



Flexibility of Nonparametric Regression Spline Truncated on Data without a Specific Pattern

Andrea Tri Rian Dani^{1}, Narita Yuri Adrianingsib², Alifita Ainurrochmah³, Riry Sriningsib⁴*

^{1,2,3,4} Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, East Java, Indonesia

* Correspondence: E-mail: andreatriandani98@gmail.com

ABSTRACTS

The form of the relationship pattern between the predictor variable and the response variable is known, but in fact some is unknown. If the form of the relationship pattern between the response variable and the predictor variable is unknown, the nonparametric regression approach is the most appropriate approach. The nonparametric regression approach does not depend on the assumption of a certain regression curve shape, so it will provide high flexibility. One of the well-known nonparametric regression estimators is the truncated spline. Truncated splines are polynomial pieces that have segmented and continuous properties. In this study, the relationship pattern between the two variables will be simulated, namely responses and predictors that do not have a specific pattern, which is then approached with two regression approaches, namely parametric and nonparametric. It will be shown that the nonparametric regression approach has high flexibility so that it is able to adjust the form of regression curve estimation by itself. The measure of the goodness of the regression curve estimation uses the coefficient of determination.

ARTICLE INFO

Article History:

Received 18 Feb 2021

Revised 29 Mei 2021

Accepted 30 Mei 2021

Available online 31 Mei 2021

Keyword:

*Nonparametric Regression,
Parametric Regression,
Spline Truncated*

ABSTRAK

Bentuk pola hubungan antara variabel prediktor dan variabel respon ada yang diketahui, namun pada nyatanya ada pula yang tidak diketahui. Apabila bentuk pola hubungan antara variabel respon dan variabel prediktor tidak diketahui, pendekatan regresi nonparametrik merupakan pendekatan yang paling sesuai. Pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva regresi tertentu, sehingga akan memberikan fleksibilitas yang tinggi. Salah satu estimator regresi nonparametrik yang terkenal adalah spline truncated. Spline truncated merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Pada penelitian ini, akan disimulasikan pola hubungan antara kedua variabel yaitu respon dan prediktor yang tidak memiliki pola tertentu, yang kemudian didekati dengan dua pendekatan regresi, yaitu parametrik dan nonparametrik. Berdasarkan ukuran kebaikan estimasi kurva regresi menggunakan koefisien determinasi diperoleh hasil bahwa pendekatan regresi nonparametrik lebih baik daripada pendekatan regresi parametrik. Hal ini dikarenakan pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi sehingga mampu menyesuaikan sendiri bentuk estimasi kurva regresi.

Keyword:

*Regresi Nonparametrik
Regresi Parametrik,
Spline Truncated*

1. PENDAHULUAN

Analisis regresi adalah salah satu metode statistika yang digunakan untuk mengetahui pola hubungan antara satu atau lebih variabel. Variabel yang digunakan dalam analisis regresi terdiri atas variabel respon dan variabel prediktor (Budiantara, 2019). Selanjutnya tujuan utama dalam melakukan analisis regresi adalah untuk mencari bentuk estimasi kurva regresi. Terdapat tiga pendekatan yang biasa digunakan untuk mengestimasi bentuk kurva regresi, yaitu pendekatan regresi parametrik, nonparametrik, dan semiparametrik. Pendekatan regresi parametrik merupakan pendekatan yang digunakan ketika bentuk hubungannya diketahui dan diasumsikan mengikuti pola data tertentu (Suparti, dkk., 2018). Pendekatan regresi nonparametrik merupakan pendekatan yang digunakan ketika bentuk hubungannya tidak diketahui polanya yang selanjutnya kurva regresinya diasumsikan mulus (Eubank, 1999). Sedangkan pada pendekatan regresi semiparametrik merupakan kombinasi dari nonparametrik dan parametrik.

Pada analisis regresi, jika sampel acak antara variabel respon dan prediktor digambarkan dalam suatu koordinat kartesius maka akan diperoleh diagram pencar. Dengan mengamati diagram pencar yang dihasilkan, maka akan didapatkan informasi apakah data mengikuti suatu pola tertentu atau tidak, misalnya jika titik-titik dalam diagram pencar mengikuti pola garis lurus, maka berhubungan secara linear, dan jika titik-titik dalam diagram pencar menunjukkan pola kuadrat, maka menunjukkan bahwa berhubungan secara kuadrat. Hal tersebut biasa ditemui pada pendekatan regresi parametrik. Begitu juga halnya dengan pendekatan regresi nonparametrik, apabila pola data tidak diketahui maka pendekatan ini dapat diusulkan. Terdapat beberapa metode estimator dalam regresi nonparametrik yang terkenal, diantaranya adalah spline truncated, kernel dan deret Fourier. Beberapa pemodelan dengan menggunakan pendekatan regresi nonparametrik yang telah dikembangkan oleh peneliti diantaranya pernah dilakukan oleh Nisa, dkk., (2018); Rahasia, dkk., (2020); Suparti

dan Prahutama, (2016); Dani, dkk., (2020) dan Saputro, dkk., (2019).

Mengingat adanya kelemahan-kelemahan pada pendekatan regresi parametrik, maka perlu dikembangkan suatu fungsi yang dapat mengatasi kelemahan tersebut. Spline truncated adalah suatu fungsi yang memodifikasi fungsi polinomial menjadi bentuk potongan-potongan polinomial. Spline truncated mempunyai beberapa kelebihan diantara lain adalah dapat menangani karakteristik data yang mempunyai perilaku berubah-ubah pada sub-sub interval tertentu, mempunyai interpretasi statistik yang sederhana dan memiliki penyajian visual yang sangat baik. Dengan menggunakan titik knot, kurva regresi dengan spline truncated mampu mengatasi permasalahan pola data yang menunjukkan naik turunnya perubahan pola data yang tajam sehingga dapat dihasilkan estimasi kurva yang relatif mulus (Eubank, 1999). Beberapa penelitian sebelumnya yang menggunakan model regresi nonparametrik spline truncated pernah dilakukan oleh Pratiwi (2017); Suparti, dkk., (2018) ; Yanthi dan Budiantara (2016) dan Astiti, dkk., (2016).

Pada penelitian ini, akan dilakukan simulasi dengan membentuk pola hubungan antara variabel respon dan prediktor yang tidak memiliki pola tertentu. Bentuk pola hubungan tersebut akan dimodelkan dengan dua pendekatan, yaitu pendekatan regresi parametrik (linear, kuadrat, dan kubik) dan pendekatan regresi nonparametrik yaitu spline truncated. Tujuan dalam penelitian ini adalah menunjukkan bahwa pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi sehingga mampu menyesuaikan sendiri bentuk estimasi kurva regresi tanpa dipengaruhi oleh subyektifitas peneliti (Eubank, 1999).

2. METODOLOGI PENELITIAN

Langkah-langkah yang digunakan pada penelitian ini sebagai berikut:

1. Mendefinisikan model regresi nonparametrik spline truncated $y_i = m(x_i) + \varepsilon_i$.
2. Membangkitkan data sebanyak n , dimana n pada penelitian ini adalah 500.

3. Membangkitkan variabel prediktor x_i secara independen dari distribusi Uniform, sehingga $x_i \sim U(0,1)$ sebanyak n .
4. Menetapkan bentuk kurva regresi $m(x_i)$ dengan fungsi $m(x) = \frac{\sin(2\pi x^2)^5}{\sin(\pi x^3)}$.
5. Membangkitkan error secara independen dari distribusi Normal $(0, \sigma^2)$ dengan $\sigma^2 = 0.1$ sehingga $\varepsilon_i \sim N(0,0.1)$ sebanyak n .
6. Maka akan kita dapatkan variabel respon dari $y_i = m(x_i) + \varepsilon_i$.
7. Membuat diagram pencar antara variabel prediktor x_i terhadap variabel respon y_i .
8. Memodelkan variabel respon y_i dengan variabel prediktor x_i dengan pendekatan regresi parametrik (linear, kuadratik dan kubik).
 Linear: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
 Kuadratik: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$
 Kubik: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$
9. Memodelkan variabel respon y_i dengan variabel prediktor x_i dengan pendekatan regresi nonparametrik spline truncated. Banyaknya titik knot yang akan dicobakan adalah 1 hingga 4 titik knot.

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

dengan fungsi truncated nya:

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

Menentukan titik knot yang optimal pada pendekatan regresi nonparametrik spline truncated menggunakan Generalized Cross Validation (GCV). Nilai GCV yang terkecil menghasilkan titik knot yang optimal.

$$GCV(k) = \frac{MSE(k)}{n^{-1} \text{trace}[\mathbf{I} - \mathbf{H}(\mathbf{K})]^2}$$

10. Menghitung koefisien determinasi untuk masing-masing proses pemodelan yang dilakukan.

$$R^2 = 1 - \frac{SSE}{SST}$$

dengan SSE adalah $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ dan SST

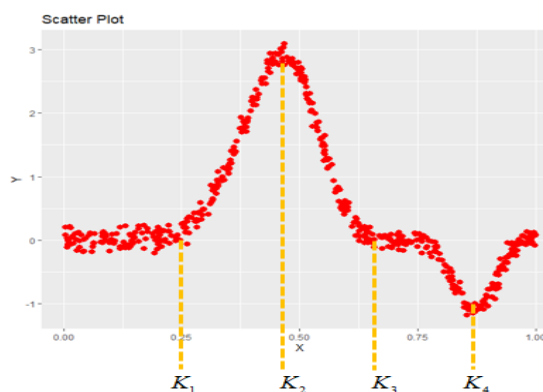
$$\text{adalah } \sum_{i=1}^n (y_i - \bar{y})^2.$$

11. Menentukan model terbaik dalam mengestimasi bentuk kurva regresi yang terbentuk berdasarkan nilai koefisien determinasi (R^2).

3. HASIL DAN PEMBAHASAN

3.1. Hasil Analisis

Diagram pencar antara variabel respon y_i dan variabel prediktor x_i dengan fungsi trigonometri $m(x) = \frac{\sin(2\pi x^2)^5}{\sin(\pi x^3)}$ disajikan pada Gambar 1.



Gambar 1. Diagram pencar hubungan antara variabel respon dan prediktor

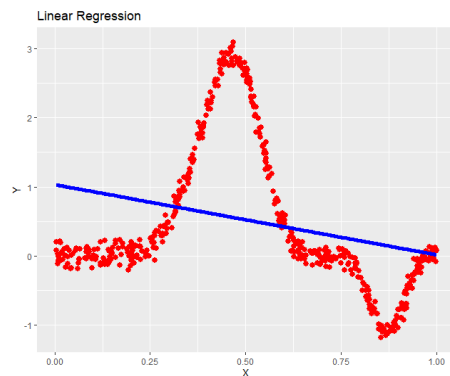
Gambar 1 menunjukkan bahwa pola hubungan antara variabel prediktor dan respon pada simulasi ini berubah-ubah pada sub-sub interval tertentu. Terlihat pola data yang cenderung naik secara lambat pada bagian awal $x < K_1$, kemudian naik secara tajam hingga menuju puncaknya pada $K_1 \leq x < K_2$. Selanjutnya terjadi perubahan pola data, dimana turun curam hingga titik tertentu yaitu pada interval $K_2 \leq x < K_3$, dan mengalami penurunan secara lambat pada interval $K_3 \leq x < K_4$, kemudian naik pada bagian akhir yaitu $x \geq K_4$.

Selanjutnya akan dilakukan pemodelan untuk mengestimasi bentuk kurva regresi. Metode estimasi yang digunakan pada

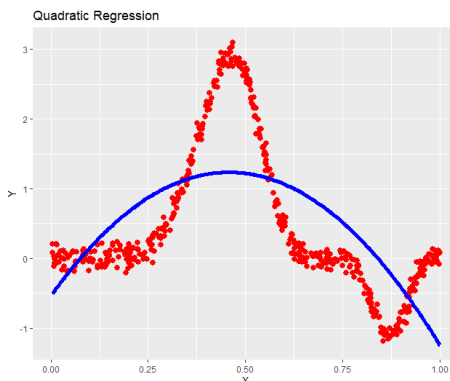
penelitian ini adalah Ordinary Least Squares (OLS).

a. Pendekatan Regresi Parametrik

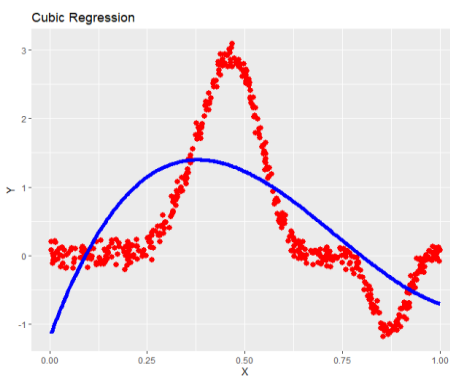
Pendekatan dengan regresi parametrik sangat populer saat ini dan familiar, hal ini dikarenakan pendekatan ini memiliki sifat yang sangat baik dari sisi statistika inferensi. Adapun hasil pemodelan data simulasi dengan pendekatan ini ditampilkan pada Gambar 2.



(a) Regresi Linear



(b) Regresi Kuadratik



(c) Regresi Kubik

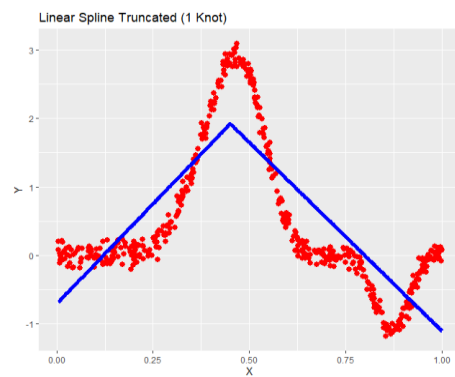
Gambar 2. Hasil estimasi kurva regresi dengan pendekatan parametrik

Gambar 2 menunjukkan bahwa pendekatan dengan menggunakan regresi parametrik secara linier, kuadratik, dan kubik

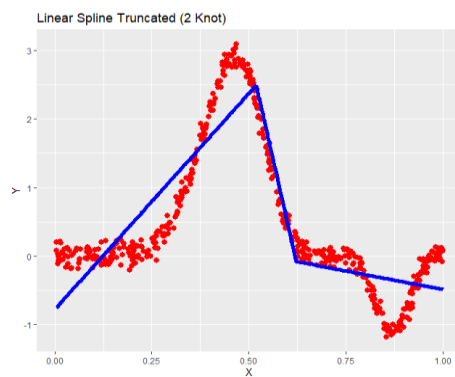
kurang tepat. Hal ini disebabkan karena belum menggambarkan dengan baik bentuk pola data yang sebenarnya antara variabel respon dan variabel prediktor.

b. Pendekatan Regresi Nonparametrik

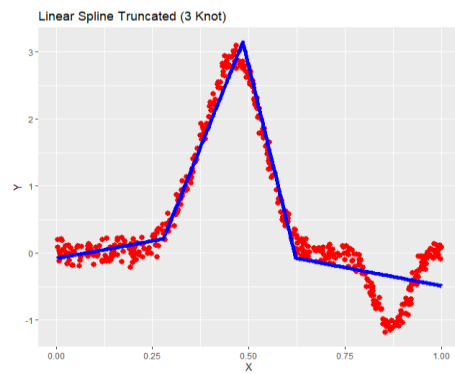
Kelemahan dari pendekatan parametrik adalah bersifat kaku dan memiliki asumsi mengenai bentuk pola hubungan antara variabel prediktor dan variabel respon. Dalam prakteknya pada kasus real, pola hubungan antara variabel respon dan prediktor tidak selalu memenuhi asumsi yang mendasari pendekatan parametrik. Hasil pemodelan data simulasi dengan pendekatan regresi nonparametrik spline truncated ditampilkan pada Gambar 3.



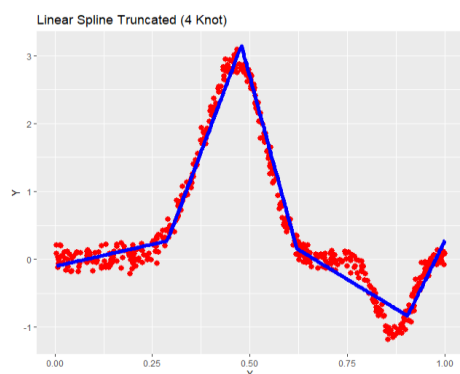
(a) Spline Truncated 1 Knot



(b) Spline Truncated 2 Knot



(c) Spline Truncated 3 Knot



(d) Spline Truncated 4 Knot

Gambar 3. Hasil estimasi kurva regresi dengan pendekatan nonparametrik

Berdasarkan Gambar 3(d), terlihat jika pendekatan dengan menggunakan regresi nonparametrik, khususnya spline truncated lebih mendekati bentuk kurva regresi yang sebenarnya antara variabel respon dengan variabel prediktor.

3.2. Pembahasan

Berdasarkan data simulasi dan setelah dilakukan estimasi kurva regresi dengan menggunakan dua pendekatan, yaitu pendekatan regresi parametrik dan nonparametrik maka diperoleh hasil nilai koefisien determinasi (R^2) sebagai berikut:

Tabel 1. Koefisien Determinasi untuk setiap hasil pemodelan

Pendekatan	Kurva Regresi	Koefisien Determinasi
Regresi Parametrik	Linear Regression	7.60%
	Quadratic Regression	45.30%
	Cubic Regression	50.50%
Regresi Nonparametrik	Linear Spline Truncated 1 Knot	67.88%
	Linear Spline Truncated 2 Knot	83.74%
	Linear Spline Truncated 3 Knot	93.99%
	Linear Spline Truncated 4 Knot	97.41%

Ukuran kebaikan estimasi kurva regresi pada penelitian ini adalah Koefisien Determinasi (R^2). Dengan pendekatan regresi parametrik, diperoleh hasil yang tidak cukup baik. Nilai

Koefisien determinasi tertinggi diperoleh dengan bentuk kurva regresi adalah Cubic Regression atau polinomial derajat 3, dengan nilai R^2 hanya sebesar 50.50%. Dapat diingat kembali bahwa pendekatan parametrik bersifat kaku dan memiliki asumsi mengenai bentuk pola hubungan antara variabel respon dan variabel prediktor. Tentunya dengan pola data seperti pada Gambar 1, maka pendekatan parametrik tidak akan cukup baik untuk mengestimasi bentuk kurva regresinya.

Selain itu bentuk polinomial dalam regresi parametrik seperti Quadratic Regression atau polinomial derajat 2 dan Cubic Regression atau polinomial derajat 3 memiliki osilasi yang rendah, sehingga akan sangat sulit untuk menghampiri bentuk kurva regresi yang mempunyai ketajaman lokal.

Dengan pendekatan regresi nonparametrik, diperoleh hasil yang cukup baik. Pada penelitian ini, titik knot yang akan dicobakan adalah dari 1 hingga 4 titik knot. Hal ini dikarenakan, peneliti menduga pola perubahan data atau perilaku data dari data simulasi mengalami 4 perubahan, sehingga diperlukan 4 titik knot untuk menangkap pola perubahan tersebut. Selanjutnya, pada penelitian ini digunakan metode GCV untuk mendapatkan lokasi titik knot optimal.

Sebagai catatan, banyaknya titik knot yang akan dicobakan perlu didasari dari banyaknya perubahan pola data. Apabila terlalu banyak titik knot yang digunakan, maka dampaknya akan terjadi pada model yang memuat banyak sekali parameter, sehingga model tersebut tidak parsimoni.

Berdasarkan hasil analisis dengan pendekatan regresi nonparametrik, nilai R^2 terbesar didapatkan saat banyaknya titik knot yang digunakan adalah 4. Hal ini sesuai dengan dugaan awal, bahwa perilaku data dari data simulasi mengalami 4 perubahan dengan nilai R^2 sebesar 97.41%. Dari Gambar 3, dapat terlihat bahwa kurva regresi yang didapat dari pendekatan regresi nonparametrik spline truncated cenderung mengikuti pola perubahan data. Hal ini sesuai dengan sifat regresi nonparametrik yang menyatakan bahwa

pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi.

4. KESIMPULAN

Berdasarkan hasil simulasi dan analisis yang dilakukan, terbukti jika pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva regresi tertentu, sehingga memberikan fleksibilitas yang tinggi. Pada kasus ini, estimator spline truncated memiliki kemampuan yang sangat baik dalam menangani data yang perilakunya berubah-ubah. Dengan menggunakan titik knot, kurva regresi dengan spline truncated mampu mengatasi permasalahan pola data yang menunjukkan naik turunnya perubahan pola data yang tajam sehingga dapat dihasilkan estimasi kurva yang relatif mulus.

5. REFERENSI

- Astiti, D. A. W., Sumarjaya, I. W., dan Susilawati, M. (2016). Analisis Regresi Nonparametrik Spline Multivariat untuk Pemodelan Indikator Kemiskinan di Indonesia. *E-Jurnal Matematika*, 05(03), pp. 111-116.
- Budiantara, I. N. (2019). Regresi Nonparametrik Spline Truncated. Surabaya: ITS Press.
- Dani, A. T. R., Adrianingsih, N. Y., dan Ainurrochmah, A. (2020). Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi. *Jambura Journal of Probability and Statistics*, 01(02), pp. 98-106.
- Eubank, R. L. (1999). Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker Inc.
- Nisa, K., Herawati, N., dan Setiawan, E. (2018). Analisis Regresi Nonparametrik dengan Teknik Smoothing. *Seminar dan Rapat Tahunan (Semirata) Medan*, pp. 01-20.
- Pratiwi, L. P. S. (2017). Pemodelan Spline Truncated dalam Regresi Nonparametrik Birespon. *Makalah Konferensi Nasional Sistem dan Informatika*, STMIK STIKOM Bali.
- Rahasia, Z., Resmawan, dan Isa, D. R. (2020). Pemodelan data Time Series dengan Pendekatan Regresi Nonparametrik B-Spline. *Aksioma: Jurnal Matematika dan Pendidikan Matematika*, 11(01), pp. 09-16.
- Saputro, D. R. S., Sukmayanti, A. dan Widyaningsih, P. (2019). The Nonparametric Regression model Using Fourier Series Approximation and Penalized Least Squares (PLS). 2nd *International Conference on Statistics, Mathematics, Teaching and Research, Journal of Physics: Conf. Series*, pp. 01-04.
- Suparti dan Prahutama, A. (2016). Pemodelan Regresi Nonparametrik menggunakan Pendekatan Polinomial Lokal pada Beban Listrik di Kota Semarang. *Media Statistika*, 09(02), pp. 85-93.
- Suparti, Prahutama, A., dan Santoso, R. (2018). Mix Local Polynomial and Spline Truncated: The Development of Nonparametric Regression Model. *ISNPINSA-7, Journal of Physics: Conf. Series*, pp. 01-07.

Suparti, Santoso, R., Prahutama, A., dan Devi, A. R. (2018). Regresi Nonparametrik. Ponorogo: Wade Group.

Yanthi, N. P. D. dan Budiantara, I. N. (2016). Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia menggunakan Regresi Nonparametrik Spline di Jawa Tengah. *Jurnal Sains dan Seni ITS*, 05(02), pp. 157-162.