

Regresi Nonparametrik *Spline* untuk Memodelkan Faktor-faktor yang Memengaruhi Indeks Pembangunan Gender (IPG) di Jawa Barat Tahun 2020

Nisrina Fajriati Rahayu*, Lisnur Wachidah

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*nisrinafr99@gmail.com, wachidah.lisnur07@gmail.com

Abstract. Regression analysis is a statistical method used to determine the pattern of the relationship between the independent variable and the dependent variable. There are three kinds of regression analysis, namely parametric regression analysis, semiparametric regression analysis and nonparametric regression analysis. Parametric regression analysis can be used when the assumptions are met but not all data can meet the parametric assumptions, an alternative to parametric regression is nonparametric regression because its use does not require strict assumptions. Spline nonparametric regression is a method used to get the estimated regression curve through the estimation of the data pattern according to the movement. The selection of the best model for Spline regression is seen from the Generalized Cross Validation (GCV) value using the minimum knot point. In this study, the dependent variable used is the Gender Development Index (GDI) in West Java Province in 2020 which consists of 18 districts and 9 cities with the independent variables consisting of the average length of schooling for women, the expected length of schooling for women, the open unemployment rate for women, female labor force participation rate, women with health complaints and sex ratio. The results of the analysis obtained that the best nonparametric Spline regression model was using the order of one and three knot points with the minimum GCV value of 0.2471, and the coefficient of determination was 99.98%. The six independent variables used have a significant influence on GPA in West Java in 2020.

Keywords: *Nonparametric Spline Regression, Gender Development Index (GDI), Knot Point, Generalized Cross Validation (GCV).*

Abstrak. Analisis regresi adalah metode statistika yang digunakan untuk menentukan pola hubungan antara variabel bebas dengan variabel terikat. Terdapat tiga macam analisis regresi, yaitu analisis regresi parametrik, analisis regresi semiparametrik dan analisis regresi nonparametrik. Analisis regresi parametrik dapat digunakan ketika asumsi terpenuhi akan tetapi tidak semua data dapat memenuhi asumsi parametrik, alternatif dari regresi parametrik adalah regresi nonparametrik karena penggunaannya tidak memerlukan asumsi yang ketat. Regresi nonparametrik *Spline* merupakan metode yang digunakan untuk mendapatkan dugaan kurva regresi melalui pendekatan estimasi pola data sesuai pergerakannya. Pemilihan model terbaik pada regresi *Spline* dilihat dari nilai *Generalized Cross Validation* (GCV) dengan menggunakan titik knot yang paling minimum. Pada penelitian ini variabel terikat yang digunakan adalah Indeks Pembangunan Gender (IPG) di Provinsi Jawa Barat Tahun 2020 yang terdiri dari 18 kabupaten dan 9 kota dengan variabel bebas yang terdiri dari rata-rata lama sekolah perempuan, harapan lama sekolah perempuan, tingkat pengangguran terbuka perempuan, tingkat partisipasi angkatan kerja perempuan, perempuan yang memiliki keluhan kesehatan dan rasio jenis kelamin. Hasil dari analisis diperoleh model regresi nonparametrik *Spline* yang terbaik adalah dengan menggunakan orde satu dan tiga titik knot dengan nilai GCV yang paling minimum 0,2471, serta didapatkan nilai koefisien determinasi sebesar 99,98%. Dengan ke enam variabel bebas yang digunakan memiliki pengaruh yang signifikan terhadap IPG di Jawa Barat tahun 2020.

Kata Kunci: *Regresi Nonparametrik Spline, Indeks Pembangunan Gender (IPG), Titik Knot, Generalized Cross Validation (GCV).*

A. Pendahuluan

Analisis regresi merupakan uji statistik untuk mengetahui pengaruh antara variabel bebas dengan variabel terikat [1]. Terdapat tiga macam analisis regresi diantaranya analisis regresi parametrik, regresi semiparametrik, dan regresi nonparametrik [2]. Ketika asumsi dalam analisis regresi parametrik terpenuhi maka analisis tersebut bisa digunakan, misalnya asumsi bentuk kurva segresi yang harus diketahui. Selain asumsi bentuk kurva yang harus dipenuhi, regresi parametrik harus mempunyai sifat *Best Linier Unbiased Estimator* (BLUE) yaitu data berdistribusi normal, nonautokorelasi, nonmultikolinearitas dan homoskedastisitas. Tetapi tidak semua data yang akan dijadikan penelitian memenuhi semua asumsi regresi parametrik.

Regresi nonparametrik digunakan sebagai alternatif regresi parametrik. Hal ini karena regresi tersebut tidak memerlukan asumsi yang ketat. Regresi nonparametrik adalah analisis untuk mengetahui hubungan antara variabel terikat dengan variabel bebas tanpa harus diketahui bentuk kurva regresi dan bentuk fungsinya. Fungsi dari regresi nonparametrik diasumsikan mulus (*smooth*) atau termuat dalam suatu ruang fungsi tertentu. Regresi nonparametrik merupakan analisis regresi yang sangat fleksibel dalam memodelkan pola data [3]. Terdapat berbagai macam regresi nonparametrik yaitu *kernel*, *Spline*, deret *fourier*, *wavelets*, tetapi pada penelitian skripsi ini akan menggunakan regresi nonparametrik *Spline*.

Menurut [4] regresi nonparametrik *Spline* merupakan analisis yang digunakan untuk memperoleh dugaan kurva regresi dengan mengestimasi pola data sesuai pergerakannya. Metode nonparametrik *Spline* adalah model polinom terpotong yang memiliki sifat tersegmentasi pada orde k dan terbentuk pada titik-titik knot. Titik knot sendiri merupakan titik yang menentukan terjadinya perubahan pola data atau disebut titik perpaduan.

Titik knot yang optimal dilihat dari hasil *Generalized Cross Validation* (GCV) yang paling minimum. Nilai GCV minimum digunakan juga dalam pemilihan model terbaik regresi *Spline*. *Mean Square Error* (MSE) kecil diakibatkan karena nilai GCV tersebut juga minimum. Analisis regresi nonparameterik *Spline* banyak digunakan karena memiliki kelebihan yaitu model memiliki kecenderungan untuk mendapatkan estimasi data kemanapun knot bergerak dengan adanya kecenderungan tersebut, regresi nonparameterik *Spline* menawarkan fleksibilitas yang tinggi, sehingga dapat menyesuaikan diri terhadap karakteristik lokal data [5].

IPG menurut [6] merupakan indeks yang mengamati kesetaraan *Gender* antara laki-laki dan perempuan yang diukur berdasarkan pencapaian pembangunan dasar manusia. Data yang dirilis BPS menyebutkan sebesar 91,06 nilai IPG Indonesia pada tahun 2020, Provinsi Jawa Barat memiliki nilai IPG sebesar 89,20 yang merupakan salah satu provinsi dengan nilai IPG dibawah IPG Nasional. Sementara itu Jawa Barat adalah provinsi yang memiliki penduduk dengan jumlah terbesar di Indonesia.

Berdasarkan latar belakang yang telah diuraikan, maka tujuan dari penelitian ini adalah:

1. Untuk mengetahui model Indeks Pembangunan *Gender* (IPG) di Provinsi Jawa Barat Tahun 2020 dengan analisis regresi *Spline*.
2. Untuk mengetahui faktor yang memiliki pengaruh kepada variabel Indeks Pembangunan *Gender* (IPG) di Provinsi Jawa Barat Tahun 2020.

B. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data Indeks Pembangunan *Gender* (IPG) Tahun 2020 yang bersumber dari BPS Jawa Barat. Obyek penelitian ini berupa 27 Kota/Kabupaten di Provinsi Jawa Barat, terdiri dari 18 kabupaten dan 9 kota dan enam variabel bebas yaitu Rata-rata Lama Sekolah (RLS) perempuan Harapan Lama Sekolah (HLS) perempuan, Tingkat Pengangguran Terbuka (TPT) perempuan, Tingkat Partisipasi Angkatan Kerja (TPAK) perempuan, Perempuan yang memiliki keluhan kesehatan dan Rasio jenis kelamin.

Metode dan prosedur pengolahan pada data yang digunakan yaitu sebagai berikut:

1. Menjelaskan karakteristik IPG Jawa Barat serta variabel-variabel bebas yang mempengaruhi dengan analisis statistika deskriptif,
2. Memodelkan IPG Kota/Kabupaten di Jawa Barat menggunakan regresi nonparametrik *Spline*, yang memiliki tahapan yaitu:

- Membuat grafik *scatterplot* dari setiap variabel bebas terhadap variabel terikat untuk mengetahui bentuk pola hubungan.
- Memodelkan variabel terikat dan enam variabel bebas dengan regresi nonparametrik *Spline* berbagai titik knot.
- Menentukan titik knot yang optimal berdasar pada nilai minimum dari GCV.
- Mendapatkan model regresi terbaik didasarkan pada titik knot yang optimal yang telah ditentukan pada langkah c.
- Pengujian signifikansi secara simultan dan parsial terhadap parameter.
- Pengujian asumsi residual untuk memenuhi asumsi IIDN (Identik, Independen dan Distribusi Normal). Jika pengujian asumsi residual tidak terpenuhi maka dilakukan transformasi, setelah itu lakukan kembali langkah a-f.
- Menghitung koefisien determinasi (R^2).
- Interpretasi hasil analisis dan penarikan kesimpulan.

C. Hasil Penelitian dan Pembahasan

Analisis Statistik Deskriptif

Analisis statistik deskriptif merupakan langkah pertama yang dilakukan sebelum melakukan proses pengolahan data dengan analisis regresi nonparametrik *Spline*. Statistik deskriptif yang digunakan pada penelitian ini yaitu rata-rata, varians, nilai minimum dan nilai maksimum dari variabel.

Tabel 1. Statistik Deskriptif IPG Jawa Barat dan Faktor yang Mempengaruhinya

Variabel	Rata-rata	Varians	Minimum	Maksimum
x_1	8,13	16,36	5,53	10,92
x_2	12,86	2,16	11,64	14,45
x_3	9,31	0,67	4,72	14,30
x_4	47,98	7,36	36,58	63,85
x_5	34,99	41,10	23,13	50,95
x_6	101,71	33,55	97,27	106,16
y	88,96	5,43	79,06	95,18

Berdasarkan Tabel 4.1 didapatkan hasil sebagai berikut:

Untuk variabel x_1 merupakan variabel Rata-rata Lama Sekolah (RLS) Perempuan di Jawa Barat tahun 2020 dengan nilai rata-rata sebesar 8,13 tahun. Kabupaten Indramayu menjadi kabupaten dengan nilai RLS yang paling kecil yaitu hanya sebesar 5,53 atau 6 tahun, sedangkan RLS terbesar yaitu Kota Bekasi dengan nilai sebesar 10,92 atau 11 tahun.

Pada variabel x_2 yang merupakan variabel Harapan Lama Sekolah (HLS) Perempuan di Jawa Barat Tahun 2020 dengan nilai rata-rata sebesar 12,86 tahun. Nilai varians sebesar 2,16. Kabupaten Cianjur memiliki nilai HLS terkecil yaitu 11,64 tahun, sebaliknya Kota Bandung menjadi kota dengan nilai HLS perempuan di Jawa Barat terbesar yaitu 14,45 tahun.

Untuk variabel x_3 atau variabel Tingkat Pengangguran Terbuka (TPT) Perempuan di Jawa Barat mewakili aspek ekonomi yang memiliki rata-rata 9,31%. Angka TPT tertinggi yaitu sebesar 14,30% yang dimiliki oleh Kota Bogor. Sedangkan angka TPT terendah yaitu 4,72% berada pada Kabupaten Pangandaran.

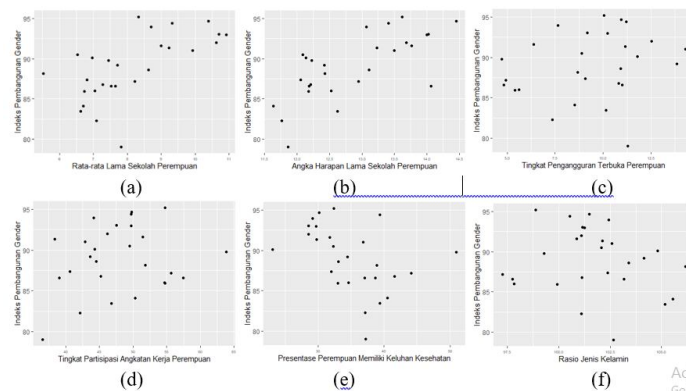
Variabel x_4 atau variabel Tingkat Partisipasi Angkatan Kerja (TPAK) Perempuan di Jawa Barat yang memiliki rata-rata 47,98%. Angka TPAK tertinggi yaitu sebesar 68,35% yang dimiliki oleh Kabupaten Pangandaran. Sedangkan angka TPAK terendah yaitu 36,58% berada pada Kabupaten Bandung Barat.

Variabel x_5 yaitu variabel persentase perempuan yang memiliki keluhan kesehatan yang dengan nilai rata-rata sebesar 34,99%. Kabupaten Pangandaran merupakan kota tertinggi yang memiliki persentase keluhan penduduk perempuan yaitu sebesar 50,95%, sedangkan Kabupaten Karawang memiliki persentase keluhan kesehatan penduduk perempuan terkecil yaitu sebesar 23,13%.

Pada variabel x_6 merupakan variabel rasio jenis kelamin yang mewakili orientasi *Gender* dengan rata-rata 101,71%. Nilai rasio jenis kelamin yang paling tinggi sebesar 106,16% yaitu Kabupaten Indramayu sedangkan nilai rasio jenis kelamin yang paling kecil sebesar 97,27% yaitu Kota Banjar.

Variabel y IPG Jawa Barat tahun 2020 memiliki nilai rata-rata sebesar 88,96. Nilai IPG tertinggi di Jawa Barat sebesar 95,18 yaitu Kabupaten Sumedang dan sebaliknya nilai IPG terendah yaitu Kabupaten Bandung Barat dengan nilai sebesar 79,06. Jika dilihat dari nilai varians sebesar 5,43, angka ini cukup kecil yang artinya IPG di Jawa Barat cukup merata pada tahun 2020.

Scatterplot Pola Hubungan Indeks Pembangunan Gender (IPG) dengan Faktor yang Mempengaruhi



Gambar 1. Scatterplot antara RLS Perempuan (a), HLS Perempuan (b), TPT Perempuan (c), TPAK Perempuan (d), Presentase Perempuan Memiliki Keluhan Kesehatan (e) dan Rasio Jenis Kelamin(f) terhadap IPG Jawa Barat tahun 2020.

Dari hasil grafik *scatterplot* untuk melihat pola hubungan pada setiap variabel bebas yang digunakan terhadap variabel terikat, dua variabel menunjukkan adanya perubahan pola data pada interval tertentu dan sisanya tidak menunjukkan adanya kecenderungan terhadap pola-pola kurva regresi tertentu.

Pemilihan Titik Knot Optimal dan GCV Minimum

Tabel 2. Titik Knot Optimal dan GCV Minimum

Nilai GCV Minimum	Titik Knot Optimal
8,5196	Satu Knot
8,0602	Dua Knot
0,2471	Tiga Knot
0,2471	Kombinasi Knot

Dari hasil Tabel 4.6 dengan menggunakan *software* R 3.5.3 didapatkan nilai GCV yang minimum terdapat pada tiga knot dan kombinasi knot dengan nilai GCV yang sama yaitu 0,2471, maka model terbaik dari regresi nonparametrik *Spline* untuk data IPG di Jawa Barat menggunakan tiga knot dengan titik knot optimal untuk enam variabel bebas yaitu pada variabel RLS perempuan (x_1) ada pada titik knot 6,4283, 7,3267 dan 9,5725; variabel HLS perempuan (x_2) ada pada titik knot 12,1083, 12,5767 dan 13,7475; variabel TPT perempuan (x_3) ada pada titik knot 6,3169, 7,9133 dan 11,9050. Variabel (x_4) atau TPAK perempuan ada pada titik knot 41,1250, 45,6700 dan 57,0325; variabel (x_5) atau variabel perempuan memiliki keluhan kesehatan ada pada titik knot 27,7667, 32,4033 dan 43,9950 serta variabel terakhir yaitu variabel rasio jenis kelamin (x_6) ada pada titik knot 98,7517, 100,2333 dan 103,9375. Model regresi nonparametrik *Spline* terbaik untuk dilakukan estimasi parameter sebagai berikut:

$$\begin{aligned}\hat{y} = & -0,0663 - 0,2729x_1 + 1,5173(x_1 - 6,4283)_+^1 + 0,8452(x_1 - 7,3267)_+^1 \\ & - 7,0924(x_1 - 9,5725)_+^1 + 10,9489x_2 - 14,1262(x_2 - 12,1083)_+^1 \\ & + 2,4784(x_2 - 12,5767)_+^1 + 0,2732(x_2 - 13,7475)_+^1 + 6,6368x_3 \\ & - 3,7926(x_3 - 6,3167)_+^1 - 2,6738(x_3 - 7,9133)_+^1 \\ & + 0,3349(x_3 - 11,9050)_+^1 + 2,0829x_4 - 2,7577(x_4 - 41,1250)_+^1 \\ & + 2,2230(x_4 - 45,67)_+^1 + 2,4681(x_4 - 57,0325)_+^1 + 1,0902x_5 \\ & - 3,3236(x_5 - 27,7667)_+^1 + 2,4128(x_5 - 32,4033)_+^1 \\ & - 3,6079(x_5 - 43,9950)_+^1 - 2,0269x_6 + 4,2637(x_6 - 98,7517)_+^1 \\ & - 1,5105(x_6 - 100,2333)_+^1 - 4,0880(x_6 - 103,9375)_+^1 + \varepsilon_i\end{aligned}$$

Pengujian Parameter Model Regresi Nonparametri *Spline*

Pengujian parameter model dilakukan untuk mengetahui apakah model suatu variabel bebas mempengaruhi signifikan terhadap variabel terikat secara simultan maupun parsial.

Uji parameter simultan dilakukan untuk mengetahui signifikansi model secara bersamaan. Hipotesis untuk pengujian parameter simultan yaitu:

$H_0 : \beta_{11} = \beta_{12} = \dots = \beta_{64} = 0$; (variabel bebas secara bersama-sama tidak berperan nyata terhadap variabel respon)

$H_1 : \text{minimal ada satu } \beta_{jl} \neq 0, j = 1, 2, \dots, 6; l = 1, 2, 3, 4$; (variabel bebas secara bersama-sama berperan nyata terhadap variabel respon)

Didapatkan hasil pengujian simultan sebagai berikut pada Tabel 3.

Tabel 3. ANOVA Regresi Nonparametrik *Spline*

Sumber Variasi	Derajat Bebas (db)	Jumlah Kuadrat (JK)	Kuadrat Tengah (KT)	F-hitung	p-value
Regresi	24	425,331	17,7221	430,3021	0,0023
Error	2	0,082	0,0412		
Total	26	425,414			

Dari Tabel 4.7 diperoleh *p-value* sebesar 0,0023 dan dengan menggunakan taraf signifikan $\alpha = 0,05$ dan karena nilai *p-value* $< \alpha$ maka didapatkan kesimpulan bahwa keputusan tolak H_0 . Artinya variabel bebas secara bersama-sama berperan nyata terhadap variabel respon.

Setelah keputusan dari uji simultan didapatkan bahwa seluruh variabel bebas secara bersama-sama memiliki pengaruh terhadap variabel respon maka akan dilanjutkan dengan uji parsial. Hipotesis uji parsial sebagai berikut.

$H_0 : \beta_{jl} = 0$ (tidak ada pengaruh signifikan variabel bebas terhadap variabel respon)

$H_1 : \beta_{jl} \neq 0, j = 1, 2, \dots, 6; l = 1, 2, 3, 4$ (terdapat pengaruh signifikan variabel bebas terhadap variabel respon)

Didapatkan hasil pengujian parsial sebagai berikut pada Tabel 4.

Tabel 4. Hasil Uji Parsial

Variabel	Beta	t-hitung	p-value	Keputusan
x_1	β_{11}	-0,3304	0,7437	Terima H_0
	β_{12}	2,8480	0,0085	Tolak H_0
	β_{13}	1,3873	0,1771	Terima H_0
	β_{14}	-14,7575	3,7719e-14	Tolak H_0
x_2	β_{21}	8,0163	1,7027e-8	Tolak H_0
	β_{22}	-5,6849	5,5705e-6	Tolak H_0
	β_{23}	2,0873	0,0468	Tolak H_0
	β_{24}	0,4008	0,6918	Terima H_0
x_3	β_{31}	12,2636	2,5856e-12	Tolak H_0
	β_{32}	-4,8079	5,5802e-5	Tolak H_0

	β_{33}	-6,5766	5,6579e-7	Tolak H_0
	β_{34}	1,4797	0,1510	Terima H_0
x_4	β_{41}	19,6790	3,8644e-17	Tolak H_0
	β_{42}	-15,8141	7,4344e-15	Tolak H_0
	β_{43}	12,6497	1,2913e-12	Tolak H_0
	β_{44}	1,9540	0,0615	Terima H_0
x_5	β_{51}	8,2300	1,0356e-8	Tolak H_0
	β_{52}	-12,0404	3,8903e-12	Tolak H_0
	β_{53}	10,0430	1,9387e-10	Tolak H_0
	β_{54}	-3,0118	0,0056	Tolak H_0
x_6	β_{61}	-10,5483	6,8972e-11	Tolak H_0
	β_{62}	12,2716	2,5485e-12	Tolak H_0
	β_{63}	-3,8888	0,0006	Tolak H_0
	β_{64}	-3,7254	0,0010	Tolak H_0

Dengan menggunakan taraf signifikan $\alpha = 0.05$ dan dari Tabel 4. diperoleh nilai t -hitung dan p -value untuk seluruh parameter yaitu sebanyak 24 parameter, terdapat 5 parameter yang menghasilkan keputusan terima H_0 atau tidak signifikan yang artinya parameter tidak berpengaruh signifikan terhadap variabel respon karena memiliki nilai p -value yang lebih besar dari α yaitu pada parameter $\beta_{11}, \beta_{13}, \beta_{24}, \beta_{34}$ dan β_{44} . Tetapi walau terdapat 5 parameter yang tidak signifikan, secara umum keenam variabel bebas memiliki pengaruh terhadap model karena masih ada parameter yang memiliki kesimpulan signifikan pada masing-masing variabel.

Pengujian Asumsi Residual

Setelah dilakukan pengujian parameter baik secara simultan maupun secara parsial maka selanjutnya dilakukan pengujian asumsi residual untuk mengetahui apakah residual memenuhi asumsi IIDN (Identik, Independen, Distribusi Normal).

Uji asumsi identik dilakukan untuk melihat homogenitas dari residual. Salah satu cara untuk mendeteksi heterokedastisitas dengan uji *Glejser*. Hipotesis uji *Glejser* sebagai berikut.

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_{27}^2 = \sigma^2$ (asumsi residual memiliki kasus homoskedastisitas)

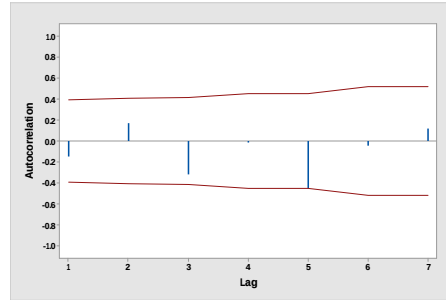
H_1 : minimal terdapat satu $\sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, 27$ (asumsi residual tidak memiliki kasus homoskedastisitas)

Tabel 5. Hasil Uji *Glejser*

Sumber Variasi	Derajat Bebas (db)	Jumlah Kuadrat (JK)	Kuadrat Tengah (KT)	F -hitung	p -value
Regresi	24	0,03	0,0012	0,9708	0,6278
Error	2	0,002	0,013		
Total	26	0,032			

Dengan menggunakan persamaan (2.23) dan bantuan *software* R didapatkan hasil uji *Glejser* yang disajikan pada Tabel 4.9 dengan nilai p -value sebesar 0,6278 dan dengan menggunakan $\alpha = 0,05$. Karena nilai p -value yang didapat lebih besar dari α , sehingga disimpulkan keputusan Terima H_0 yang artinya asumsi residual memiliki kasus homoskedastisitas atau asumsi identik terpenuhi.

Setelah asumsi identik terpenuhi selanjutnya asumsi independen. Asumsi independen digunakan untuk mendeteksi adanya autokorelasi antar residual atau nilai dugaan tidak berhubungan dengan nilai dugaan lainnya. Pada Gambar 4.8 diperoleh hasil dari plot ACF dengan menggunakan *software* Minitab 17 yang menunjukkan bahwa pada lag berapapun tidak ada nilai yang keluar dari batas bawah maupun batas atas interval konfidensi maka tidak terjadi autokorelasi antar residual.



Gambar 2. Plot ACF Residual

Untuk memperkuat hasil dari plot ACF dilanjutkan dengan uji *DW*. Hipotesis untuk uji *Durbin Watson* yaitu:

$H_0 : \rho = 0$ (tidak ada otokorelasi)

$H_1 : \rho \neq 0$ (terdapat otokorelasi)

Dengan persamaan (2.24) statistik uji *DW* serta bantuan *software R*, didapatkan hasil perhitungan pada Tabel 6.

Tabel 6. Hasil Uji *Durbin Watson*

<i>DW</i> -hitung	<i>p-value</i>
2,5620	0,204

Pada Tabel 4.10 disajikan hasil uji *Durbin Watson* dengan nilai *p-value* sebesar 0,204 dan dengan taraf signifikan yang digunakan $\alpha = 0,05$, maka keputusan terima H_0 karena *p-value* lebih besar dari α . Artinya tidak terdapat autokorelasi pada residual atau asumsi independen terpenuhi.

Asumsi residual berdistribusi normal adalah asumsi terakhir yang harus dipenuhi residual setelah asumsi identik dan asumsi independen. Pengujian asumsi distribusi normal dilakukan dengan uji *Kolmogorov Smirnov*, berikut hipotesis uji *Kolmogorov Smirnov*.

$H_0 : F_0(x) = F(x)$ (asumsi residual mengikuti distribusi normal)

$H_1 : F_0(x) \neq F(x)$ (asumsi residual tidak mengikuti distribusi normal)

Tabel 7. Hasil Uji *Kolmogorov Smirnov*

<i>D</i> -hitung	<i>p-value</i>
0,0897	0,9680

Berdasarkan statistik uji pada persamaan (2.25) dan dengan *software R* diperoleh hasil uji pada Tabel 4.11 didapatkan nilai *p-value* 0,9680. Dari taraf signifikan yang digunakan $\alpha = 0,05$ maka diambil keputusan bahwa H_0 diterima karena *p-value* lebih besar dari α , artinya asumsi residual mengikuti distribusi normal atau asumsi distribusi normal terpenuhi.

Koefisien Determinasi (R^2)

$$R^2 = \frac{JK_{regresi}}{JK_{total}} = \frac{425.331}{425.414} = 0.999804$$

Dengan menggunakan persamaan pada (2.26) dan perhitungan yang telah dilakukan dengan *software R*, didapatkan nilai R^2 sebesar 0,9998 atau 99,98% yang memiliki arti bahwa sebesar 99,98% model IPG di Jawa Barat dapat dijelaskan menggunakan enam variabel bebas dan tiga titik knot yang digunakan.

Interpretasi Model Regresi Nonparametrik *Spline*

Setelah dilakukan pengujian terhadap parameter dan pengujian asumsi residual yang telah terpenuhi, maka dilakukan interpretasi model regresi nonparametrik *Spline* terbaik dari nilai GCV minimum yang telah diperoleh sebelumnya. Berikut interpretasi model regresi

nonparametrik *Spline* untuk setiap variabel.

Pengaruh RLS perempuan (x_1) terhadap IPG di Jawa Barat dengan variabel lainnya (x_2, x_3, x_4, x_5, x_6) dianggap konstan yaitu sebagai berikut.

$$\hat{y} = -0,0663 - 0,2729x_1 + 1,5173(x_1 - 6,4283)_+^1 + 0,8452(x_1 - 7,3267)_+^1 - 7,0924(x_1 - 9,5725)_+^1$$

$$\hat{y} = \begin{cases} -0,0663 - 0,2729x_1; & x_1 < 6,4283 \\ -9,8199 + 1,2444x_1; & 6,4283 \leq x_1 < 7,3267 \\ -16,0142 + 2,0896; & 7,3267 \leq x_1 < 9,5725 \\ 51,8779 - 5,0028x_1; & x_1 \geq 9,5725 \end{cases}$$

Dengan membandingkan nilai RLS pada data di Lampiran 1 dan didapatkan interval pertama ketika RLS perempuan di kota/kabupaten di Jawa Barat kurang dari 6,4283 tahun dan mengalami kenaikan satu satuan serta variabel lain dianggap konstan, maka IPG di Jawa Barat cenderung turun sebesar 0,2729. Terdapat satu daerah yang termasuk pada interval ini yaitu Kabupaten Indramayu. Lalu pada interval kedua jika RLS perempuan di kota/Kabupaten di Jawa Barat memiliki nilai antara 6,4283 tahun hingga 7,3267 tahun dan mengalami kenaikan satu satuan serta variabel dianggap konstan, maka IPG di Jawa Barat akan naik 1,2444. Terdapat sembilan daerah yang berada pada interval kedua yaitu Kabupaten Subang, Kabupaten Cirebon, Kabupaten Cianjur, Kabupaten Majalengka, Kabupaten Sukabumi, Kabupaten Tasikmalaya, Kabupaten Karawang, Kabupaten Garut dan Kabupaten Kuningan. Jika RLS perempuan memiliki nilai antara 7,3267 tahun hingga 9,5725 tahun dan mengalami pertambahan satu satuan serta variabel lain diasumsikan konstan maka IPG di Jawa Barat akan naik sebesar 2,0896. Pada interval ketiga terdapat 12 daerah yang termasuk seperti Kabupaten Pangandaran, Kabupaten Ciamis, Kabupaten Purwakarta, Kabupaten Bogor, Kabupaten Bandung Barat, Kota Banjar, Kabupaten Sumedang, Kota Bekasi, Kabupaten Bandung, Kota Tasikmalaya, Kota Sukabumi dan Kota Cirebon. Pada interval terakhir jika RLS perempuan lebih dari 9,5725 tahun dan mengalami kenaikan satu satuan serta variabel lain diasumsikan konstan, maka IPG di Jawa Barat cenderung turun sebesar 5,0028. Daerah yang termasuk pada interval terkakhir ini adalah Kota Bogor, Kota Bandung, Kota Cimahi, Kota Depok dan Kota Bekasi. Variabel lainnya dapat diinterpretasikan dengan cara yang sama dengan x_1 .

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Model regresi nonparametrik *Spline* yang terbaik untuk memodelkan IPG di Provinsi Jawa Barat tahun 2020 adalah dengan menggunakan tiga titik knot. Nilai GCV minimum yang diperoleh 0,2471 dan nilai R^2 sebesar 99,98% serta semua asumsi residual yang telah terpenuhi. Dengan model regresi nonparametrik *Spline* yaitu:

$$\hat{y} = -0,0663 - 0,2729x_1 + 1,5173(x_1 - 6,4283)_+^1 + 0,8452(x_1 - 7,3267)_+^1 - 7,0924(x_1 - 9,5725)_+^1 + 10,9489x_2 - 14,1262(x_2 - 12,1083)_+^1 + 2,4784(x_2 - 12,5767)_+^1 + 0,2732(x_2 - 13,7475)_+^1 + 6,6368x_3 - 3,7926(x_3 - 6,3167)_+^1 - 2,6738(x_3 - 7,9133)_+^1 + 0,3349(x_3 - 11,9050)_+^1 + 2,0829x_4 - 2,7577(x_4 - 41,1250)_+^1 + 2,2230(x_4 - 45,67)_+^1 + 2,4681(x_4 - 57,0325)_+^1 + 1,0902x_5 - 3,3236(x_5 - 27,7667)_+^1 + 2,4128(x_5 - 32,4033)_+^1 - 3,6079(x_5 - 43,9950)_+^1 - 2,0269x_6 + 4,2637(x_6 - 98,7517)_+^1 - 1,5105(x_6 - 100,2333)_+^1 - 4,0880(x_6 - 103,9375)_+^1$$

2. Pada analisis ini terdapat enam variabel yang digunakan 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. Dari ke enam variabel memiliki pengaruh yang signifikan terhadap IPG di Jawa Barat pada tahun 2020. Dengan nilai rata-rata IPG di Jawa Barat tahun 2020 sebesar 88,89. Kabupaten Sumedang menjadi yang terbesar memiliki nilai IPG yaitu 95,18 dan nilai IPG terendah yaitu Kabupaten Bandung Barat dengan nilai sebesar 79,06.

Acknowledge

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah banyak membantu dalam pelaksanaan penelitian ini, khususnya kepada dosen pembimbing Ibu Dr. Lisnur Wachidah, Dra., M.Si., yang selalu memberikan waktunya untuk membantu dan memberikan saran kepada penulis serta para dosen Statistika Unisba yang sudah memberikan ilmu pengetahuan kepada penulis. Terima kasih juga kepada rekan-rekan yang membantu bertukar pikiran hingga penelitian ini dapat terselesaikan dengan baik.

Daftar Pustaka

- [1] Budiantara, I. N. (2000). Metode U, GML, CV dan GCV dalam Regresi Nonparametrik *Spline*. *Jurnal Majalan Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 6, No. 1, 285-290.
- [2] Qudratullah, M. F. (2013). Analisis Regresi Terapan Teori, Contoh Kasus dan Aplikasi dengan SPSS. Yogyakarta: Andi Offset.
- [3] Eubank, R. L. (1988). Nonparametric Regression and *Spline* Smoothing Second.
- [4] Wulandari, H., & dkk. (2017). Penerapan Analisis Regresi *Spline* Untuk Menduga Harga Cabai di Jakarta. *Indonesian Journal of Statistics and Its Applications*, Vol. 1 No. 1, 1-12.
- [5] Budiantara, I. N. (2009). *Spline* dalam Regresi Nonparametrik dan Semiparametrik. Pidato Pengukuhan Untuk Jabatan Guru Besar pada Jurusan Statistika, Institut Teknologi Sepuluh Nopember (ITS). Surabaya.
- [6] BPS Jawa Barat. (2020). Indeks Pembangunan *Gender* Provinsi Jawa Barat. Bandung: Badan Pusat Statistik Provinsi Jawa Barat
- [7] Khoeriyah, Risti Yulianti. (2021). *Regresi Terboboti Geografis Semiparametrik (RTG-S) untuk Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Sumatera Utara*, Jurnal Riset Statistika, 1(1), 43-50.