

Pemodelan Indeks Pembangunan Gender Provinsi Jawa Barat Menggunakan Regresi Nonparametrik *Penalized Spline*

Attala Rania Insiro¹, Sri Sulistijowati Handajani², Sri Subanti³

Program Studi Statistika FMIPA UNS^{1,2,3}
attalarania@student.uns.ac.id

Abstrak—Indeks Pembangunan Gender (IPG) merupakan salah satu indikator untuk mengukur keberhasilan pembangunan manusia berbasis gender. Indeks pembangunan gender menggambarkan perbandingan capaian Indeks Pembangunan Manusia (IPM) laki-laki dan perempuan. Pengukuran IPG di Indonesia menunjukkan bahwa kesetaraan gender di Indonesia masih rendah dan terjadi ketidakmerataan pencapaian IPG pada masing-masing provinsi di Indonesia. Satu-satunya provinsi di Pulau Jawa yang mempunyai IPG lebih rendah dari IPG Indonesia adalah Jawa Barat, yaitu dengan capaian IPG sebesar 89,36. Oleh karena itu, penelitian ini menganalisis faktor-faktor yang memengaruhi indeks pembangunan gender di Provinsi Jawa Barat untuk memperoleh model IPG di Provinsi Jawa Barat. Variabel terikat yang digunakan dalam penelitian ini yaitu indeks pembangunan gender (IPG) di Provinsi Jawa Barat tahun 2021 yang terdiri dari 27 kabupaten/kota, sedangkan variabel bebasnya adalah Angka Harapan Hidup (AHH) perempuan, Harapan Lama Sekolah (HLS) perempuan, Rata-Rata Lama Sekolah (RLS) perempuan, Angka Partisipasi Sekolah (APS) SMA, persentase penduduk miskin, dan rasio jenis kelamin. Pola data variabel bebas terhadap variabel terikat tidak mengikuti pola tertentu sehingga dilakukan pemodelan menggunakan regresi nonparametrik *penalized spline*. Model regresi nonparametrik *penalized spline* yang dihasilkan memiliki nilai *Mean Squared Error* (MSE) sebesar 3,321185 dan koefisien determinasi (R^2) sebesar 78,84%. Keenam variabel bebas yang digunakan memiliki pengaruh yang signifikan terhadap IPG Jawa Barat tahun 2021 dengan jumlah knot yang bervariasi pada setiap variabel bebas.

Kata kunci: IPG, *Penalized Spline*, GCV, Knot

I. PENDAHULUAN

Kesetaraan gender merupakan hal penting dalam pembangunan dan kemajuan suatu negara. Hal ini dibuktikan dengan kesetaraan gender yang menjadi tujuan kelima dari Tujuan Pembangunan Berkelanjutan (TPB) atau *Sustainable Development Goals* (SDGs), yaitu “Mencapai Kesetaraan Gender dan Memberdayakan Kaum Perempuan” [1]. Dalam hal ini kesetaraan gender adalah kondisi dimana laki-laki dan perempuan mendapatkan hak, kesempatan, penghargaan, dan penghormatan yang sama sehingga sama-sama dapat berperan, berpartisipasi, dan berkontribusi dalam pencapaian pembangunan manusia yang seutuhnya. Salah satu cara untuk mengukur kesetaraan pembangunan manusia berbasis gender adalah dengan Indeks Pembangunan Gender (IPG). Indeks pembangunan gender adalah indeks yang menggambarkan perbandingan antara capaian Indeks Pembangunan Manusia (IPM) laki-laki dan perempuan.

Badan Pusat Statistik (BPS) mencatat IPG di Indonesia selama lima tahun terakhir tidak menunjukkan peningkatan yang signifikan, yaitu dari tahun 2017 hingga 2021 hanya mengalami peningkatan sebesar 0,31 persen [2]. Laporan *Human Development Report* (HDR) tahun 2022 juga menyatakan capaian IPG di Indonesia tahun 2021 sebesar 91,27 masih di bawah capaian IPG dunia sebesar 95,8 dan berada dalam kelompok tiga yaitu kelompok negara dengan capaian IPG sedang [3]. Pada level ASEAN, IPG di Indonesia berada di peringkat sembilan dari sepuluh negara. Hal ini menunjukkan bahwa kesetaraan gender di Indonesia masih rendah diantara negara-negara ASEAN dan dunia.

Pengukuran IPG menunjukkan adanya ketidakmerataan pencapaian IPG pada masing-masing provinsi di Indonesia. Jawa Barat merupakan satu-satunya provinsi di Pulau Jawa yang mempunyai IPG lebih rendah dari IPG Indonesia meskipun Jawa Barat adalah provinsi dengan penduduk terbanyak di pulau Jawa. IPG Jawa Barat pada tahun 2021 adalah sebesar 89,36 dan termasuk dalam sepuluh provinsi dengan IPG terendah di Indonesia, sedangkan nilai IPM Jawa Barat sebesar 72,45 dan termasuk dalam sepuluh provinsi dengan IPM tertinggi di Indonesia. Hal ini menunjukkan bahwa keberhasilan pembangunan manusia di

Jawa Barat belum diiringi dengan kesetaraan gender. Untuk itu, perlu dilakukan identifikasi faktor-faktor yang berpengaruh terhadap IPG, agar upaya peningkatan nilai IPG bisa terfokus pada faktor tersebut.

Variabel yang digunakan dalam penelitian ini merujuk pada penelitian yang telah dilakukan sebelumnya. Penelitian sebelumnya tentang Indeks Pembangunan Gender (IPG) telah dilakukan oleh Rahayu dan Wachidah pada tahun 2022 dengan menganalisis faktor-faktor yang memengaruhi Indeks Pembangunan Gender (IPG) di Jawa Barat Tahun 2020 menggunakan regresi nonparametrik spline [4]. Pada penelitian ini terdapat enam variabel yang memiliki pengaruh signifikan terhadap IPG yaitu rata-rata lama sekolah perempuan, harapan lama sekolah perempuan, perempuan yang memiliki keluhan kesehatan, tingkat pengangguran terbuka perempuan, tingkat partisipasi angkatan kerja perempuan, dan rasio jenis kelamin. Pada tahun 2018, Sari melakukan penelitian tentang aplikasi metode regresi nonparametrik spline multivariabel untuk pemodelan faktor-faktor yang memengaruhi indeks pembangunan gender di provinsi Jawa Barat, didapatkan semua variabel bebas signifikan memengaruhi IPG di Jawa Barat antara lain, angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran berdasarkan gender, dan keterlibatan perempuan di parlemen [5].

Salah satu metode yang dapat digunakan untuk mengetahui hubungan antara variabel bebas dan variabel terikat adalah analisis regresi [6]. Terdapat tiga pendekatan dalam analisis regresi, yaitu pendekatan parametrik, semiparametrik, dan nonparametrik. Regresi nonparametrik digunakan jika bentuk pola data tidak diketahui. Salah satu pendekatan nonparametrik yang sering digunakan adalah spline. Spline memiliki kelebihan yaitu model cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak, karena terdapat titik knot yang menunjukkan terjadinya perubahan pola data. Salah satu contoh regresi spline adalah *penalized spline* (P-Spline) [7]. Estimator *penalized spline* diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS) dan pemilihan model regresi *penalized spline* dilakukan dengan memilih orde, titik knot, dan parameter penghalus (λ) yang optimal dengan *generalized cross validation* (GCV) minimum. Pada *penalized spline* titik knot terletak pada titik-titik kuantil dari nilai tunggal variabel bebas, sehingga lebih efektif untuk data yang berjumlah besar. Oleh karena itu, penelitian ini menggunakan regresi nonparametrik *penalized spline* karena dianggap sesuai untuk memodelkan IPG Jawa Barat yang bentuk pola datanya tidak membentuk suatu pola tertentu.

Tujuan dari penelitian ini adalah untuk memperoleh model indeks pembangunan gender di Provinsi Jawa Barat menggunakan regresi nonparametrik *penalized spline* dan memperoleh nilai kebaikan model yang dihasilkan. Adapun manfaat dari penelitian ini antara lain, diharapkan akan memberi kajian baru dan referensi dalam pemanfaatan metode regresi nonparametrik *penalized spline* dalam pemodelan faktor-faktor yang memengaruhi Indeks Pembangunan Gender serta dapat memberikan informasi dan bahan masukan bagi instansi pemerintahan khususnya Pemerintah Provinsi Jawa Barat terkait upaya peningkatan Indeks Pembangunan Gender (IPG).

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder tahun 2021 yang diperoleh dari publikasi Badan Pusat Statistik (BPS) Provinsi Jawa Barat yang terdapat pada website BPS dalam indikator Sosial dan Kependudukan. Unit observasi adalah seluruh Kabupaten/Kota di Provinsi Jawa Barat yaitu sebanyak 27 kabupaten/kota. Variabel terikat yang digunakan dalam penelitian ini yaitu Indeks Pembangunan Gender (Y), sedangkan variabel prediktor yang digunakan antara lain, Angka Harapan Hidup (AHH) perempuan (X_1), Harapan Lama Sekolah (HLS) perempuan (X_2), Rata-Rata Lama Sekolah (RLS) perempuan (X_3), Angka Partisipasi Sekolah (APS) SMA (X_4), persentase penduduk miskin (X_5), dan rasio jenis kelamin (X_6).

B. Landasan Teori

1) Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu metode statistika yang digunakan untuk mengetahui hubungan antara variabel terikat dan variabel bebas dimana bentuk kurva data tidak diketahui sebelumnya [8]. Model umum regresi nonparametrik ditulis sebagai :

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

dengan y_i adalah variabel terikat pada pengamatan ke- i , x_i adalah variabel bebas pada pengamatan ke- i , $f(x_i)$ adalah fungsi regresi pada pengamatan ke- i yang tidak diketahui bentuknya, dan ε_i adalah residu pada pengamatan ke- i .

2) *Regresi Spline*

Regresi spline adalah pendekatan dengan mencocokkan data dan tetap memperhitungkan kemulusan kurva. Spline merupakan metode analisis regresi yang merupakan potongan-potongan polinomial dengan sifat tersegmen yang digabungkan oleh beberapa titik knot dan menghasilkan fungsi regresi yang sesuai dengan data [8]. Secara umum bentuk fungsi spline dengan p orde dan k titik knot dinyatakan sebagai :

$$\hat{f}(x_i) = \sum_{q=0}^p \beta_q x_i^q + \sum_{j=1}^k \beta_{(p+j)} (x_i - K_j)_+^p \tag{2}$$

dengan $\hat{f}(x_i)$ merupakan fungsi dari variabel ke- i , β_q dan $\beta_{(p+j)}$ adalah koefisien regresi spline dengan $q = 1, 2, \dots, p$ dan $j = 1, 2, \dots, k$ serta $(x_i - K_j)_+^p$ adalah fungsi *truncated* pada variabel ke- i dengan letak titik knot K_j pada orde ke- p .

3) *Regresi Penalized Spline*

Penalized spline merupakan salah satu bentuk estimator dari regresi spline yang diperoleh dengan meminimumkan fungsi *penalized least square (PLS)* [7]. *Penalized least square* adalah fungsi pendugaan yang menggabungkan antara fungsi *least square* dan kemulusan kurva (*smooth*) seperti berikut.

$$PLS = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{j=1}^k \beta_{(p+j)}^2 \tag{3}$$

dengan λ merupakan parameter pemulus bebas x_i dan $\lambda \geq 0$. Menggunakan metode *Ordinary Least Square* (OLS) diperoleh fungsi *penalized spline* dari variabel bebas x_i dalam bentuk matriks berikut.

$$\hat{f}(X) = X(X^T X + n\lambda D)^{-1} X^T Y \tag{4}$$

4) *Pemilihan Parameter Penghalus (λ) dan Orde Optimal*

Pemilihan parameter penghalus (λ) dan orde yang optimal dapat menggunakan kriteria *Generalized Cross Validation (GCV)* minimum. Fungsi GCV dapat dinyatakan sebagai :

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

dengan $H(\lambda) = X(X^T X + n\lambda D)^{-1} X^T$ dan $MSE = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

5) *Pemilihan Jumlah Knot Optimal*

Dalam *penalized spline*, knot terletak pada titik-titik kuantil dari nilai *unique* (tunggal) variabel bebas $\{x_i\}_{i=1}^n$ yang membagi segugus pengamatan menjadi $(k + 1)$ bagian yang sama. Pada penelitian ini, algoritma yang digunakan untuk memilih jumlah knot (k) optimal adalah algoritma *Full-Search*, dimana jumlah knot yang akan dihitung tidak akan dipilih, melainkan dihitung semua yaitu 1,2,3,... untuk $k < (n_{uniq} - p - 1)$, dengan n_{uniq} adalah banyaknya nilai *unique* dari variabel bebas $\{x_i\}_{i=1}^n$ dan p adalah orde polinomial [9].

C. *Langkah Analisis*

1. Melakukan analisis statistik deskriptif untuk mengetahui karakteristik data dalam penelitian meliputi rata-rata, variansi, nilai minimum, dan nilai maksimum dari variabel terikat dan masing-masing variabel bebas.
2. Membuat *scatterplot* antara variabel terikat dengan masing-masing variabel bebas untuk mengetahui pola hubungan antar variabel terikat dan bebas.
3. Menentukan nilai estimasi dari $\hat{f}(X_i)$ untuk masing-masing bebas berdasarkan estimator *penalized spline* untuk mendapatkan orde polinomial, jumlah knot, dan parameter penghalus optimal dengan kriteria GCV minimum.
4. Mengestimasi parameter model aditif regresi nonparametrik berdasarkan estimator *penalized spline* dengan orde polinomial dan titik knot optimal yang diperoleh dari langkah sebelumnya.
5. Melakukan pemodelan regresi nonparametrik *penalized spline* untuk masing-masing variabel bebas.
6. Melakukan pemodelan regresi nonparametrik *penalized spline* secara lengkap.
7. Mengukur nilai kebaikan model regresi nonparametrik *penalized spline*.

8. Melakukan uji asumsi normalitas.
9. Menarik kesimpulan dari analisis yang dilakukan.

III. HASIL DAN PEMBAHASAN

A. Analisis Statistik Deskriptif

Karakteristik dari variabel yang digunakan dapat diketahui dengan melihat statistik deskriptif, yaitu rata-rata, variansi, nilai minimum, dan nilai maksimum.

TABEL 1. STATISTIK DESKRIPTIF VARIABEL

Variabel	Rata-Rata	Variansi	Minimum	Maksimum
Y	89,02	16,29	79,07	95,18
X_1	74,29	1,84	71,59	76,89
X_2	12,90	0,69	11,65	14,46
X_3	8,25	2,26	5,7	11,11
X_4	70,02	54,75	55,52	83,76
X_5	8,97	8,37	2,58	13,13
X_6	102,4	2,31	100,3	105,7

Tabel 1 menunjukkan bahwa 27 kabupaten/kota di Provinsi Jawa Barat memiliki rata-rata Indeks Pembangunan Gender (Y) sebesar 89,02. Nilai variansi sebesar 16,29 termasuk cukup besar, yang menunjukkan bahwa terdapat ketidakmerataan pembangunan manusia berbasis gender di Jawa Barat pada tahun 2021, hal ini ditandai dengan Kabupaten Bandung Barat yang memiliki nilai IPG terendah sebesar 79,07 dan Kabupaten Sumedang yang memiliki nilai IPG tertinggi sebesar 95,18.

Variabel X_1 yaitu Angka Harapan Hidup (AHH) perempuan memiliki rata-rata sebesar 74,29. Nilai variansi sebesar 1,84 menandakan bahwa data AHH perempuan memiliki variansi yang kecil atau setiap kabupaten/kota di Jawa Barat tidak terlalu bervariasi. AHH perempuan tertinggi tercatat pada nilai 76,89 di Kota Bekasi dan terendah sebesar 71,59 di Kabupaten Tasikmalaya.

Variabel X_2 yaitu Harapan Lama Sekolah (HLS) perempuan memiliki rata-rata sebesar 12,90 dan variansi 0,69. Secara rata-rata anak perempuan usia 7 tahun yang masuk jenjang pendidikan formal di Jawa Barat memiliki peluang untuk bersekolah selama 12,90 tahun atau setara dengan Diploma I. Kabupaten Cianjur memiliki nilai HLS perempuan terkecil yaitu 11,65 tahun dan Kota Bandung memiliki nilai HLS perempuan terbesar yaitu 14,46 tahun.

Variabel X_3 yaitu Rata-Rata Lama Sekolah (RLS) perempuan memiliki rata-rata sebesar 8,25 dan variansi 2,26. Nilai RLS perempuan terendah sebesar 5,77 di Kabupaten Indramayu dan tertinggi sebesar 11,11 di Kota Bekasi. Berdasarkan nilai terendah dan tertinggi, dapat diketahui bahwa pendidikan formal terendah yang diperoleh penduduk perempuan di Jawa Barat adalah jenjang Sekolah Dasar (SD) dan yang tertinggi adalah jenjang Sekolah Menengah Atas (SMA).

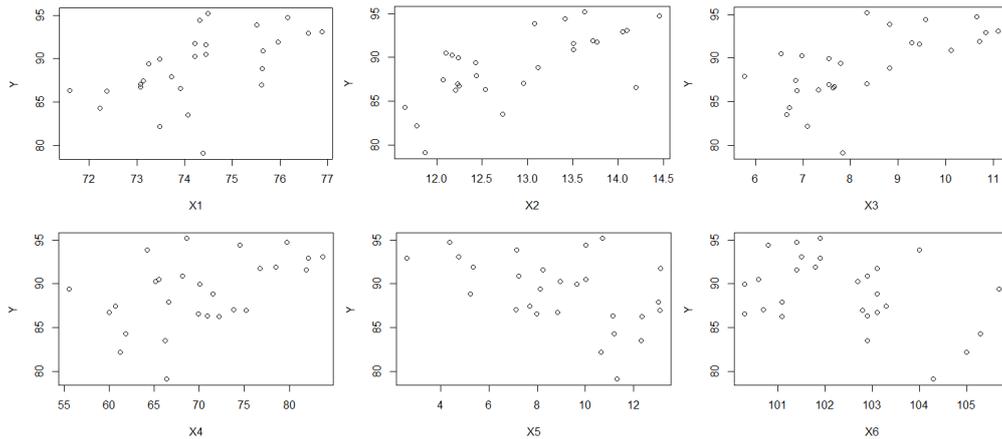
Variabel X_4 yaitu Angka Partisipasi Sekolah (APS) SMA memiliki rata-rata sebesar 70,02% dan variansi 54,75. Nilai APS SMA terendah yaitu 55,52 yang dimiliki oleh Kabupaten Bogor, sedangkan nilai APS SMA tertinggi sebesar 83,76 pada Kota Bekasi.

Variabel X_5 yaitu persentase penduduk miskin memiliki rata-rata sebesar 8,97 dan variansi 8,37. Persentase penduduk miskin terendah yaitu 2,58 pada kota Depok dan tertinggi sebesar 13,13 pada kota Tasikmalaya.

Variabel X_6 yaitu rasio jenis kelamin memiliki rata-rata sebesar 102,4, yang berarti jumlah penduduk laki-laki di Jawa Barat lebih banyak daripada jumlah penduduk perempuan karena nilai rasio lebih dari 100. Rasio jenis kelamin terendah berada di Kabupaten Ciamis yaitu 100,3 yang berarti jumlah penduduk laki-laki lebih banyak dibandingkan penduduk perempuan, sedangkan rasio jenis kelamin tertinggi berada di Kabupaten Bogor yaitu 105,7 yang berarti jumlah penduduk laki-laki lebih banyak dibandingkan penduduk perempuan.

B. Scatterplot Variabel Bebas Terhadap Variabel Terikat

Pola hubungan antara IPG dengan enam variabel yang memengaruhinya yaitu AHH perempuan (X_1), HLS perempuan (X_2), RLS perempuan (X_3), APS SMA (X_4), persentase penduduk miskin (X_5), dan rasio jenis kelamin (X_6) dapat diketahui dengan melihat *scatterplot*.



GAMBAR 1. SCATTERPLOT ANTARA $X_1, X_2, X_3, X_4, X_5,$ DAN X_6 TERHADAP Y

Gambar 1 menggambarkan pola data antara IPG Jawa Barat dengan masing-masing variabel bebas yang memengaruhi tidak membentuk pola tertentu, seperti linier, kuadratik, dan kubik. Sebaran titik-titik data acak dan tidak beraturan, sehingga dapat diasumsikan bahwa semua variabel bebas merupakan komponen nonparametrik dan dapat dilakukan pemodelan regresi nonparametrik.

C. Estimasi Model Hubungan IPG Jawa Barat Pada Masing-Masing Variabel Bebas

Model regresi nonparametrik *penalized spline* terbaik diperoleh dari model regresi dengan menggunakan orde polinomial, parameter penghalus (λ), dan titik knot optimal yang diperoleh dari nilai GCV minimum. Berdasarkan analisis yang dilakukan dengan program *software* R, didapatkan parameter penghalus optimal untuk masing-masing variabel bebas.

TABEL 2. PARAMETER PENGHALUS OPTIMAL VARIABEL BEBAS

Bebas	Orde	Jumlah Knot	Lamdha	GCV	Titik Knot 1	Titik Knot 2	Titik Knot 3
X_1	1	1	27	12,82	74,31	-	-
X_2	1	3	0	9,04	12,22	12,63	13,6
X_3	1	1	0,7	10,80	7,66	-	-
X_4	1	2	27	14,18	66,29	72,75	-
X_5	2	3	0	13,62	7,17	8,89	11,17
X_6	3	1	0,6	13,71	102,7	-	-

Tabel 2 memperlihatkan bahwa variabel $X_1, X_2, X_3,$ dan X_4 memiliki orde polinomial optimal bernilai satu atau berbentuk linier. X_5 memiliki orde polinomial optimal bernilai dua atau berbentuk kuadratik, dan X_6 memiliki orde polinomial optimal bernilai tiga atau berbentuk kubik. Orde polinomial, titik knot, dan parameter penghalus optimal digunakan untuk pemodelan regresi nonparametrik *penalized spline* dan didapatkan nilai $\hat{\beta}$ untuk masing-masing variabel bebas seperti berikut :

$$\hat{\beta}_1^T = (-358,012; 0,026; 0,004)$$

$$\hat{\beta}_2^T = (-69,038; 12,913; 19,668; 14,381; -9,645)$$

$$\hat{\beta}_3^T = (-89,699; 0,111; 0,263)$$

$$\hat{\beta}_4^T = (73,521; 0,217; 0,038; 0,037)$$

$$\hat{\beta}_5^T = (91,106; 1,394; -0,228; 1,598; -3,601; 7,224)$$

$$\hat{\beta}_6^T = (-9102,038; 115,886; 0,387; -0,006; 0,776)$$

Dari hasil pada Tabel 2 dan nilai $\hat{\beta}$ dapat dilakukan pemodelan *penalized spline* seperti pada persamaan (2).

Persamaan regresi nonparametrik *penalized spline* untuk masing-masing variabel adalah :

$$1. \hat{f}(X_1) = -358,012 + 0,026X_1 + 0,004(X_1 - 74,31)_+ \quad (6)$$

$$2. \hat{f}(X_2) = -69,038 + 12,913X_2 + 19,668(X_2 - 12,22)_+ + 14,381(X_2 - 12,63)_+ - 9,645(X_2 - 13,6)_+ \quad (7)$$

$$3. \hat{f}(X_3) = -89,699 - 0,111X_3 + 0,263(X_3 - 7,66)_+ \quad (8)$$

$$4. \hat{f}(X_4) = 73,521 + 0,217X_4 + 0,038(X_4 - 66,29)_+ + 0,037(X_4 - 72,75)_+ \quad (9)$$

$$5. \hat{f}(X_5) = 91,106 + 1,394X_5 - 0,228X_5^2 + 1,598(X_5 - 7,17)_+^2 - 3,601(X_5 - 8,89)_+^2 + 7,224(X_5 - 11,17)_+^2 \quad (10)$$

$$6. \hat{f}(X_6) = -9102,038 + 115,886X_6 + 0,387X_6^2 - 0,006X_6^3 + 0,776(X_6 - 102,7)_+^3 \quad (11)$$

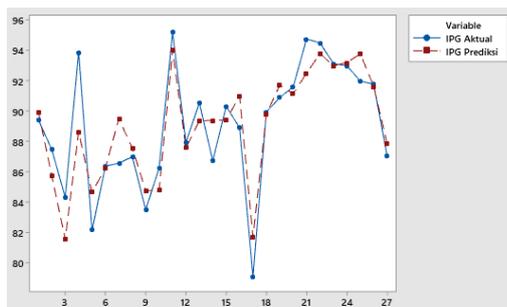
Secara lengkap bentuk estimator fungsi nonparametrik *penalized spline* pada data Indeks Pembangunan Gender (IPG) di Jawa Barat Tahun 2021 dapat dinyatakan:

$$\begin{aligned} \hat{Y} &= \hat{f}(X_1) + \hat{f}(X_2) + \hat{f}(X_3) + \hat{f}(X_4) + \hat{f}(X_5) + \hat{f}(X_6) \\ \hat{Y} &= -358,012 + 0,026X_1 + 0,004(X_1 - 74,31)_+ - 69,038 + 12,913X_2 + 19,668(X_2 - 12,22)_+ \\ &\quad + 14,381(X_2 - 12,63)_+ - 9,645(X_2 - 13,6)_+ - 89,699 - 0,111X_3 \\ &\quad + 0,263(X_3 - 7,66)_+ + 73,521 + 0,217X_4 + 0,038(X_4 - 66,29)_+ \\ &\quad + 0,037(X_4 - 72,75)_+ + 91,106 + 1,394X_5 - 0,228X_5^2 + 1,598(X_5 - 7,17)_+^2 \\ &\quad - 3,601(X_5 - 8,89)_+^2 + 7,224(X_5 - 11,17)_+^2 - 9102,038 + 115,886X_6 \\ &\quad + 0,387X_6^2 - 0,006X_6^3 + 0,776(X_6 - 102,7)_+^3 \quad (12) \end{aligned}$$

D. Validasi Model

Untuk mengukur kebaikan model digunakan nilai *Mean Squared Error* (MSE) dan koefisien determinasi (R^2). Model regresi nonparametrik *penalized spline* pada data Indeks Pembangunan Gender (IPG) di Jawa Barat memiliki nilai MSE sebesar 3,321185 dan R^2 sebesar 78,84%. Nilai R^2 tersebut menunjukkan bahwa pengaruh variabel AHH perempuan, HLS perempuan, RLS perempuan, APS SMA, persentase penduduk miskin, dan rasio jenis kelamin terhadap IPG Provinsi Jawa Barat tahun 2021 adalah sebesar 78,84%, sedangkan 21,12% IPG Provinsi Jawa Barat tahun 2021 dipengaruhi oleh faktor lain yang tidak diteliti pada penelitian ini.

Adapun plot antara nilai IPG aktual dan nilai IPG hasil prediksi dapat dilihat pada Gambar 2.



GAMBAR 2. PLOT PERBANDINGAN ANTARA IPG AKTUAL DAN IPG PREDIKSI

Gambar 2 menunjukkan bahwa penggunaan regresi nonparametrik *penalized spline* menghasilkan tingkat fleksibilitas yang baik dalam mengatasi pola data yang naik atau turun secara tajam dengan bantuan titik knot karena regresi nonparametrik memiliki fleksibilitas yang tinggi atau dapat mencari bentuk estimasi kurva regresinya sendiri dengan menyesuaikan data tanpa dipengaruhi faktor dari peneliti [10]. Hal ini dapat dilihat dari pola hasil prediksi yang cenderung mengikuti pola pada data IPG aktual.

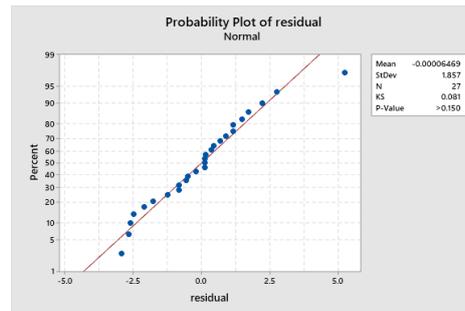
E. Uji Asumsi Normalitas

Setelah diperoleh model regresi nonparametrik *penalized spline* pada data indeks pembangunan gender di Provinsi Jawa Barat tahun 2021 secara lengkap, langkah selanjutnya adalah menguji distribusi dari residu yang diperoleh menggunakan metode *Kolmogorov Smirnov* dengan hipotesis sebagai berikut.

H_0 : Residu berdistribusi normal

H_1 : Residu tidak berdistribusi normal

Tingkat Signifikansi yang digunakan adalah *alpha* (α) sebesar 0,05. Daerah kritisnya yaitu H_0 akan ditolak jika nilai *p-value* < α . Dari analisis yang dilakukan, didapatkan hasil statistik uji seperti pada Gambar 3.



GAMBAR 3. PLOT PROBABILITAS RESIDU

Gambar 3 menunjukkan bahwa sebaran titik-titik data mendekati garis lurus, sehingga dapat disimpulkan bahwa data mengikuti distribusi normal dan diketahui bahwa nilai *p-value* > 0,150 > $\alpha = 0,05$, sehingga H_0 gagal ditolak yang artinya residu berdistribusi normal.

F. Interpretasi Model Regresi Nonparametrik Penalized Spline

Setelah pengujian asumsi normalitas terpenuhi, maka dilakukan interpretasi model regresi nonparametrik *penalized spline* untuk setiap variabel bebas. Berdasarkan persamaan (6), (7), (8), (9), (10), dan (11) diperoleh interpretasi sebagai berikut.

1. Pengaruh AHH perempuan (X_1) terhadap IPG Jawa Barat dengan variabel lainnya (X_2, X_3, X_4, X_5 , dan X_6) dianggap konstan yaitu :

Persamaan (6) dapat dituliskan dalam model *truncated* sebagai :

$$\hat{f}(X_1) = \begin{cases} -358,012 + 0,026X_1, & X_1 < 74,31 \\ -358,309 + 0,03X_1, & X_1 \geq 74,31 \end{cases} \quad (13)$$

Dari persamaan (13) didapatkan interpretasi bahwa pada saat Angka Harapan Hidup (AHH) perempuan kurang dari 74,31%, maka jika AHH perempuan naik satu persen, nilai indeks pembangunan gender (IPG) cenderung naik sebesar 0,026%. Terdapat 14 daerah yang termasuk dalam interval ini yaitu Kab. Bogor, Kab. Sukabumi, Kab. Cianjur, Kab. Garut, Kab. Tasikmalaya, Kab. Ciamis, Kab. Cirebon, Kab. Majalengka, Kab. Indramayu, Kab. Purwakarta, Kab. Karawang, Kab. Pangandaran, Kota Tasikmalaya, dan Kota Banjar. Pada saat AHH perempuan sama dengan atau lebih dari 74,31%, maka jika AHH perempuan naik satu persen, nilai indeks pembangunan gender (IPG) cenderung naik sebesar 0,03%. Terdapat 13 daerah dalam interval kedua ini. Hasil interpretasi mengenai hubungan antara AHH perempuan dengan IPG Jawa Barat menunjukkan bahwa peningkatan AHH perempuan akan menyebabkan peningkatan pada IPG Jawa Barat.

2. Pengaruh HLS perempuan (X_2) terhadap IPG Jawa Barat dengan variabel lainnya (X_1, X_3, X_4, X_5 , dan X_6) dianggap konstan yaitu :

Persamaan (7) dapat dituliskan dalam model *truncated* sebagai :

$$\hat{f}(X_2) = \begin{cases} -69,038 + 12,913X_2, & X_2 < 12,22 \\ -309,381 + 32,581X_2, & 12,22 \leq X_2 < 12,63 \\ -491,013 + 46,962X_2, & 12,63 \leq X_2 < 13,6 \\ -359,841 + 37,317X_2, & X_2 \geq 13,6 \end{cases} \quad (14)$$

Dari persamaan (14) didapatkan interpretasi bahwa pada saat Harapan Lama Sekolah (HLS) perempuan kurang dari 12,22, maka jika HLS perempuan naik satu satuan, nilai indeks pembangunan

gender (IPG) cenderung naik sebesar 12,913. Terdapat 7 daerah yang termasuk dalam interval ini. Pada saat HLS perempuan sama dengan atau lebih dari 12,22 dan kurang dari 12,63, maka jika HLS perempuan naik satu satuan, nilai IPG cenderung naik sebesar 32,581. Terdapat 6 daerah dalam interval ke dua ini. Pada saat HLS perempuan sama dengan atau lebih dari 12,63 dan kurang dari 13,6, maka jika HLS perempuan naik satu satuan, nilai indeks pembangunan gender (IPG) cenderung naik sebesar 46,962. Terdapat 7 daerah dalam interval ketiga ini. Pada saat HLS perempuan sama dengan atau lebih dari 13,6, maka jika HLS perempuan naik satu satuan, nilai indeks pembangunan gender (IPG) cenderung naik sebesar 37,317. Terdapat 6 daerah dalam interval keempat ini. Hasil interpretasi mengenai hubungan antara HLS perempuan dengan IPG Jawa Barat menunjukkan bahwa peningkatan HLS perempuan akan menyebabkan peningkatan pada IPG Jawa Barat.

3. Variabel lainnya dapat diinterpretasikan dengan cara yang sama seperti interpretasi X_1 dan X_2 .

IV. SIMPULAN DAN SARAN

Berdasarkan analisa dan pembahasan yang telah dilakukan, diperoleh kesimpulan bahwa keenam variabel bebas yang digunakan, yaitu Angka Harapan Hidup (AHH) perempuan, Harapan Lama Sekolah (HLS) perempuan, Rata-Rata Lama Sekolah (RLS) perempuan, Angka Partisipasi Sekolah (APS) SMA, persentase penduduk miskin, dan rasio jenis kelamin memiliki pengaruh yang signifikan terhadap Indeks Pembangunan Gender (IPG) Jawa Barat tahun 2021. Persamaan regresi dugaan untuk model regresi nonparametrik *penalized spline* terbaik diperoleh pada saat X_1 , X_2 , X_3 dan X_4 berorde 1 atau linier, X_5 berorde 2 atau kuadrat, dan X_6 berorde 3 atau kubik, serta jumlah knot yang bervariasi pada setiap variabel bebas. Model regresi nonparametrik *penalized spline* untuk data IPG di Provinsi Jawa Barat pada tahun 2021 adalah :

$$\begin{aligned} \hat{Y} = & -358,012 + 0,026X_1 + 0,004(X_1 - 74,31)_+ - 69,038 + 12,913X_2 + 19,668(X_2 - 12,22)_+ \\ & + 14,381(X_2 - 12,63)_+ - 9,645(X_2 - 13,6)_+ - 89,699 - 0,111X_3 \\ & + 0,263(X_3 - 7,66)_+ + 73,521 + 0,217X_4 + 0,038(X_4 - 66,29)_+ \\ & + 0,037(X_4 - 72,75)_+ + 91,106 + 1,394X_5 - 0,228X_5^2 + 1,598(X_5 - 7,17)_+^2 \\ & - 3,601(X_5 - 8,89)_+^2 + 7,224(X_5 - 11,17)_+^2 - 9102,038 + 115,886X_6 \\ & + 0,387X_6^2 - 0,006X_6^3 + 0,776(X_6 - 102,7)_+^3 \end{aligned}$$

Persamaan regresi dugaan untuk model regresi nonparametrik *penalized spline* untuk data Indeks Pembangunan Gender (IPG) di Provinsi Jawa Barat pada tahun 2021 memiliki nilai *Mean Square Error* (MSE) sebesar 3,321185 dan koefisien determinasi (R^2) sebesar 78,84%, serta asumsi normalitas yang terpenuhi.

Saran untuk penelitian selanjutnya yaitu lebih baik menambahkan variabel bebas lain yang mungkin dapat memengaruhi Indeks Pembangunan Gender (IPG) di Provinsi Jawa Barat seperti perempuan yang memiliki keluhan kesehatan, tingkat pengangguran terbuka perempuan, dan tingkat partisipasi angkatan kerja perempuan, agar memperoleh model dengan nilai akurasi yang lebih tinggi. Peneliti juga menyarankan untuk penelitian selanjutnya dapat mencoba menggunakan metode regresi nonparametrik dengan pendekatan yang lain seperti *spline truncated* dan *B-Spline*.

DAFTAR PUSTAKA

- [1] Badan Perencanaan Pembangunan Nasional. 2017. Ringkasan Metadata Indikator Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs) Indonesia. Jakarta : Badan Perencanaan Pembangunan Nasional.
- [2] Badan Pusat Statistik (BPS). 2022. Indeks Pembangunan Gender (IPG). <https://www.bps.go.id/indicator/40/463/1/indeks-pembangunan-gender-ipg-.html>. Diakses tanggal 20 September 2022.
- [3] Human Development Report. *Gender Development Index* (GDI). Gender Development Index | Human Development Reports (undp.org). Diakses tanggal 25 September 2022.
- [4] Rahayu, N. F., dan Wachidah, L. 2022. Regresi Nonparametrik Spline untuk Memodelkan Faktor-faktor yang Memengaruhi Indeks Pembangunan Gender (IPG) di Jawa Barat Tahun 2020. *Bandung Conference Series: Statistics*. 2(2) : 273-281.
- [5] Sari, S. U. R. 2018. Aplikasi Metode Regresi Nonparametrik Spline Multivariabel untuk Pemodelan Faktor-Faktor yang Memengaruhi Indeks Pembangunan Gender di Provinsi Jawa Barat. *Jurnal Unimus*. 6(2) : 119-129.
- [6] Budiantara, I. N. 2000. Metode U, GML, CV dan GCV dalam Regresi Nonparametrik Spline. *Jurnal Majalan Ilmiah Himpunan Matematika Indonesia (MIHMI)*. 6(1) : 285-290.
- [7] Ruppert, D., Wand, M. P., dan Carrol, R. J. 2003. *Semiparametric Regression*, New York : Cambridge University Press.
- [8] Eubank, R. L. 1998. *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker, Inc.

- [9] Ruppert, D. 2002. Selecting The Number of Knots for Penalized Spline. *Journal of Computational and Graphical Statistics*. 11 : 735-757.
- [10] Eubank, R. L. 1999. *Nonparametric Regression and Spline Smoothing, Second Edition*. New York : Marcel Dekker, Inc.