INFERENSI, Vol. 5(1), March 2022. ISSN: 0216-308X (Print) 2721-3862 (Online)

Penerapan Keluarga Model Spline Truncated Polinomial pada Regresi Nonparametrik

Andrea Tri Rian Dani[1*], Ludia Ni'matuzzahroh[2]
[1]Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, Kalimantan Timur, Indonesia
[2]Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember, Surabaya, Jawa Timur, Indonesia

*Corresponding author: andreatririandani98@gmail.com

Received: 16 Maret 2022
Accepted: 30 Maret 2022
Published: 6 April 2022

ABSTRAK – **Salah satu pendekatan yang sering digunakan oleh para peneliti untuk mengetahui bentuk pola hubungan antara variabel respon dan variabel prediktor dalam analisis regresi, yaitu pendekatan nonparametrik, dimana pendekatan tersebut digunakan ketika kurva regresi diduga polanya tidak diketahui. Spline truncated merupakan model polinomial dalam regresi nonparametrik yang memiliki sifat tersegmen, dimana sifat tersebut memberikan fleksibilitas yang lebih bagus dibandingkan model polinomial biasa, serta mampu menangani data yang perilakunya berubah pada sub-sub interval tertentu karena adanya titik-titik knot di dalamnya. Pada penelitian ini bertujuan untuk menerapkan keluarga model spline truncated polinomial pada regresi nonparametrik dalam kasus data otomotif. Metode estimasi yang digunakan yaitu *Ordinary Least Square* (OLS). Banyaknya titik knot yang dicobakan yaitu sebanyak 1 hingga 4 titik knot dengan derajat $p = 1, 2, 3$. Berdasarkan hasil analisis, diperoleh model terbaik yang menghasilkan nilai GCV terkecil yaitu model regresi nonparametrik spline truncated kuadratik dengan 4 titik knot, dimana menghasilkan nilai GCV sebesar 522,27 serta nilai koefisien determinasi sebesar 79,77%.**

Kata Kunci– **GCV, Regresi Nonparametrik, Spline Truncated Polinomial**

I. PENDAHULUAN

Analisis regresi merupakan metode analisis dalam statistika yang berperan dalam melihat pola atau dengan kata lain *pattern* antara variabel respon dan prediktor. Bentuk pola hubungan antara variabel respon dan variabel prediktor dalam analisis regresi tidak selalu dapat didekati dengan pendekatan parametrik, yang mana bentuk kurva regresi diasumsikan dapat diketahui yaitu seperti kuadratik, linier, kubik, dan lain sebagainya. Akan tetapi juga terdapat banyak kasus, dimana bentuk pola hubungan antar variabel respon dan variabel prediktor perlu didekati menggunakan pendekatan semiparametrik dan pendekatan nonparametrik. Pendekatan regresi semiparametrik digunakan apabila bentuk kurva regresi sebagian dapat diketahui dan sebagian tidak dapat diketahui. Sedangkan pendekatan regresi nonparametrik dapat digunakan ketika bentuk kurva regresi diasumsikan tidak diketahui.

Model regresi nonparametrik yang sering digunakan oleh para peneliti, salah satunya yaitu spline. Model regresi nonparametrik spline merupakan model yang memiliki interpretasi visual khusus dan interpretasi statistik yang sangat baik [1]. Spline sendiri adalah model polinomial yang mempunyai sifat tersegmen, dimana sifat tersebut memberikan fleksibilitas yang lebih bagus dibandingkan model polinomial biasa. Budiantara (2005) telah mengembangkan estimator spline dalam pendekatan regresi nonparametrik menggunakan basis fungsi keluarga spline truncated. Kelebihan dari spline truncated yaitu memiliki kemampuan sangat baik dalam menangani data yang perilakunya berubah pada sub-sub interval tertentu [3], serta dengan adanya titik-titik knot dapat menghasilkan kurva regresi yang relatif mulus karena mampu mengatasi pola data yang menunjukkan pola naik dan pola turun yang tajam [4]. Menurut Montoya, dkk (2014), dalam membangun model regresi spline truncated, terdapat beberapa hal yang perlu diperhatikan, diantaranya menentukan jumlah derajat dari model regresi, jumlah titik knot, dan lokasi titik knot berada.

Menurut Budiantara (2019), regresi nonparametrik dengan fungsi spline truncated ini dikembangkan karena adanya kelemahan-kelemahan yang dimiliki oleh fungsi polinomial, salah satunya yaitu bersifat global. Fungsi spline truncated merupakan suatu fungsi yang masih mempertahankan sifat dari fungsi polinomial, yang dirancang dengan cara memodifikasi fungsi polinomial. Berdasarkan kejadian-kejadian khusus dari fungsi spline truncated dengan derajat p, dimana p = 1, 2, 3; maka akan diperoleh tiga bentuk fungsi spline truncated. Apabila p = 1 maka akan diperoleh fungsi spline truncated linier, apabila p = 2 maka akan diperoleh fungsi spline truncated kuadratik, dan apabila p = 3 maka akan diperoleh fungsi spline truncated kubik. Pendekatan regresi nonparametrik spline truncated telah digunakan dan dikembangkan dalam penelitian sebelumnya, beberapa diantaranya yaitu oleh [7]–[17].

Pendekatan regresi nonparametrik spline truncated tidak hanya dapat diterapkan pada bidang kesehatan, sosial, ekonomi, dan lain sebagainya. Akan tetapi juga dapat diterapkan pada bidang ilmu yang membahas tentang sistem kendaraan bermotor, yaitu otomotif. Beberapa variabel yang menyangkut kasus otomotif yaitu diantaranya variabel akselerasi sepeda motor dan variabel waktu.

Berdasarkan uraian di atas, maka pada penelitian ini akan menggunakan bentuk pola hubungan antara variabel akselerasi sepeda motor (y) dan variabel waktu (x) untuk menerapkan keluarga model spline truncated polinomial pada regresi nonparametrik. Kemudian bentuk pola hubungan antara kedua variabel tersebut akan dimodelkan dengan pendekatan regresi nonparametrik spline truncated dengan menggunakan kombinasi banyaknya titik knot dan derajat p. Penelitian ini bertujuan untuk membuktikan bahwa pendekatan regresi nonparametrik spline truncated mampu memberikan fleksibilitas yang baik dalam proses estimasi kurva regresi.

II. METODOLOGI PENELITIAN

A. Regresi Nonparametrik

Pada saat bentuk kurva regresi antara variabel respon dan variabel prediktor tidak mengacu pada suatu pola tertentu atau tidak terdapat informasi di masa lalu yang lengkap tentang pola data, maka dapat digunakan pendekatan regresi nonparametrik. Pendekatan regresi nonparametrik mempunyai fleksibilitas yang tinggi, dikarenakan data diharapkan mampu mencari sendiri bentuk estimasinya tanpa dipengaruhi subyektifitas peneliti [3]. Secara umum, model dari pendekatan regresi nonparametrik seperti pada Persamaan (1).

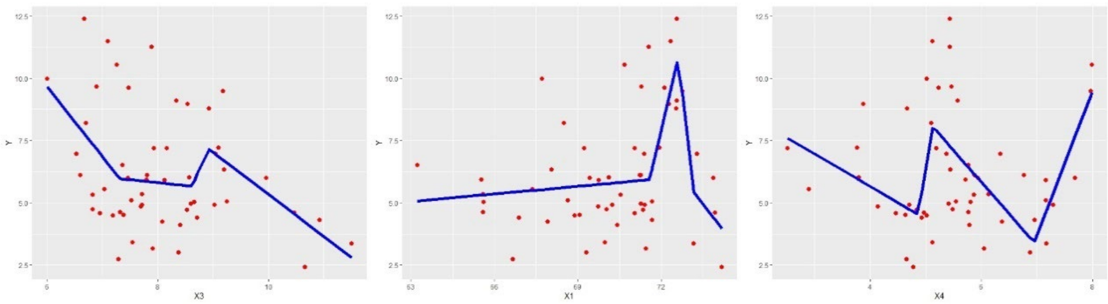
$$y_i = m(x_i) + \varepsilon_i \quad ; i = 1,2, \dots , n \tag{1}$$

dimana:

y_i : variabel respon
x_i : variabel prediktor
$m(x_i)$: fungsi dari kurva regresi yang akan dihampiri
ε_i : *error random* yang diasumsikan IIDN dengan *mean* bernilai nol dan variasi σ^2.

B. Regresi Nonparametrik Spline Truncated

Pendekatan regresi nonparametrik spline truncated atau estimator spline truncated cenderung memiliki kemampuan yang baik dalam menangani/mengatasi pola data yang perilakunya menunjukkan perubahan pada sub-sub interval tertentu. Hal tersebut terjadi karena dalam fungsi spline truncated terdapat titik-titik knot. Ilustrasi dari titik knot ditampilkan pada Gambar 1.



Gambar 1 Ilustrasi dari Titik Knot

Secara umum, model regresi nonparametrik spline truncated univariabel dapat dituliskan seperti pada Persamaan (2).

$$y_i = m(x_i) + \varepsilon_i \quad ; i = 1,2, \dots , n \tag{2}$$

dengan,

$$m(x_i) = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{k=1}^{r} \beta_{p+k} (x_i - K_k)_+^p$$

dan komponen truncatednya:

$$(x_i - K_k)_+^p = \begin{cases} (x_i - K_k)^p & ; x \geq K_k \\ 0 & ; x < K_k \end{cases}$$

dimana:

y_i : variabel respon
x_i : variabel prediktor
$m(x_i)$: fungsi regresi nonparametrik spline truncated
ε_i : *error random* yang diasumsikan IIDN dengan *mean* bernilai nol dan variasi σ^2

$f(x_i) = \sum_{j=0}^{p} \beta_j x_i^j$: komponen polinomial

$\sum_{k=1}^{r} \beta_{p+k} (x_i - K_k)_+^p$: komponen truncated

DOI: 10.12962/j27213862.v5i1.12537

Berdasarkan model pada Persamaan (2) dapat dijabarkan untuk setiap $i = 1,2,\dots,n$

$$\begin{aligned}
i = 1\text{ , maka} \quad & y_1 = \sum_{j=0}^{p} \beta_j x_1^j + \sum_{k=1}^{r} \beta_{p+k}(x_1 - K_k)_+^p + \varepsilon_1 \\
i = 2\text{ , maka} \quad & y_2 = \sum_{j=0}^{p} \beta_j x_2^j + \sum_{k=1}^{r} \beta_{p+k}(x_2 - K_k)_+^p + \varepsilon_2 \\
\vdots \quad & \\
i = n\text{ , maka} \quad & y_n = \sum_{j=0}^{p} \beta_j x_n^j + \sum_{k=1}^{r} \beta_{p+k}(x_n - K_k)_+^p + \varepsilon_n
\end{aligned} \tag{3}$$

Penjabaran model regresi nonparametrik spline truncated pada Persamaan (3) dapat dibentuk matriks:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \dots & x_1^p & \vdots & (x_1 - K_1)_+^p & \dots & (x_1 - K_r)_+^p \\ 1 & x_2^1 & x_2^2 & \dots & x_2^p & \vdots & (x_2 - K_1)_+^p & \dots & (x_2 - K_r)_+^p \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & x_n^2 & \dots & x_n^p & \vdots & (x_n - K_1)_+^p & \dots & (x_n - K_r)_+^p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p+r} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{4}$$

Kemudian dapat dibentuk dalam notasi matriks sebagai berikut:

$$\tilde{y} = \boldsymbol{X}\tilde{\beta} + \tilde{\varepsilon} \tag{5}$$

dimana $\tilde{y}$ merupakan vektor dari variabel respon berukuran $n \times 1$, $\boldsymbol{X}$ merupakan matriks berukuran $n \times (m + r + 1)$, $\tilde{\beta}$ merupakan vektor parameter koefisien regresi yang akan diestimasi dan berukuran $(m + r + 1) \times 1$, serta $\tilde{\varepsilon}$ merupakan vektor *error random* berukuran $n \times 1$. Pada penelitian ini, parameter $\tilde{\beta}$ diestimasi menggunakan metode OLS (*Ordinary Least Square*), yaitu estimator didapatkan dengan cara meminimumkan jumlah kuadrat *error*. Berdasarkan Persamaan (5), maka estimator $\tilde{\beta}$ dapat diperoleh dengan cara menyelesaikan optimasi OLS sebagai berikut:

$$min\{Q(\tilde{\beta})\} = min\left\{(\tilde{y} - \boldsymbol{X}\tilde{\beta})^T(\tilde{y} - \boldsymbol{X}\tilde{\beta})\right\} \tag{6}$$

Optimasi pada Persamaan (6) dapat diselesaikan sebagai berikut:

$$\begin{aligned} Q(\tilde{\beta}) &= (\tilde{y} - \boldsymbol{X}\tilde{\beta})^T(\tilde{y} - \boldsymbol{X}\tilde{\beta}) \\ &= (\tilde{y}^T - \boldsymbol{X}^T\tilde{\beta}^T)(\tilde{y} - \boldsymbol{X}\tilde{\beta}) \\ &= \tilde{y}^T\tilde{y} - 2\tilde{\beta}^T\boldsymbol{X}^T\tilde{y} + \tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta} \end{aligned} \tag{7}$$

Kemudian melakukan derivatif parsial $Q(\tilde{\beta})$ terhadap $\tilde{\beta}$ untuk mendapatkan estimator $\tilde{\beta}$**.**

$$\begin{aligned} \frac{\partial Q(\tilde{\beta})}{\partial \tilde{\beta}} &= \frac{\partial(\tilde{y}^T\tilde{y} - 2\tilde{\beta}^T\boldsymbol{X}^T\tilde{y} + \tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta})}{\partial \tilde{\beta}} \\ &= -2\boldsymbol{X}^T\tilde{y} + 2\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta} \end{aligned} \tag{8}$$

Selanjutnya derivatif parsial pada Persamaan (8) disama dengankan nol, sehingga:

$$-2\boldsymbol{X}^T\tilde{y} + 2\boldsymbol{X}^T\boldsymbol{X}\hat{\tilde{\beta}} = 0 \tag{9}$$

Persamaan (9) dapat ditulis:

$$\boldsymbol{X}^T\boldsymbol{X}\hat{\tilde{\beta}} = \boldsymbol{X}^T\tilde{y} \tag{10}$$

Berdasarkan Persamaan (10), maka diperoleh estimator $\hat{\tilde{\beta}}$ sebagai berikut:

$$\hat{\tilde{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{y} \tag{11}$$

Berdasarkan model pada Persamaan (2) serta apabila menggunakan satu variabel respon dan satu variabel prediktor, maka dapat dibuat model regresi nonparametrik spline truncated derajat p dimana $(p = 1,2,3)$ dengan titik-titik knot pada titik-titik K_1, K_2, K_3, K_4 sebagai berikut:

1. Model regresi nonparametrik spline truncated linier $(p = 1)$ dengan 1 titik knot (K_1)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(x_1 - K_1)_+$
2. Model regresi nonparametrik spline truncated kuadratik $(p = 2)$ dengan 1 titik knot (K_1)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3(x_1 - K_1)_+^2$
3. Model regresi nonparametrik spline truncated kubik $(p = 3)$ dengan 1 titik knot (K_1)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4(x_1 - K_1)_+^3$
4. Model regresi nonparametrik spline truncated linier $(p = 1)$ dengan 2 titik knot (K_2)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(x_1 - K_1)_+ + \hat{\beta}_3(x_1 - K_2)_+$
5. Model regresi nonparametrik spline truncated kuadratik $(p = 2)$ dengan 2 titik knot (K_2)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3(x_1 - K_1)_+^2 + \hat{\beta}_4(x_1 - K_2)_+^2$
6. Model regresi nonparametrik spline truncated kubik $(p = 3)$ dengan 2 titik knot (K_2)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4(x_1 - K_1)_+^3 + \hat{\beta}_5(x_1 - K_2)_+^3$
7. Model regresi nonparametrik spline truncated linier $(p = 1)$ dengan 3 titik knot (K_3)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(x_1 - K_1)_+ + \hat{\beta}_3(x_1 - K_2)_+ + \hat{\beta}_4(x_1 - K_3)_+$
8. Model regresi nonparametrik spline truncated kuadratik $(p = 2)$ dengan 3 titik knot (K_3)
   $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3(x_1 - K_1)_+^2 + \hat{\beta}_4(x_1 - K_2)_+^2 + \hat{\beta}_5(x_1 - K_3)_+^2$

9. Model regresi nonparametrik spline truncated kubik ($p = 3$) dengan 3 titik knot (K_3)
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4 (x_1 - K_1)_+^3 + \hat{\beta}_5 (x_1 - K_2)_+^3 + \hat{\beta}_6 (x_1 - K_3)_+^3$
10. Model regresi nonparametrik spline truncated linier ($p = 1$) dengan 4 titik knot (K_4)
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - K_1)_+ + \hat{\beta}_3 (x_1 - K_2)_+ + \hat{\beta}_4 (x_1 - K_3)_+ + \hat{\beta}_5 (x_1 - K_4)_+$
11. Model regresi nonparametrik spline truncated kuadratik ($p = 2$) dengan 4 titik knot (K_4)
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 (x_1 - K_1)_+^2 + \hat{\beta}_4 (x_1 - K_2)_+^2 + \hat{\beta}_5 (x_1 - K_3)_+^2 + \hat{\beta}_6 (x_1 - K_4)_+^2$
12. Model regresi nonparametrik spline truncated kubik ($p = 3$) dengan 4 titik knot (K_4)
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4 (x_1 - K_1)_+^3 + \hat{\beta}_5 (x_1 - K_2)_+^3 + \hat{\beta}_6 (x_1 - K_3)_+^3 + \hat{\beta}_7 (x_1 - K_4)_+^3$

C. Penentuan Titik Knot Optimal

Penentuan titik knot yang optimal merupakan hal yang sangat penting dalam regresi nonparametrik spline truncated. Titik knot merupakan titik perpaduan bersama dimana adanya perubahan perilaku pola data interval yang berbeda [18]. Terdapat beberapa metode dalam menentukan titik knot optimal yang telah dikembangkan oleh peneliti, salah satunya yaitu menggunakan *Generalized Cross-Validation* (GCV) [19]. Model regresi nonparametrik spline truncated yang terbaik adalah model dengan titik knot optimal yang diperoleh dari nilai GCV terkecil [2], [20]-[21]. Rumus GCV adalah sebagai berikut:

$$GCV(\tilde{K}) = n^{-1} \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\left[n^{-1} trace\left(\boldsymbol{I} - \boldsymbol{A}(\tilde{K})\right)\right]^2} \tag{12}$$

dimana:

$GCV(\tilde{K})$: vektor yang memuat nilai GCV dari titik-titik knot

n^{-1} : banyaknya pengamatan sebanyak n

$\boldsymbol{A}(\tilde{K})$: matriks yang memuat titik knot $\tilde{K} = (K_1, K_2, \dots, K_r)^T$ diperoleh $\boldsymbol{X}(\tilde{K})\left[\boldsymbol{X}(\tilde{K})^T \boldsymbol{X}(\tilde{K})\right]^{-1} \boldsymbol{X}(\tilde{K})^T$

D. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari https://vincentarelbundock.github.io/Rdatasets/datasets.html terkait *Simulated Motorcycle Accident* . Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon dan satu variabel prediktor.

Variabel respon:
y : akselerasi sepeda motor

Variabel prediktor:
x : waktu

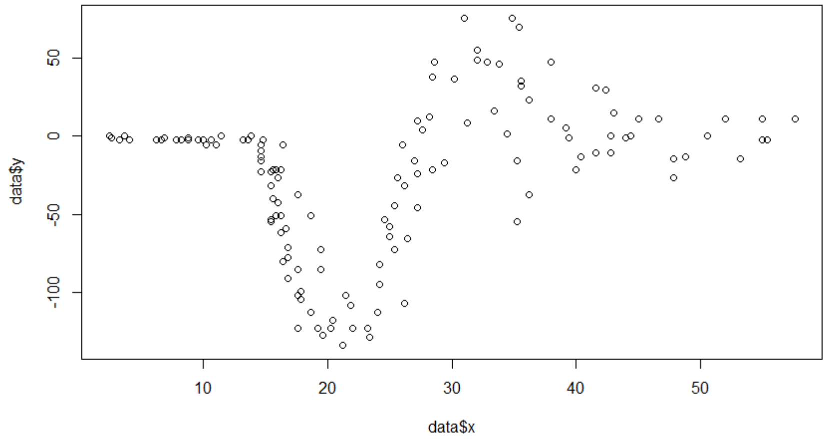
E. Langkah-Langkah Penelitian

Langkah-langkah yang dilakukan dalam penelitian ini disusun untuk menyelesaikan tujuan penelitian, yaitu menerapkan keluarga spline truncated polinomial pada regresi nonparametrik dalam kasus data otomotif untuk membuktikan bahwa pendekatan regresi nonparametrik spline truncated mampu memberikan fleksibilitas yang baik dalam proses estimasi kurva regresi.

1. Membuat *scatter plot* antara variabel respon dan variabel prediktor.
2. Memodelkan data penelitian menggunakan model regresi nonparametrik spline truncated, dengan titik knot antara 1, 2, 3, hingga 4 titik knot dan derajat $p = 1$,2, 3 (linier, kuadratik, kubik).
3. Menentukan titik knot optimal dengan menggunakan metode *Generalized Cross-Validation* (GCV), dimana titik knot yang optimal diperoleh dari hasil nilai GCV yang terkecil.
4. Menghitung nilai koefisien determinasi (R^2) sebagai kriteria kebaikan model.
5. Melakukan interpretasi dari model nonparametrik spline truncated yang terbaik.

III. ANALISIS DAN PEMBAHASAN

Analisis dimulai dengan menampilkan diagram pencar (*scatter plot*) untuk melihat pola dari variabel prediktor terhadap variabel respon. Diagram pencar ditampilkan pada Gambar 2.

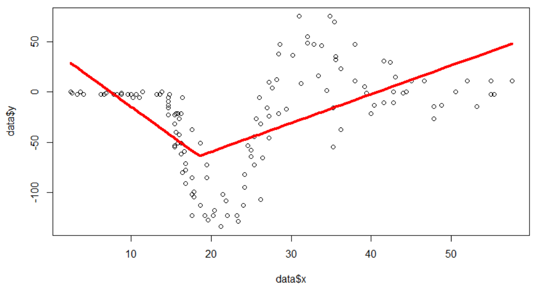


Gambar 2 *Scatter Plot*

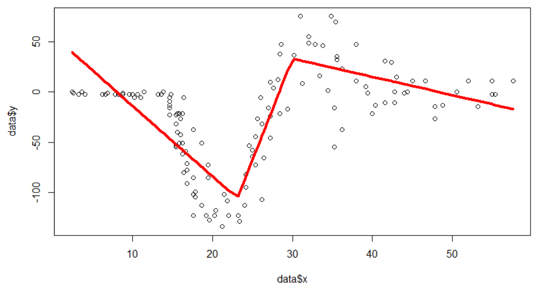
Berdasarkan Gambar 2, terlihat jika pola hubungan antara variabel prediktor terhadap variabel respon menunjukkan karakteristik yaitu berubah-ubah pada sub-sub interval tertentu. Sehingga pada penelitian ini, akan dilakukan pemodelan dengan menggunakan keluarga model spline truncated polinomial.

A. Pemodelan Keluarga Model Spline Truncated Polinomial

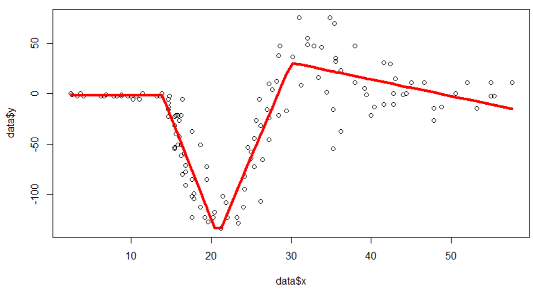
Seperti yang telah dijelaskan sebelumnya, jika pendekatan regresi nonparametrik spline truncated cenderung memiliki kemampuan yang bagus dalam menangani data yang mempunyai perilaku berubah-ubah pada sub-sub interval tertentu. Berdasarkan Gambar 2, kondisi tersebut terpenuhi. Hasil pemodelan regresi nonparametrik dengan menggunakan keluarga model spline truncated polinomial dengan derajat *p=1* dan jumlah titik knot yang dicobakan sebanyak K_1, K_2, K_3, K_4 ditampilkan pada Gambar 3.



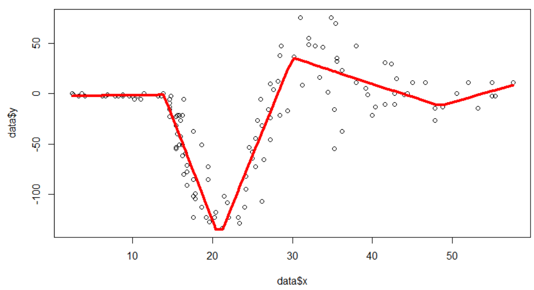
(a) 1 Titik Knot



(b) 2 Titik Knot



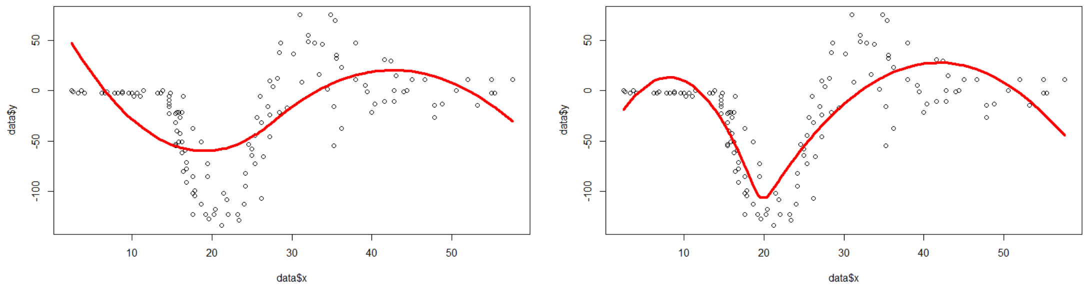
(c) 3 Titik Knot



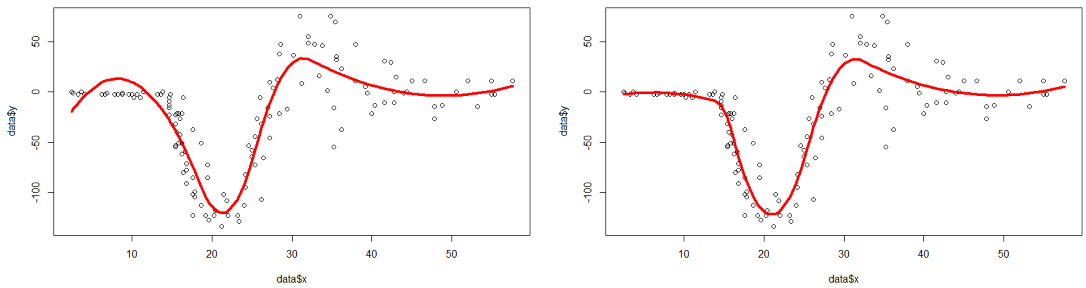
(d) 4 Titik Knot

Gambar 3 Hasil Estimasi Kurva Regresi Spline Truncated *p=1* (Linear)

Hasil pemodelan regresi nonparametrik dengan menggunakan keluarga model spline truncated polinomial dengan derajat *p=2* dan jumlah titik knot yang dicobakan adalah sama ditampilkan pada Gambar 4.



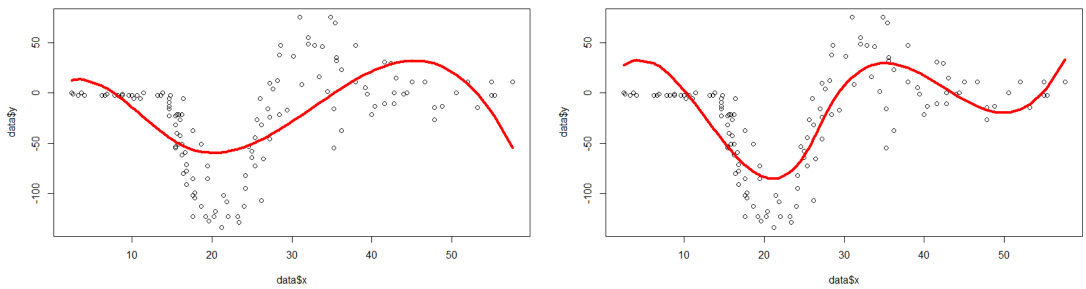
(a) 1 Titik Knot *(b) 2 Titik Knot*



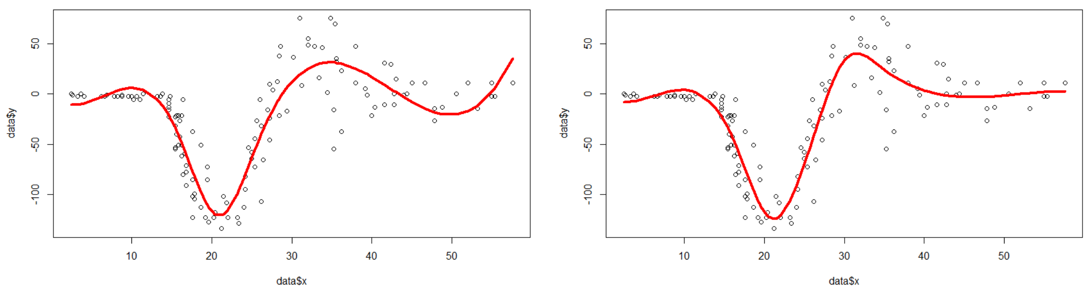
(c) 3 Titik Knot *(d) 4 Titik Knot*

Gambar 4 Hasil Estimasi Kurva Regresi Spline Truncated *p=2* (Kuadratik)

Hasil pemodelan regresi nonparametrik dengan menggunakan keluarga model spline truncated polinomial dengan derajat *p=3* dan jumlah titik knot yang dicobakan adalah sama ditampilkan pada Gambar 4.



(a) 1 Titik Knot *(b) 2 Titik Knot*



(c) 3 Titik Knot *(d) 4 Titik Knot*

Gambar 5 Hasil Estimasi Kurva Regresi Spline Truncated *p=3* (Kubik)

Berdasarkan Gambar 3, 4, dan 5, hasil estimasi kurva regresi dengan menggunakan pendekatan regresi nonparametrik spline truncated keluarga polinomial cenderung mengikuti pola data aktualnya. Terlihat juga semakin banyak jumlah titik knot yang digunakan, maka akan mengikuti pola hubungannya.

B. Penentuan Model Terbaik

Penentuan model terbaik dari model regresi nonparametrik spline truncated polinomial dilakukan berdasarkan nilai GCV terkecil yang didukung dengan nilai Koefisien Determinasi (R^2) yang lebih besar. Pada Tabel 1, ditampilkan kembali hasil estimasi dari model-model spline truncated polinomial:

Tabel 1 Tabulasi Hasil Estimasi

Jumlah Knot	Model Regresi	GCV Minimum	MSE	Koefisien Determinasi
1	*Spline* Linear	1578,23	1507,83	34,93%
	Spline Kuadratik	1447,86	1362,08	41,22%
	Spline Kubik	1510,80	1421,29	39,03%
2	*Spline* Linear	783,75	737,32	68,18%
	Spline Kuadratik	917,51	849,82	63,32%
	Spline Kubik	891,32	825,56	65,11%
3	*Spline* Linear	548,83	508,34	78,06%
	Spline Kuadratik	561,93	512,37	77,88%
	Spline Kubik	593,08	540,78	76,59%
4	*Spline* Linear	536,50	489,18	78,89%
	***Spline* Kuadratik**	**522,27**	**468,75**	**79,77%**
	Spline Kubik	534,10	479,36	79,27%

Berdasarkan Tabel 1, terlihat model regresi nonparametrik spline truncated polinomial yang terbaik adalah model regresi spline truncated kuadratik dengan 4 titik knot, dimana memiliki nilai GCV yang terkecil sebesar 522,27 dan nilai Koefisien Determinasi (R^2) yaitu sebesar 79,77%. Model akhir nya dapat dituliskan sebagai berikut:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 (x_1 - K_1)_+^2 + \hat{\beta}_4 (x_1 - K_2)_+^2 + \hat{\beta}_5 (x_1 - K_3)_+^2 + \hat{\beta}_6 (x_1 - K_4)_+^2$$

$$\hat{y} = -5{,}70 + 1{,}63x_1 - 0{,}13x_1^2 - 6{,}25(x_1 - 13{,}9)_+^2 + 9{,}68(x_1 - 16{,}2)_+^2 - 5{,}73(x_1 - 25{,}4)_+^2 + 2{,}54(x_1 - 32{,}3)_+^2$$

IV. KESIMPULAN

Penerapan model regresi nonparametrik spline truncated polinomial pada penelitian ini telah berhasil dilakukan. Terlihat jika memang benar bahwa pendekatan regresi nonparametrik spline truncated memberikan fleksibilitas dalam proses estimasi kurva regresi. Model terbaik didapatkan berdasarkan nilai GCV yang terkecil dengan didukung nilai Koefisien Determinasi (R^2).

REFERENSI

[1] I. N. Budiantara, "Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter,"Prosiding Seminar Nasional FMIPA Undhiksa, pp. 09–28, 2011.

[2] I. N. Budiantara, "Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik,"Berkala MIPA, vol. 15(3), pp. 55–61, 2005.

[3] R. L. Eubank, Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker, 1999.

[4] W. Hardle, Applied Nonparametric Regression. New York: Cambridge University Press, 1990.

[5] E. Montoya, N. Ulloa, dan V. Miller, "A Simulation Study Comparing Knot Slection Methods with Equally Spaced Knots in a Penalized Regression Spline,"International Journal of Statistics and Probability, vol. 03(03), pp. 96–110, 2014.

[6] I. N. Budiantara, Regresi Nonparametrik Spline Truncated. Surabaya: ITS Press, 2019.

[7] A. Tripena, "Analisis Regresi Spline Kuadratik," dipresentasikan dalam Seminar Nasional Matematika dan Pendidikan Matematika, Jurusan Pendidikan Matematika, FMIPA UNY, 2011.

[8] I. N. Budiantara, M. Ratna, I. Zain, dan W. Wibowo, "Modeling the Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach,"International Journal of Basic & Applied Sciences IJBAS-IJENS, vol. 12, no. 06, pp. 119–124, 2012.

[9] R. K. Dewi dan I. N. Budiantara "Faktor-Faktor yang Mempengaruhi Angka Gizi Buruk di Jawa Timur dengan Pendekatan Regresi Nonparametrik Spline,"JURNAL SAINS DAN SENI ITS, vol. 1, no. 1, pp. D177–D182, 2012.

[10] M. F. Bintariningrum dan I. N. Budiantara "Pemodelan Regresi Nonparametrik Spline Truncated dan Aplikasinya pada Angka Kelahiran Kasar di Surabaya,"JURNAL SAINS DAN SENI POMITS, vol. 3, no. 1, pp. D-7–D-12, 2014.

[11] N. Arfan dan I. N. Budiantara, "Pendekatan Spline untuk Estimasi Kurva Regresi Nonparametrik (Studi Kasus pada Data Angka Kematian Maternal di Jawa Timur),"JURNAL SAINS DAN SENI POMITS, vol. 3, no. 1, pp. D-13–D-17, 2014.

[12] A. Islamiyati,"Spline Polynomial Truncated dalam Regresi Nonparametrik,"Jurnal Matematika, Statistika, & Komputasi, vol. 14, no. 1, pp. 54–60, 2017.

[13] D. R. S. Saputro, K. R. Demu, dan P. Widyaningsih, "Nonparametric Truncated Spline Regression Model on Data of Human Development Index (HDI) in Indonesia,"IOP Conf. Series: Journal of Physics, vol. 1188, pp. 01–01, 2018.

[14] A. T. R. Dani, N. Y. Adrianingsih, dan A. Ainurrochmah, "Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi,"JAMBURA Journal of Probability and Statistics, vol. 01, pp. 98–106, 2020.

[15] A. T. R. Dani, N. Y. Adrianingsih, A. Ainurrochmah, dan R. Sriningsih, "Flexibility of Nonparametric Regression Spline Truncated on Data without a Specific Pattern,"Jurnal Litbang Edusaintech, vol. 2(1), pp. 37–43, 2021.

[16] A. T. R. Dani, L. Ni'matuzzahroh, V. Ratnasari, dan I. N. Budiantara, "Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal,"INFERENSI, vol. 4(1), pp. 47–55, 2021.
[17] A. T. R. Dani dan L. Ni'matuzzahroh, "Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated,"J Statistika, vol. 14, no. 1, pp. 24–29, 2021.
[18] W. Hardle, Applied Nonparametric Regression. Berlin: Humboldt-Universitat zu Berlin, 1994.
[19] G. Wahba, Spline Models for Observational Data. Pennsylvania: SIAM, 1990.
[20] A. T. R. Dani, V. Ratnasari, and I. N. Budiantara, "Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1115, no. 1, p. 012020, 2021, doi: 10.1088/1757-899x/1115/1/012020.
[21] V. Ratnasari, I. N. Budiantara, and A. T. R. Dani, "Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized Cross- Validation (GCV), and Unbiased Risk (UBR) Methods," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 11, no. 6, pp. 2400–2406, 2021.