Namun, pada IPM tinggi, kenaikan IPM yang hanya didorong oleh

kuantitas tanpa memperhatikan kualitas dapat menurunkan IKLH akibat dampak negatif pembangunan terhadap lingkungan.

Jika variabel selain X_2 dianggap konstan, maka pengaruh kepadatan penduduk X_2 terhadap IKLH sebagai berikut.

$$\hat{y} = 0,032 - 0,141x_2 + 0,001(x_2 - 1657,755)_+ + 0,075(x_2 - 14839,8)_+$$

Apabila suatu provinsi memiliki kepadatan penduduk kurang dari 1657,755 penduduk/km² maka kenaikan satu angka kepadatan penduduk akan menyebabkan IKLH turun sebesar 0,141 satuan. Apabila suatu provinsi memiliki kepadatan penduduk di antara 1657,755 hingga 14839,8 maka kenaikan satu angka kepadatan penduduk akan menyebabkan IKLH turun sebesar 0,14 satuan. Apabila suatu daerah memiliki kepadatan penduduk lebih dari 14839,8 penduduk/km² maka kenaikan satu angka kepadatan penduduk akan menyebabkan IKLH turun sebesar 0,066 satuan. Pada tingkat kepadatan penduduk rendah hingga menengah, peningkatan kepadatan penduduk menyebabkan penurunan IKLH yang lebih terkendali. Namun, pada tingkat kepadatan penduduk yang sangat tinggi, penurunan IKLH lebih besar, mencerminkan dampak negatif yang lebih signifikan pada kualitas lingkungan. Secara keseluruhan, kenaikan kepadatan penduduk tanpa perencanaan pembangunan yang berkelanjutan akan memperburuk kualitas lingkungan jika variabel selain X_3 dianggap konstan, maka pengaruh sanitasi layak (X_3) terhadap IKLH sebagai berikut.

$$\hat{y} = 0,032 + 0,344x_3 + 0,108(x_3 - 46,04)_+ + 0,026(x_3 - 91,6491)_+$$

Apabila suatu provinsi memiliki sanitasi layak kurang dari 46,0412% maka kenaikan satu angka sanitasi layak akan menyebabkan IKLH naik sebesar 0,344 satuan. Apabila suatu provinsi memiliki sanitasi layak di antara

46,0412% hingga 91,6491% maka kenaikan satu angka sanitasi layak akan menyebabkan IKLH naik sebesar 0,452 satuan. Apabila suatu daerah memiliki sanitasi layak lebih dari 91,6491% maka kenaikan satu angka sanitasi layak akan menyebabkan IKLH naik sebesar 0,478 satuan. Peningkatan akses sanitasi layak secara konsisten memberikan dampak positif terhadap IKLH. Pada tingkat akses sanitasi yang rendah, perbaikan sanitasi memiliki pengaruh signifikan karena secara langsung meningkatkan kualitas hidup masyarakat dengan memenuhi kebutuhan dasar kesehatan. Di sisi lain, pada tingkat sanitasi yang tinggi, dampaknya tetap positif dan bahkan lebih besar karena mencerminkan keberhasilan pembangunan yang berkelanjutan dan peningkatan standar kualitas lingkungan hidup.

Jika variabel selain X_4 dianggap konstan, maka pengaruh Persentase Penduduk Miskin (X_4) terhadap IKLH sebagai berikut.

$$\hat{y} = 0,032 + 0,26x_4 + 0,121(x_4 - 6,8024)_+ - 0,475(x_4 - 24,9820)_+$$

Apabila suatu provinsi memiliki PPM kurang dari 6,8024% maka kenaikan satu angka PPM akan menyebabkan IKLH naik sebesar 0,141 satuan. Apabila suatu provinsi memiliki PPM di antara 6,8024% hingga 24,982% maka kenaikan satu angka PPM akan menyebabkan IKLH naik sebesar 0,381 satuan. Apabila suatu daerah memiliki PPM lebih dari 24,982% maka kenaikan satu angka PPM akan menyebabkan IKLH turun sebesar 0,094 satuan. Hal ini mencerminkan bahwa pada wilayah dengan PPM rendah hingga menengah, peningkatan PPM dapat menunjukkan adanya distribusi pembangunan yang lebih merata, sehingga berdampak positif pada IKLH. Sebaliknya, pada wilayah dengan PPM tinggi, kenaikan PPM yang berlebihan dapat mencerminkan tekanan sosial dan ekonomi yang meningkat, sehingga berdampak negatif pada kualitas lingkungan.

Jika variabel selain X_5 dianggap konstan, maka pengaruh PDRB (X_5) terhadap IKLH sebagai berikut.

$$\hat{y} = 0,032 - 0,005x_5 + 0,001(x_5 - 29608)_+ + 0,001(x_5 - 161729,6)_+$$

Apabila suatu provinsi memiliki PDRB kurang dari 29608 maka kenaikan satu angka PDRB akan menyebabkan IKLH turun sebesar 0,005 satuan. Apabila suatu provinsi memiliki PDRB di antara 29608 hingga 161729,6 maka kenaikan satu angka PDRB akan menyebabkan IKLH turun sebesar 0,003 satuan. Apabila suatu daerah memiliki PDRB lebih dari 161729,6 maka kenaikan satu angka PDRB akan menyebabkan IKLH turun sebesar 0,003 satuan.

Hal ini sejalan dengan pengaruh pembangunan terhadap lingkungan, di mana pada tingkat PDRB rendah hingga menengah, peningkatan PDRB akan berdampak positif pada peningkatan IKLH. Sementara itu, pada tingkat PDRB yang rendah dan tidak didukung oleh pengelolaan sumber daya yang berkualitas atau keberlanjutan lingkungan akan memperburuk kualitas lingkungan.

Pemodelan dengan MARS

Dalam penelitian ini pemodelan IKLH di 34 provinsi Indonesia menggunakan metode MARS dengan berdasarkan 5 variabel prediktor dimana model terbaik yang dihasilkan berada pada nilai fungsi basis 10.

Didapatkan model MARS untuk mengestimasi IKLH di Indonesia sebagai berikut.

$$\hat{Y} = 73,187 - 0,009BF_1 + 0,002BF_3 + 0,472 \times 10^{-7}BF_4;$$

Persamaan di atas dapat disusun kembali dengan memasukkan nilai fungsi basis sebagai berikut.

$$\hat{Y} = 73,187 - 0,009(X_1 - 264) + 0,002(X_4 - 4,53)(264 - X_1) + 0,472 \times 10^{-7}(X_5 - 13092,81)(X_1 - 264)$$

Uji Signifikansi Model MARS

Uji signifikansi terdiri dari simultan dan parsial yang dijelaskan pada tabel 8 dan 9.

Tabel 8. Uji Simultan Koefisien Fungsi Basis Model MARS

Statistik Uji	Nilai
Statistik Uji <i>F</i>	38,271
<i>P-value</i>	$0,22329 \times 10^{-9}$

Hal ini dapat diartikan bahwa model yang diperoleh telah sesuai dan menunjukkan adanya hubungan antar koefisien fungsi basis dengan variabel respon.

Tabel 9. Uji Parsial Koefisien Fungsi Basis Model MARS

Parameter	Estimate	S.E.	<i>P-value</i>
<i>Constant</i>	73,187	0,684	$0,999 \times 10^{-15}$
<i>BF</i> ₁	-0,009	0,002	$0,199 \times 10^{-4}$
<i>BF</i> ₃	0,002	$0,401 \times 10^{-3}$	$0,798 \times 10^{-4}$
<i>BF</i> ₄	$0,472 \times 10^{-7}$	$0,107 \times 10^{-7}$	$0,121 \times 10^{-3}$

Hal ini diartikan bahwa semua fungsi basis berpengaruh signifikan terhadap model. Selanjutnya, mencari kepentingan tiap variabel yang paling memberikan kontribusi seperti pada Tabel 10.

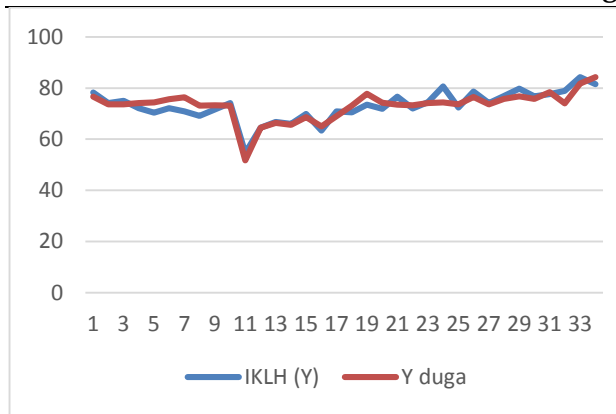
Tabel 10. Tingkat Kepentingan Variabel

Variabel	Tingkat Kepentingan	Pengurangan GCV
<i>X</i> ₂	100%	34,747
<i>X</i> ₅	46,277%	20,026
<i>X</i> ₃	43,104%	19,494
<i>X</i> ₁	0%	16,014
<i>X</i> ₄	0%	16,014

Berdasarkan Tabel 10, variabel prediktor yang paling berpengaruh terhadap variabel respon adalah variabel kepadatan penduduk (*X*₂) dengan tingkat kepentingan sebesar 100%. Selain itu, variabel kepadatan penduduk dapat mengurangi nilai GCV sebesar 34,747, apabila variabel tersebut dimasukkan ke dalam model.

Interpretasi Model MARS Terbaik

Selanjutnya dilakukan interpretasi dari model MARS yang telah didapatkan.



Gambar 4. Plot Y dan Y dugaan Model MARS

Berdasarkan Gambar 4, dapat diketahui bahwa nilai hasil dugaan (Y') mendekati nilai dari data faktual, yaitu IKLH (Y). Selain itu, berdasarkan model terbaik yang telah didapatkan, maka hasil interpretasinya adalah sebagai berikut.

$$BF_1 = \begin{cases} (X_1 - 264); & \text{untuk } X_1 > 264 \\ 0; & \text{untuk } X_1 \text{ yang lain} \end{cases}$$

Interpretasi nilai fungsi basis satu (BF_1) dengan koefisien $-0,009$ bermakna bahwa setiap kenaikan satu satuan BF_1 akan menurunkan IKLH sebesar $0,009$, dengan fungsi basis BF_3 dan BF_4 dianggap konstan. Selain itu, makna lainnya adalah provinsi yang memiliki nilai kepadatan penduduk lebih dari 264 jiwa/ km^2 akan memberikan kontribusi signifikan pada penurunan IKLH. Hal ini selaras dengan kondisi yang ada di lapangan yaitu semakin tinggi kepadatan penduduk, maka peluang IKLH juga menurun.

$$BF_3 = \begin{cases} (X_4 - 4,53)(264 - X_1); & X_4 > 4,53 \text{ dan } X_1 < 264 \\ 0; & X_1 \text{ dan } X_4 \text{ yang lain} \end{cases}$$

Interpretasi nilai fungsi basis tiga (BF_3) dengan koefisien $0,002$ bermakna bahwa setiap kenaikan satu satuan BF_3 akan menaikkan IKLH sebesar $0,002$, dengan fungsi basis BF_1 dan BF_4 dianggap konstan. Selain itu, makna lainnya adalah provinsi yang memiliki nilai persentase penduduk miskin lebih dari $4,53\%$ dan kepadatan penduduk kurang dari 264 jiwa/ km^2 akan memberikan kontribusi signifikan pada peningkatan IKLH.

$$BF_4 = \begin{cases} (X_5 - 13092,8)(X_1 - 264); & X_5 > 13092,8, X_1 > 264 \\ 0; & X_1 \text{ dan } X_5 \text{ yang lain} \end{cases}$$

Interpretasi nilai fungsi basis empat (BF_4) dengan koefisien $4,72 \times 10^{-8}$ bermakna bahwa setiap kenaikan satu satuan BF_4 akan menaikkan IKLH sebesar $4,72 \times 10^{-8}$, dengan fungsi basis BF_3 dianggap konstan. Selain itu, makna lainnya adalah provinsi yang memiliki nilai persentase PDRB lebih dari $13092,808$ dan kepadatan penduduk lebih dari 264 jiwa/ km^2 akan memberikan kontribusi signifikan pada peningkatan IKLH. Namun, kondisi di lapangan menunjukkan bahwa semakin tinggi kepadatan penduduk, dampak negatif terhadap lingkungan cenderung meningkat, sehingga peluang untuk meningkatkan IKLH menjadi lebih sulit..

Pemilihan Model Terbaik

Selanjutnya dilakukan pemilihan metode terbaik dengan membandingkan GCV, R^2 , dan MSE seperti pada Tabel 11.

Tabel 11. Pemilihan Model Terbaik

Model	GCV	R^2 (%)	MSE
Regresi	10,41895	82,63382	5,63308
Nonparametrik <i>Spline Truncated</i>			
MARS	16,01	79,3	7,685

Berdasarkan Tabel 12, dapat diperoleh nilai MSE terkecil dari masing-masing model adalah $5,63308$ yang diperoleh dari model *Spline Truncated*. Sehingga, model terbaik yaitu regresi nonparametrik *Spline Truncated*.

PENUTUP

Kesimpulan

Berdasarkan hasil penelitian, model regresi nonparametrik *Spline Truncated* menunjukkan kinerja terbaik dibandingkan model MARS dalam memodelkan Indeks Kualitas Lingkungan Hidup (IKLH) terhadap variabel-variabel prediktor, yaitu Indeks Pembangunan Manusia (IPM), kepadatan penduduk, sanitasi layak, persentase kemiskinan, dan Produk Domestik Regional Bruto (PDRB). Model regresi nonparametrik *spline truncated* menghasilkan kriteria evaluasi terbaik dengan nilai *Mean Square Error* (MSE)

sebesar 5,63308, *Generalized Cross Validation* (GCV) sebesar 10,41895, dan koefisien determinasi (R^2) sebesar 82,63%. Temuan ini mengindikasikan bahwa regresi nonparametrik spline truncated mampu membentuk hubungan yang lebih fleksibel dan akurat antara IKLH dan faktor-faktor prediktornya. Hasil penelitian ini dapat dijadikan sebagai dasar dalam perumusan kebijakan untuk peningkatan kualitas lingkungan hidup di Indonesia.

Saran

Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel prediktor lain yang relevan dengan kualitas lingkungan, seperti tingkat pencemaran atau luas ruang terbuka hijau, serta menggunakan data panel untuk menangkap perubahan lingkungan dari waktu ke waktu. Selain itu, dapat dilakukan eksplorasi metode nonparametrik lain guna memperoleh model yang lebih akurat dan mendukung perumusan kebijakan peningkatan kualitas lingkungan hidup secara berkelanjutan di Indonesia.

DAFTAR PUSTAKA

B. W. Silverman. (1985). Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1), 1–52. <https://doi.org/10.1111/j.2517-6161.1985.tb01327.x>

Badan Perencanaan Pembangunan Nasional (Bappenas). (2017). Pilar Pembangunan Lingkungan. Badan Perencanaan Pembangunan Nasional. <https://sustainability.undip.ac.id/wp-content/uploads/2020/09/Pilar-Pembangunan-Lingkungan.pdf>

Chamidah, N., Lestari, B., Budiantara, I.

N., Saifudin, T., Rulaningtyas, R., Aryati, A., Wardani, P., & Aydin, D. (2022). Consistency and Asymptotic Normality of Estimator for Parameters in Multiresponse Multipredictor Semiparametric Regression Model. *Symmetry*, 14(2), 336. <https://doi.org/10.3390/sym14020336>

Dani, A. T. R., & Ni'matuzzahroh, L. (2021). Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 14(1), 24–29. <https://doi.org/10.36456/jstat.vol14.n01.a3840>

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1). <https://doi.org/10.1214/aos/1176347963>

Hardle, W. (1994). *Applied Nonparametric Regression*. Cambridge University Press.

Islamiyati, A., Anisa, A., Raupong, R., Massalesse, J., Sirajang, N., Sahriman, S., & Wahyuni, A. (2022). Estimasi Model Regresi Spline Kubik Tersegmen dengan Metode Penalized Least Square. *Al-Khwarizmi : Jurnal Pendidikan Matematika Dan Ilmu Pengetahuan Alam*, 10(2), 139–148. <https://doi.org/10.24256/jpmipa.v10i2.3197>

Kementerian Lingkungan Hidup Dan Kehutanan. (2023). Profil Indeks Kualitas Lingkungan Hidup 2022. <https://ppkl.menlhk.go.id/website/index.php?q=1158&s=9bd7909ea32c89f68e82ed0cbe1f30a2e1d72b1c>

Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia. (2022). Kualitas Lingkungan Hidup Indonesia Meningkatkan dalam Lima Tahun Terakhir. <https://Ppid.Menlhk.Go.Id/Berita/Siaran-Pers/6972/Kualitas-Lingkungan->

[Hidup-Indonesia-Meningkat-Dalam-Lima-Tahun-Terakhir](#)

EPI Results - Environmental Performance Index. Yale Center for Environmental Law & Policy. <https://epi.yale.edu/epi-results/2022/component/epi>

Khoirunnisa, W., Fatekurohman, M., & Tirta, I. M. (2024). Analisis Ketahanan Hidup Pasien COVID-19 Menggunakan Pendekatan Multivariate Adaptive Regression Spline (MARS). *Jurnal Statistika Dan Komputasi*, 3(1), 11–21. <https://doi.org/10.32665/statkom.v3i1.2700>

Lestari, B., Fatmawati, Budiantara, I. N., & Chamidah, N. (2018). Estimation of Regression Function in Multi-Response Nonparametric Regression Model Using Smoothing Spline and Kernel Estimators. *Journal of Physics: Conference Series*, 1097, 012091. <https://doi.org/10.1088/1742-6596/1097/1/012091>

Qudratullah, M. F. (2013). *Analisis Regresi Terapan Teori, Contoh Kasus dan Aplikasi dengan SPSS*. Andi Offset.

Risambessy, S., Aulele, S. N., & Lembang, F. K. (2022). Misclassification Analysis of Elementary School Accreditation Data in Ambon City Using Multivariate Adaptive Regression Spline. *Jurnal Matematika, Statistika Dan Komputasi*, 18(3), 394–406. <https://doi.org/10.20956/j.v18i3.19451>

Steinberg, D. (1999). *MARS User Guide*. Salford System.

Wibowo, A. (2019). Multivariate Adaptive Regression Splines Modeling For Household Food Security In Central Borneo Province 2017. *Global Science Education Journal*, 1(1), 39–47. <https://jurnal.sainsglobal.com/index.php/ges/article/view/517>

Yale College. (2022). Yale Center for Environmental Law & Policy. 2022