BuletinIlmiah Mat. Stat. danTerapannya (Bimaster)
Volume 08, No. 2(2019), hal 175-184.

ESTIMASI PARAMETER REGRESI *SPLINE* DENGAN METODE *PENALIZED SPLINE*

Wahyu Kurniasari, Dadan Kusnandar, Evy Sulistianingsih

INTISARI

Regresi spline merupakan suatu pendekatan ke arah pencocokan data dengan tetap memperhitungkan kemulusan kurva. Salah satu bentuk estimator dari regresi spline ialah penalized spline. Tujuan dari penelitian ini adalah untuk mengestimasi parameter regresi spline dengan metode penalized spline untuk data yang tidak memiliki pola tertentu. Data penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik Indonesia pada tahun 2015 yaitu indeks pembangunan manusia, gini rasio, harapan lama sekolah, penduduk miskin, dan kepadatan penduduk. Hasil regresi spline yang diperoleh untuk model terbaik yaitu model spline linier pada setiap variabel dengan nilai Generalized Cross Validation (GCV) minimum. Hasil penelitian menunjukkan bahwa regresi spline dengan metode penalized spline menghasilkan estimasi parameter yang signifikan dan memperoleh nilai koefisien determinasi terkoreksi (R^2_{adj}) *sebesar 76,66% serta nilai MAPE untuk model regresi spline sebesar 1,415%.*

Kata Kunci: regresi nonparametrik, regresi *spline, penalized spline.*

PENDAHULUAN

Regresi nonparametrik merupakan metode untuk mengetahui hubungan antar variabel prediktor dan variabel respon yang tidak diketahui bentuk kurva fungsi regresi dari data yang digunakan. Bentuk kurva dari regresi nonparametrik ini berupa kurva fungsi yang tidak berpola sehingga menyebabkan regresi nonparametrik memiliki fleksibilitas yang tinggi [1]. Salah satu regresi nonparametrik yang sering digunakan adalah regresi *spline*. Regresi *spline* merupakan suatu pendekatan pada plot data dengan memperhitungkan kemulusan kurva. Pada regresi *spline*, penentuan titik knot sangat penting diperhatikan untuk memperoleh model yang optimal. Dengan bantuan titik-titik knot, regresi *spline* dapat mengatasi pola data yang menunjukkan naik/turun yang tajam [2]. Salah satu metode yang dapat digunakan yakni *penalized spline*. *Penalized spline* merupakan salah satu jenis metode estimator dalam regresi *spline* yang diperoleh dengan meminimumkan fungsi *Penalized Least Square* [3].

Adapun tujuan yang dicapai dari penelitian ini adalah menentukan estimasi parameter untuk data yang tidak memiliki pola tertentu menggunakan regresi *spline* dengan metode *penalized spline*. Batasan masalah pada penelitian ini adalah data yang digunakan adalah indeks pembangunan manusia, angka harapan hidup, gini rasio, harapan lama sekolah, penduduk miskin dan kepadatan penduduk dari 31 Provinsi di Indonesia pada tahun 2015. Metode dalam penelitian ini adalah regresi *spline* dengan menggunakan metode *penalized spline*. Dalam pemilihan parameter pemulus menggunakan *Generalized Cross Validation* (GCV), kesesuaian model yang digunakan yakni koefisien determinasi terkoreksi (R^2_{adj}) dan *Mean Absolute Percentage Error* (MAPE).

Penelitian ini berupa studi literatur yang dimulai dengan mempelajari karakteristik data, menentukan orde *spline*, banyaknya titik knot, nilai λ dan letak titik knot. Selanjutnya ditentukan nilai λ berdasarkan nilai *Generalized Cross Validation* (GCV) minimum dan banyak titik knot optimal. Kemudian mengestimasi parameter model regresi dengan meminimumkan fungsi *penalized least square*, memperoleh estimator *spline*, dan memilih model regresi *spline* terbaik dengan metode *penalized spline*. Setelah itu menguji signifikansi parameter dan uji asumsi residual model regresi *spline*.

REGRESI NONPARAMETRIK

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui hubungan antara variabel prediktor dan variabel respon yang tidak diketahui bentuk kurva fungsi regresinya. Model umum regresi nonparametrik adalah sebagai berikut [4]:

$$y_i = f\left(x_i\right) + \varepsilon_i,\ \ i = 1, 2, \ldots, n \tag{1}$$

dimana y_i adalah variabel respon pengamatan ke-i; x_i adalah variabel prediktor pengamatan ke-i; $f\left(x_i\right)$ adalah fungsi regresi yang tidak diketahui; dan ε_i adalah pengukuran residual yang tidak dapat dijelaskan dengan fungsi regresi $f\left(x_i\right)$.

REGRESI *SPLINE*

Spline merupakan potongan polinomial tersegmen yang dihubungkan oleh titik-titik knot yang dapat menjelaskan karakteristik dari data. Knot merupakan titik fokus dalam fungsi *spline* sehingga kurva yang dibentuk tersegmen pada titik tersebut. Secara umum bentuk fungsi *spline* linier dengan r orde dan m titik knot adalah sebagai berikut [4]:

$$f\left(x_i\right) = \sum_{q=0}^{r} \beta_q x_i^{\,q} + \sum_{j=1}^{m} \beta_{(r+j)} (x_i - K_j)_+^r \tag{2}$$

dimana $f\left(x_i\right)$ merupakan fungsi dari variabel ke-i, β_q dan $\beta_{(r+j)}$ menunjukkan koefisen regresi *spline* dengan $q = 1, 2, ..., r$ dan $j = 1, 2, ..., m$ serta $(x_i - K_j)_+^r$ menunjukkan fungsi linier pada variabel ke-*i* dengan letak titik knot K_j pada orde ke-*r*.

METODE *PENALIZED SPLINE*

Metode *penalized spline* merupakan salah satu bentuk estimator dari regresi *spline*. *Penalized spline* diperoleh dengan meminimumkan fungsi estimasi dari *Penalized Least Square* (PLS). Fungsi PLS dapat dituliskan dalam persamaan matriks sebagai berikut:

$$PLS = (\boldsymbol{Y} - \boldsymbol{X\beta})^T (\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda^2 \boldsymbol{\beta}^T \boldsymbol{D\beta} \tag{3}$$

dimana λ merupakan parameter penghalus pengontrol keseimbangan antara kecocokan data dan kemulusan kurva [3]. Parameter λ memetakan data dimana $\lambda \geq 0$ dan $\boldsymbol{D}$ merupakan matriks diagonal simetrik dengan banyak diagonal tergantung pada jumlah titik knot yang diperoleh berikut ini:

$$\boldsymbol{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times K} \\ \boldsymbol{0}_{K\times 2} & \boldsymbol{I}_{K\times K} \end{bmatrix}$$

Sedangkan untuk memperoleh koefisien regresi $\boldsymbol{\beta}$ dari turunan pertama fungsi PLS sebagai berikut:

$$\boldsymbol{\beta} = (\boldsymbol{X}^T \boldsymbol{X} + \lambda^2 \boldsymbol{D})^{-1} \boldsymbol{X}^T \boldsymbol{Y} \tag{4}$$

PEMILIHAN PARAMETER SMOOTHING (λ) DAN ORDE OPTIMAL

Parameter *smoothing* (λ) merupakan pengontrol antara kemulusan fungsi kurva dan kesesuaian fungsi terhadap data sehingga dalam memilih nilai λ diharapkan nilai optimal. Nilai λ optimal diperoleh dengan menghitung nilai *Generalized Cross Validation* (GCV) [3]. Parameter yang optimal diperoleh berdasarkan nilai GCV minimum. Fungsi GCV dapat dinyatakan sebagai berikut:

$$GCV = \frac{MSE}{\left(1 - tr(\boldsymbol{A}_{\lambda})/n\right)^2} \tag{5}$$

dengan $tr\left(\boldsymbol{A}_{\lambda}\right) = tr\left[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{D})^{-1}\boldsymbol{X}^T\right]$ dan $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

Sedangkan untuk pemilihan orde optimal juga diambil berdasarkan nilai GCV minimum sehingga pemilihan orde dan nilai λ saling berkaitan.

PEMILIHAN JUMLAH KNOT (K) OPTIMAL

Jumlah knot merupakan banyaknya titik knot atau banyaknya titik dimana terjadi perubahan perilaku fungsi pada interval yang berbeda [4]. Pemilihan jumlah knot (K) juga berdasarkan dari nilai λ dan orde optimal. Algoritma yang digunakan untuk memilih jumlah knot (K) optimal adalah algoritma *Full Seacrh.* Dalam algoritma *Full Seacrh*, jumlah knot (K) dihitung semua hingga batas maksimal. Untuk maksimal jumlah knot (K) yang dihitung yakni $K < (n_{unique} - r - 1)$, sedangkan untuk titik knot (K) terletak pada sampel kuantil dari nilai *unique* (tunggal) dari variabel prediktor.

PENGUJIAN SIGNIFIKANSI PARAMETER MODEL REGRESI

Pengujian signifikansi parameter bertujuan untuk mengetahui pengaruh dari variabel prediktor terhadap variabel respon [6]. Pengujian dilakukan secara serentak (simultan) kemudian dilanjutkan dengan pengujian secara individu. Berikut ini tahapan dalam pengujian signifikansi parameter:

1. Uji serentak (Uji F)
   Hipotesis yang digunakan yakni:
   H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ (tidak terdapat pengaruh antara variabel prediktor terhadap variabel respon).
   H_1 : minimal ada satu $\beta_k \neq 0$ (terdapat pengaruh antara variabel prediktor terhadap variabel respon).
   Statistik uji yang digunakan untuk uji F sebagai berikut:

$$F_{hitung} = \frac{MS_{regresi}}{MS_{error}} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 \Big/ k}{\left(y_i - \hat{y}_i\right)^2 \Big/ \left(n - k - 1\right)}$$

   Kriteria pengambilan keputusan yakni tolak H_0 jika $F_{hitung} > F_{tabel}\,(F_{\alpha;(k-1,n-1)})$ atau *p-value* $< \alpha$.

2. Uji individu (Uji t)
   Hipotesis yang digunakan yakni:
   H_0: $\beta_i = 0$ (parameter tidak signifikan)
   H_1: $\beta_i \neq 0$ (parameter siginifikan)

Statistik uji yang digunakan untuk uji t sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_i}{SE_{\hat{\beta}_i}}$$

Kriteria pengambilan keputusan tolak H_0 jika $|t_{hitung}|$ lebih besar daripada $t_{tabel}(t_{\frac{\alpha}{2},n-k})$ atau $p-value < \alpha$.

PENGUJIAN ASUMSI RESIDUAL

Berikut ini asumsi residual yang harus dipenuhi dalam regresi *spline* adalah sebagai berikut [6]:

1. Asumsi Residual Independen

   Salah satu metode yang digunakan untuk mengetahui adanya korelasi antar residual. Uji yang digunakan ialah Durbin-Watson. Hipotesis yang digunakan yakni:

   H_0: Tidak terdapat korelasi antar residual

   H_1: Terdapat korelasi antar residual

   Berikut statistik uji yang digunakan [7].

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i-1}^{n} e_i^2} \tag{6}$$

   Kriteria pengujian yakni H_0 ditolak jika $0 \le DW \le dL$ atau $4 - dL \le DW \le 4$.

2. Asumsi Residual Identik

   Asumsi residual identik digunakan untuk melihat homogenitas dari variansi residual. Untuk mendeteksi adanya heterokedastisitas menggunakan uji *Glejser*. Hipotesis yang digunakan yakni:

   H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

   H_1: minimal ada satu $\sigma_i^2 \neq \sigma^2$

   Statistik uji dari uji *Glejser* adalah sebagai berikut:

$$F_{hitung} = \frac{\sum_{i=1}^{n}(|\hat{e}_i| - |\bar{e}_i|)^2 / _{k-1}}{\sum_{i=1}^{n}(|e_i| - |\hat{e}_i|) / _{n-k}} \tag{7}$$

   Kriteria pengambilan keputusan yakni tolak H_0 jika $F_{hitung} > F_{tabel}(F_{\alpha;(k-1,n-1)})$ atau *p-value* $< \alpha$.

3. Asumsi Normalitas Residual

   Uji normalitas residual dilakukan untuk melihat apakah residual mengikuti distribusi normal. Hipotesis yang digunakan yakni sebagai berikut:

   H_0 : Residual mengikuti distribusi normal

   H_1 : Residual tidak mengikuti distribusi normal

   Statistik uji yang digunakan yaitu uji *Kolmogorov-Smirnov* sebagai berikut:

$$z_{hitung} = Sup_x|F_n(x) - F_0(x)|$$

   Kriteria pengambilan keputusan yakni tolak H_0 jika $z_{hitung} > z_{\alpha}$ atau *p-value* $< \alpha$.

KESESUAIAN MODEL

Kesesuaian model yang digunakan dalam mengestimasi model regresi dapat dilihat berdasarkan nilai *Mean Absolute Percentage Error* (MAPE). Nilai MAPE dapat dihitung sebagai berikut:

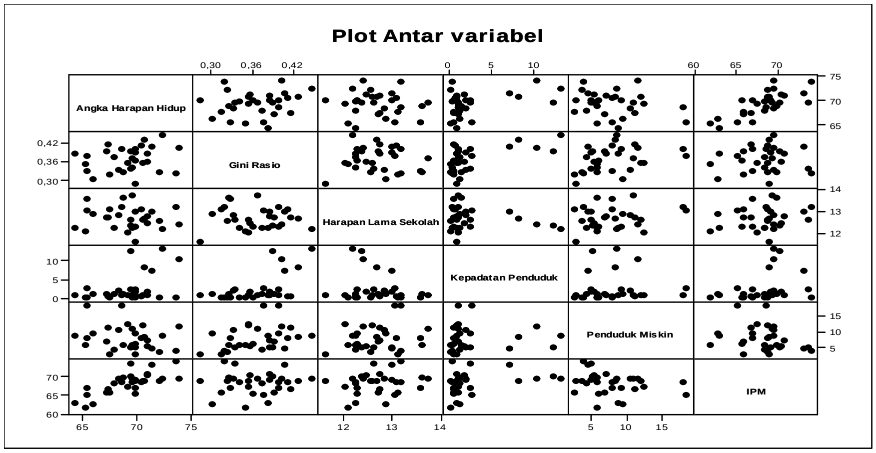
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - \bar{y}_i|}{y_i} \times 100\%$$

dengan kriteria penilaian nilai MAPE sebagai berikut:

No	Nilai MAPE	Kesimpulan
1.	MAPE< 10%	Sangat Baik
2.	10% ≤ MAPE < 20%	Baik
3.	20% ≤ MAPE < 50%	Cukup
4.	MAPE≥ 50%	Sangat Tidak Baik

METODE PENELITIAN

Data dalam penelitian ini merupakan data sekunder yang diambil dari Badan Pusat Statitik (BPS) Nasional tahun 2015. Penelitian ini menganalisis data dari 31 Provinsi di Indonesia dengan variabel-variabel yang digunakan dalam penelitian ini adalah variabel respon (Y) berupa indeks pembangunan manusia (%) menurut Provinsi di Indonesia. Kemudian terdapat pula 5 variabel prediktor (X) yaitu angka harapan hidup (X_1), gini rasio (X_2), harapan lama sekolah (X_3), penduduk miskin (X_4) dan kepadatan penduduk (X_5). Langkah awal dalam penentuan model hubungan antara indeks pembangunan manusia dengan masing-masing variabel yaitu membuat *scatter plot* antara variabel respon dan masing-masing variabel prediktor.



Gambar 1 *Scatterplot* Antarvariabel

Variabel-variabel tersebut juga tidak membentuk pola data yang spesifik. Hal ini menunjukkan dengan sebaran data yang cenderung acak, bahwa setiap variabel saling bebas.

MENENTUKAN ORDE, PARAMETER SMOOTHING (λ) DAN JUMLAH KNOT (K) OPTIMAL

Ada kriteria yang harus diperhatikan dalam membentuk model regresi *spline* ke arah *fitting* data yaitu menentukan orde untuk model, banyaknya titik knot (K) dan parameter *smoothing* (λ). Berikut nilai λ dan jumlah knot (K) optimal dari masing-masing variabel prediktor terhadap variabel respon.

Tabel 1 Nilai Optimal pada Regresi *Spline*

Variabel	**Nilai Optimal**				**Spline Terbaik**
	Orde	**Jumlah Knot**	**Parameter *Smoothing*** (λ)	**GCV**	
X_1	1	2	0,4	4,06402	*Spline* Linier
	2	1	0,33	4,33002	
X_2	1	1	0,9	9,29664	*Spline* Linier
	2	1	0,9	9,92530	
X_3	1	1	0,9	8,93831	*Spline* Linier
	2	1	0,9	9,57535	
X_4	1	1	0,9	8,83922	*Spline* Linier
	2	2	0,1	8,84770	
X_5	1	2	0,9	8,55086	*Spline* Linier
	2	1	0,9	8,75415	

Tabel 1 menunjukkan kriteria optimal yang diperoleh dengan menggunakan regresi *spline*. Diperoleh orde optimum dari setiap variabel yakni pada orde satu yang menunjukkan pola data yakni *spline* linier.

ESTIMASI PARAMETER REGRESI *SPLINE* DENGAN METODE *PENALIZED SPLINE*

Berdasarkan hasil pemilihan titik knot optimal yang dilakukan, maka nilai estimasi parameter yang diperoleh dengan metode *penalized spline* pada setiap variabel prediktor terhadap indeks pembangunan manusia. Untuk angka harapan hidup, berdasarkan Tabel 4.3 diperoleh informasi bahwa nilai GCV minimum sebesar 4,06402 terdapat pada orde (r_1) satu, jumlah knot (K_1) sebanyak dua titik yakni titik 68,5 dan 69,9 serta parameter *smoothing* (λ_1) sebesar 0,4. Sehingga diperoleh nilai $\boldsymbol{\beta}_1$ sebagai berikut

$$\boldsymbol{\beta}_1 = \begin{bmatrix} -13,2256 \\ 1,1884 \\ -0,4357 \\ -0,1443 \end{bmatrix}$$

Dengan demikian bentuk estimator $f(x_1)$ dari regresi *spline* yang diperoleh adalah:

$$f(x_1) = -13,2256 + 1,1884x_1 - 0,4357(x_1 - 68,5)_+ - 0,1443(x_1 - 69,9)_+$$

Atau dapat ditulis seperti berikut:

$$f(x_1) = \begin{cases} -13,2256 + 1,1884x_1, & x_1 < 68,5 \\ -15,9354 + 0,7527x_1, & 68,5 < x_1 < 69,9 \\ -5,8488 + 0,6384x_1, & x_1 \geq 69,9 \end{cases}$$

Untuk gini rasio, titik knot yang diperoleh yakni 0,371 dan parameter *smoothing* (λ_2) sebesar 0,9. Kemudian dihitung nilai estimasi $f(x_2)$ sehingga diperoleh nilai $\boldsymbol{\beta}_2$ sebagai berikut:

$$\boldsymbol{\beta}_2 = \begin{bmatrix} 66,67 \\ 4,502 \\ 0,004 \end{bmatrix}$$

Dengan demikian, bentuk estimator $f(x_2)$ dari regresi *spline* yang diperoleh adalah:

$$f(x_2) = 66,67 - 4,502x_2 + 0,004(x_2 - 0,371)_+$$

Atau dapat ditulis seperti berikut:

$$f(x_2) = \begin{cases} 66,67 - 4,502x_2, & x_2 < 0,371 \\ 66,6552 - 4,506x_2, & x_2 \geq 0,371 \end{cases}$$

Untuk harapan lama sekolah, titik knot yang diperoleh yakni 12,68 dan parameter *smoothing* (λ_3) sebesar 0,9. Kemudian dihitung nilai estimasi $f(x_3)$ sehingga diperoleh nilai $\boldsymbol{\beta}_3$ sebagai berikut:

$$\boldsymbol{\beta}_3 = \begin{bmatrix} 53,1942 \\ 1,1936 \\ -0,0003 \end{bmatrix}$$

Dengan demikian, bentuk estimator $f(x_3)$ dari regresi *spline* yang diperoleh adalah:

$$f(x_3) = 53,1942 + 1,1936x_3 - 0,0003(x_3 - 12,68)_+$$

Atau dapat ditulis seperti berikut:

$$f(x_3) = \begin{cases} 53,1942 + 1,1936x_3, & x_3 < 12,68 \\ 53,198 + 1,1933x_3, & x_3 \geq 12,68 \end{cases}$$

Untuk penduduk miskin, titik knot yang diperoleh yakni 7,44 dan parameter *smoothing* (λ_4) sebesar 0,9. Kemudian dihitung nilai estimasi $f(x_4)$ sehingga diperoleh nilai $\boldsymbol{\beta}_4$ sebagai berikut:

$$\boldsymbol{\beta}_4 = \begin{bmatrix} 69,7972 \\ -0,1892 \\ 0,0047 \end{bmatrix}$$

Dengan demikian, bentuk estimator $f(x_4)$ dari regresi *spline* yang diperoleh adalah:
$f(x_4) = 69,7972 - 0,1892x_4 + 0,0047(x_4 - 7,44)_+$

Atau dapat ditulis seperti berikut:

$$f(x_4) = \begin{cases} 69,7972 - 0,1892x_4, & x_4 < 7,44 \\ 69,7625 - 0,1845x_4, & x_4 \geq 7,44 \end{cases}$$

Untuk kepadatan penduduk, titik knot yang diperoleh yakni 0,78 dan 1,79 serta parameter *smoothing* (λ_5) sebesar 0,9. Kemudian, dihitung nilai estimasi $f(x_5)$ sehingga diperoleh nilai $\boldsymbol{\beta}_5$ sebagai berikut:

$$\boldsymbol{\beta}_5 = \begin{bmatrix} 67,7357 \\ 0,2361 \\ -0,0006 \\ -0,0024 \end{bmatrix}$$

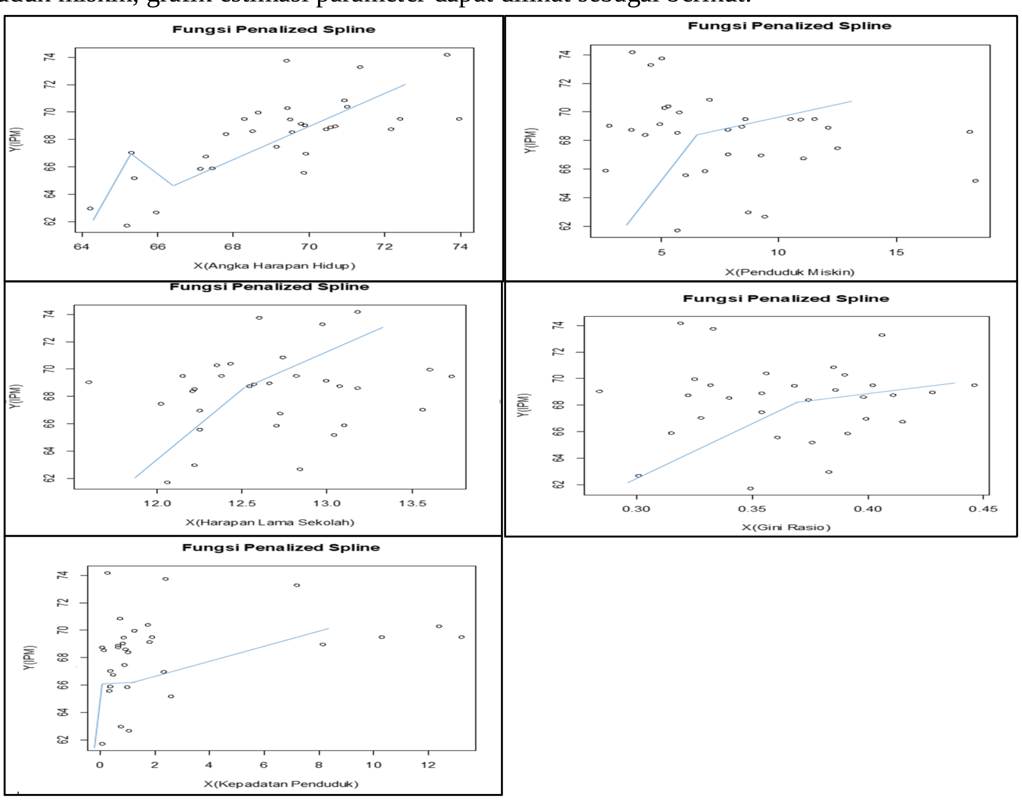
Dengan demikian, bentuk estimator $f(x_5)$ dari regresi *spline* yang diperoleh adalah:

$$f(x_5) = 67,7357 + 0,2361x_5 - 0,0006(x_5 - 0,78)_+ - 0,0024(x_5 - 1,79)_+$$

Atau dapat ditulis seperti berikut:

$$f(x_5) = \begin{cases} 67,7357 + 0,2361x_5, & x_5 < 0,78 \\ 67,7362 + 0,2355x_5, & 0,78 < x_5 < 1,79 \\ 67,7405 + 0,2331x_5, & x_5 \geq 1,79 \end{cases}$$

Berdasarkan hasil estimasi dari variabel gini rasio, harapan lama sekolah, kepadatan penduduk dan penduduk miskin, grafik estimasi parameter dapat dilihat sebagai berikut:



Gambar 4.2 Grafik Estimasi Fungsi $f(x_i)$

Dari Gambar 4.2 diketahui fungsi $f\left(x_i\right)$ menunjukkan perubahan pergerakan data berdasarkan titik-titik knot yang dihasilkan.

UJI SIGNIFIKANSI PARAMETER

Berdasarkan hasil analisis variansi (Anava), berikut ini hasil pengujian parameter secara serentak:

Tabel 2 Analisis Variansi

Model	Df	*Sum of Squares*	*Mean Squares*	F
Regression	16	253,15	15,822	2,8747
Error	14	77,055	5,5039	
Total	30	330,21		

Berdasarkan hasil Anava pada Tabel 2 dapat diketahui bahwa F_{tabel} yakni sebesar 2,092 dan nilai tersebut kurang dari nilai F_{hitung} sebesar 2,8747. Sehingga dalam hal ini dapat disimpulkan bahwa H_0 ditolak, maka terdapat pengaruh antara variabel prediktor terhadap variabel respon. Untuk mengetahui parameter yang berpengaruh dan signifikan terhadap variabel respon dilakukan pengujian secara individu. Berikut hasil pengujian terhadap parameter pada setiap variabel prediktor:

Tabel 3 Uji Parameter Model Regresi

Variabel	$\boldsymbol{\beta}$	Parameter	$\lvert t_{hitung} \rvert$	t_{tabel}	Kesimpulan
X_1	β_0	-13.2256	17,5276	1,7056	Signifikan
	β_1	1.1884	15,129		Signifikan
	β_2	-0.4357	6,507		Signifikan
	β_3	-0.1443	1,714		Signifikan
X_2	β_0	-0.6667	8,647		Signifikan
	β_1	-0.4502	5,839		Signifikan
	β_2	-0.00003	0,0003		Tidak Signifikan
X_3	β_0	53.1942	68,985		Signifikan
	β_1	1.1936	154,803		Signifikan
	β_2	-0.0003	0,004		Tidak Signifikan
X_4	β_0	69.7972	905		Signifikan
	β_1	-0.1892	0,245		Signifikan
	β_2	0.0047	0,060		Tidak Signifikan
X_5	β_0	67.7357	878,489		Signifikan
	β_1	0.2361	3,062		Signifikan
	β_2	-0.0006	0,007		Tidak Signifikan
	β_3	-0.0024	0,031		Tidak Signifikan

Tabel 3 menunjukkan bahwa dari keseluruhan parameter model yang diperoleh ternyata parameter β_0,β_1, β_2, dan β_3 pada X_1, parameter β_0 dan β_1 pada X_2, parameter β_0 dan β_1 pada X_3, parameter β_0 dan β_1 pada X_4 serta parameter β_0 dan β_1 pada X_5 yang berpengaruh signifikan terhadap model regresi *spline* linier. Hal ini diketahui karena $\lvert t_{hitung} \rvert$ lebih besar dari t_{tabel}. Parameter yang signifikan dituliskan pada model regresi *spline* linier terbaik.

PENGUJIAN ASUMSI RESIDUAL

Berdasarkan model yang telah diperoleh, untuk melihat kelayakan model dilakukan uji asumsi residual. Untuk mengetahui ada tidaknya korelasi antar residual dengan uji *Durbin-Watson* sebagai berikut:

Tabel 4 Uji *Durbin-Watson*

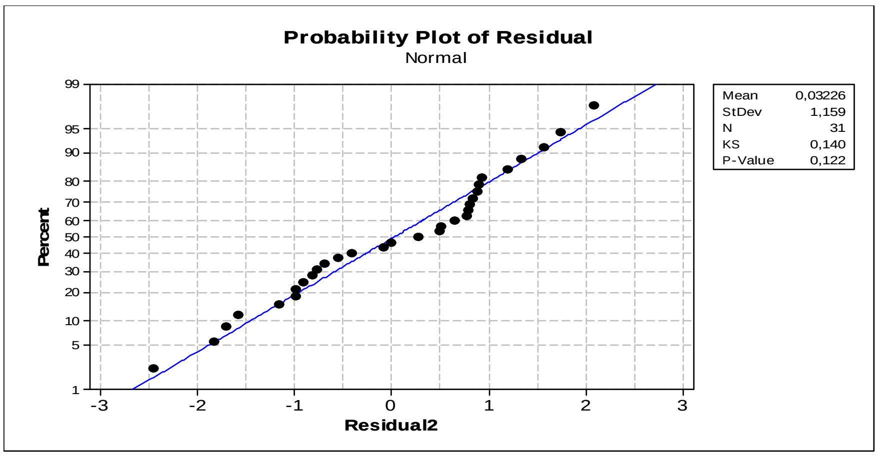
Durbin-Watson
1,7352

Dari Tabel 4 dapat dilihat bahwa nilai DW sebesar 1,7352 atau lebih dari nilai dL sebesar 1,0904 sehingga menunjukkan bahwa tidak terdapat autokorelasi pada residual. Dengan demikian dapat disimpulkan bahwa asumsi residual independen telah terpenuhi. Selanjutnya dilakukan pengujian asumsi residual yang identik. Uji ini dilakukan dengan melihat hasil uji dengan uji *Glejser*. Berikut hasil uji *Glejser* untuk residual pada model:

Tabel 5 Uji *Glejser*

Model	DF	*Sum of Square*	*Mean Square*	F
Regression	16	28,4145	1,6509	1,2808
Error	14	18,045	1,2889	
Total	30	44,4596		

Berdasarkan Tabel 5 diperoleh nilai statistik uji F sebesar 1,2808 dan F_{tabel} sebesar 2,533. Hal ini menunjukkan bahwa F_{hitung} kurang dari F_{tabel} yang berarti H_0 tidak ditolak yakni tidak terdapat heterokedastisitas atau asumsi identik pada residual terpenuhi. Kemudian, untuk melihat residual mengikuti distribusi normal dengan uji *Kolmogorov-Smirnov*. Berikut hasil uji *Kolmogorov-Smirnov* dari residual model:



Gambar 2 Grafik Uji Normalitas Residual

Berdasarkan Gambar 2 diperoleh hasil bahwa *p-value* bernilai 0,122 yang nilainya lebih besar dari α yakni 0,05. Hal ini menunjukkan bahwa residual berdistribusi normal.

MODEL ESTIMASI TERBAIK

Berdasarkan hasil estimasi parameter yang diperoleh, maka model hubungan dari regresi *spline* dengan metode *penalized spline* pada indeks pembangunan manusia dengan setiap variabel prediktor dapat dituliskan dalam persamaan regresi *spline* seperti berikut:

$$y_i = -13,2256 + 1,1884x_1 - 0,4357(x_1 - 68,5)_+ - 0,1443(x_1 - 69,9)_+ - 0,6667 - 0,45021x_2 + 53,1942 + 1,1936x_3 + 69,7972 - 0,1892x_4 + 67,7357 + 0,2361x_5 + \varepsilon_i$$

Berdasarkan hasil estimasi regresi *spline* dengan metode *penalized spline* diperoleh nilai koefisien determinasi terkoreksi (R^2_{adj}) sebesar 76,66%. Sedangkan, untuk nilai estimasi indeks pembangunan manusia dengan metode *penalized spline* memperoleh nilai MAPE sebesar 1,415%.

KESIMPULAN

Hasil estimasi parameter regresi *spline* dengan metode *penalized spline* menghasilkan model yang sesuai untuk indeks pembangunan manusia dengan fleksibilitas yang sangat baik. Hasil regresi *spline* yang diperoleh untuk model terbaik yaitu *spline* linier dengan orde satu. Pemilihan orde *spline*, parameter *smoothing* (λ) dan jumlah knot (K) berdasarkan nilai *Generalized Cross Validation* (GCV) minimum. Untuk parameter *smoothing* (λ) dan jumlah knot (K) pada variabel angka harapan hidup diperoleh jumlah knot sebanyak dua titik dengan nilai (λ) sebesar 0,4. Untuk gini rasio dan harapan lama sekolah diperoleh jumlah knot sebanyak satu titik dan nilai (λ) sebesar 0,9 pada kedua variabel. Sedangkan pada penduduk miskin dan kepadatan penduduk secara berturut-turut diperoleh satu titik knot dan dua titik knot serta nilai (λ) sebesar 0,9 pada kedua variabel. Estimasi yang diperoleh yakni dengan nilai koefisien determinasi terkoreksi (R^2_{adj}) sebesar 76,66%. Untuk model regresi *spline* yang diperoleh menghasilkan nilai MAPE sebesar 1,415% yang menunjukkan bahwa model estimasi yang diperoleh kurang dari 10% dan dapat menginterpretasikan indeks pembangunan manusia dengan sangat baik. Nilai MAPE tersebut juga menunjukkan bahwa model estimasi parameter *spline* sangat baik digunakan pada indeks pembangunan manusia. Dan model regresi *spline* dengan metode *penalized spline* terbaik ialah sebagai berikut:

$$\begin{aligned} y_i = & -13,2256 + 1,1884x_1 - 0,4357(x_1 - 68,5)_+ - 0,1443(x_1 - 69,9)_+ - 0,6667 - 0,45021x_2 + \\ & 53,1942 + 1,1936x_3 + 69,7972 - 0,1892x_4 + 67,7357 + 0,2361x_5 + \varepsilon_i \end{aligned}$$

DAFTAR PUSTAKA

[1]. Eubank RL. *Nonparametric Regression and Spline Smoothing*. New York:Marcel Dekker Inc;1999.

[2]. Tripena A. Penentuan Model Regresi Spline Terbaik. *Seminar Sewindu Statistika*. 2011. 978–979.

[3]. Griggs W. *Penalized Spline Regression and Its Applications*. United States: Whitman College. 2013.

[4]. Putra IMB, Srinadi IGA, Sumarjaya IW. Pemodelan Spline. *E-Jurnal Matematika*. 2015. **4**(3), 110-114.

[5]. Agustina N, Suparti, Mukid, M. A., Pemodelan Data Indeks Harga Saham Gabungan Menggunakan Regresi Penalized Spline. *Jurnal Gaussian*. 2015: **4**(*3*); 603–612.

[6]. Sari RS, Budiantara IN. Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Pendekatan Regresi Spline Multivariabel. *Jurnal Sains dan Seni ITS*. 2012: **1**(*1*); 236-241.

[7]. Gujarati, D.N. *Basic Econometrics*. New York: McGraw Hill Companies Inc; 2003.

WAHYU KURNIASARI	:	Jurusan Matematika FMIPA Untan, Pontianak, wahyukuurniasari@gmail.com
DADAN KUSNANDAR	:	Jurusan Matematika FMIPA Untan, Pontianak, dkusnand@untan.ac.id
EVY SULISTIANINGSIH	:	Jurusan Matematika FMIPA Untan, Pontianak, evysulistianingsih@math.untan.ac.id