

Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus: Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar)

Rahmat Syam¹, Wahidah Sanusi¹, dan Rabiatal Adawiyah^{1,a)}

¹Jurusan Matematika FMIPA Universitas Negeri Makassar, 90224

^{a)}rabiabial19@gmail.com

Abstrak. Pendekatan nonparametrik merupakan suatu pendekatan yang digunakan apabila bentuk hubungan antara variabel respon dan variabel prediktornya tidak diketahui atau tidak adanya informasi mengenai bentuk fungsi regresinya. Spline merupakan suatu teknik yang dilakukan untuk mengestimasi parameter dalam regresi nonparametrik. Penelitian ini bertujuan untuk mengetahui model hubungan antara berat badan lahir rendah dan faktor-faktor yang mempengaruhi berdasarkan model spline. Faktor-faktor tersebut adalah usia ibu, usia kehamilan, dan jarak kehamilan. Data tersebut diperoleh dari rumah sakit ibu dan anak Siti Fatimah Makassar tahun 2017. Dimana untuk mendapatkan model spline terbaik langkah awal yang dilakukan adalah menentukan knot dengan nilai Generalized Cross Validation (GCV) yang minimum. Berdasarkan penelitian yang telah dilakukan, dua variabel dinyatakan berpengaruh terhadap berat badan lahir rendah yaitu usia ibu, dan usia kehamilan. Model regresi nonparametrik dengan pendekatan Spline yang terbentuk memiliki koefisien determinasi sebesar 78,19%, serta nilai GCV dengan tiga titik knot yaitu 0,0117.

Kata kunci: Regresi Nonparametrik, Spline, Berat Badan Lahir Rendah, Generalized Cross Validation

Abstract. The non-parametric approach is an approach that is used if the form of the relationship between the response variable and the predictor variable is unknown or the absence of information about the shapes of regression functions. The Spline is a technique performed to estimate the parameters in the nonparametric regression. This study aims to determine the model of the relationship between low birth weight and the factors that affect the based on the spline model. Such factors are maternal age, gestational age, and pregnancy distance. The Data is obtained from the mother and child hospital Siti Fatimah Makassar 2017. Where to get a spline model best the initial step is to determine the knots with the value of the Generalized Cross Validation (GCV) which is a minimum. Based on the research that has been done, the two variables stated effect against low birth weight, namely age of mother, and gestational age. Nonparametric regression Model with the approach of the Spline that is formed has a coefficient of determination of 78.19 to%, as well as the value of the GCV with a three-point knot that is 0.0117.

Keyword : Nonparametric Regression, Spline, Low Birth Weight, Generalized Cross Validation

PENDAHULUAN

Analisis regresi merupakan metode analisis data yang telah diterapkan secara luas pada berbagai bidang penelitian, sebagai contoh penelitian-penelitian dalam ilmu pengetahuan terapan seperti sosial, biologi, kesehatan, dan ekonomi. Metode analisis regresi mempelajari bagaimana menentukan bentuk sebuah model atau hubungan antara variabel-variabel dari sekumpulan data untuk menjelaskan hubungan sebab akibat atau keterkaitan antarkejadian. Dalam hal ini, analisis regresi digunakan untuk memodelkan atau mencari pola hubungan antara satu atau lebih variabel prediktor (variabel independen) dengan satu atau lebih variabel respons (variabel dependen) (Budiantara, 2000).

Terdapat dua teknik pendekatan yang dapat digunakan dalam analisis regresi untuk mengestimasi kurva regresi yaitu pendekatan regresi parametrik dan regresi nonparametrik. Dalam model regresi parametrik terdapat asumsi yang harus terpenuhi yaitu bentuk kurva regresinya (pola hubungan antara variabel respons dan variabel prediktor) diketahui (Eubank, 1999). Sedangkan Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Winarti & Sony, 2010). Dalam hal ini, teknik pendekatan regresi nonparametrik bisa menjadi alternatif karena penggunaannya tidak terikat pada asumsi-asumsi seperti dalam regresi parametrik. Pendekatan nonparametrik digunakan ketika informasi mengenai bentuk kurva regresi terbatas atau tidak ada asumsi tentang bentuk kurva regresi.

Pendekatan regresi nonparametrik telah banyak dikembangkan antara lain menggunakan spline, kernel, polinomial lokal, wavelet, dan fourier (Eubank, 1999). Salah satu model regresi dengan pendekatan nonparametrik yang sangat sering digunakan untuk melakukan estimasi terhadap kurva regresi adalah regresi spline. Salah satu model regresi dengan pendekatan nonparametrik yang sangat sering digunakan untuk melakukan estimasi terhadap kurva regresi adalah regresi spline. Spline merupakan suatu potongan-potongan polinom yang memiliki sifat tersegmen pada selang k yang terbentuk pada titik-titik knot (Wang & Yang, 2009). Regresi Nonparametrik dengan pendekatan spline dapat pula diaplikasikan ke dalam kehidupan. Seperti pada penelitian ini yang mengaplikasikan Regresi Nonparametrik dengan pendekatan Spline dalam kasus Berat Badan Lahir Rendah. Pada Negara berkembang, BBLR masih menjadi salah satu permasalahan defisiensi zat gizi. Bayi Berat Badan Lahir Rendah adalah bayi yang dilahirkan dengan berat badan kurang dari 2.500 gram, tanpa memandang masa gestasi (Kosim, 2012).

Penyelesaian Model Regresi Nonparametrik telah dibahas oleh beberapa peneliti sebelumnya. Dewi & Budiantara (2012) membahas tentang faktor-faktor yang mempengaruhi angka gizi buruk di Jawa Timur dengan pendekatan regresi nonparametrik spline. Selanjutnya Alfiani (2014) membahas tentang Estimator Polinomial Lokal Kernel Pada Kasus Pertumbuhan Balita untuk menganalisis pengaruh antara berat badan dan umur. Selanjutnya Ismail (2016) membahas tentang menentukan fungsi kernel yang terbaik untuk memodelkan hubungan antara umur Ibu dan Berat Badan Bayi Lahir.

Oleh karena itu, penelitian ini akan mengkaji mengenai pemodelan regresi nonparametrik dengan kasus berat badan lahir rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar. Faktor-faktor penyebab yaitu dari umur ibu, usia kehamilan dan jarak kehamilan. Berdasarkan faktor-faktor inilah yang nantinya akan dijadikan sebagai kriteria untuk mencari model terbaik dengan menggunakan pendekatan Spline.

METODE PENELITIAN

Regresi Nonparametrik

Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Winarti & Sony, 2010). Model regresi nonparametrik adalah sebagaimana persamaan (1).

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, 3, \dots, n \quad (1)$$

dimana

y_i : variabel respon

$f(x_i)$: fungsi *smooth* yang tidak diketahui ke- i

x_i : variabel prediktor

ε_i : error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varians σ^2

Regresi Spline

Spline merupakan potongan polinomial (piecewise polynomial) tersegmen yang memiliki sifat fleksibilitas. Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda disebut (Winarti & Sony, 2010). Adapun model dari regresi spline adalah sebagaimana persamaan (2).

$$f(x_i) = \sum_{j=0}^m \beta_j x_i^j + \sum_{l=1}^p \beta_{(m+l)} (x_i - k_l)^m \quad (2)$$

dengan $(x_i - k_l)^m = \begin{cases} (x_i - k_l)^m, & x_i \geq k_l \\ 0 & ; x_i < k_l \end{cases}$

dimana

- $f(x_i)$: fungsi regresi spline
- k_1, k_2, \dots, k_k : titik knot
- x : variabel prediktor
- β : konstanta

Estimasi Parameter

Untuk mengestimasi parameter dapat menggunakan Metode *Maximum Likelihood Estimator* (MLE), dengan demikian *Maximum Likelihood Estimator* (MLE) adalah suatu teknik yang sering digunakan pada model parametrik baik untuk mencari penduga parameter maupun konstruksi statistika uji (Supranto, 1989).

fungsi *likelihood*nya dapat dituliskan sebagaimana persamaan (3).

$$L(x_1, x_2, \dots, x_n; \mu) = \prod_{i=1}^n \left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \right) \quad (3)$$

Fungsi *log-likelihood* dari persamaan (3) sebagaimana persamaan (4).

$$\ln[L(\mu)] = \ln \left\{ \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \right\} \quad (4)$$

Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pada data. Model regresi spline terbaik tergantung pada titik knot optimal (Eubank, 1998). Metode untuk mencari titik knot optimal yang sering dipakai adalah *Generalized Cross Validation* (GCV) dan *Mean Squared Error* (MSE). Titik knot optimal diperoleh dari nilai GCV yang minimum.

1. Mean Squared Error (MSE)

Kriteria sederhana yang digunakan sebagai ukuran kinerja atas penaksir yang baik adalah *Mean Squared Error* (MSE) adalah

$$MSE(k) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$$

dimana:

- x_i : variabel independen/prediktor
- y_i : variabel dependen/respon
- n : banyaknya pengamatan

2. Generalized Cross Validation (GCV)

Kriteria lain yang dapat digunakan sebagai ukuran kinerja atas penaksir yang baik adalah *Generalized Cross Validation* adalah

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

dimana:

- I : matriks identitas
- n : jumlah pengamatan
- $A(k)$ adalah matriks $X(X^T V X)^{-1} X^T V$ (Eubank, 1998).

Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline

1. Uji Serentak

Pengujian parameter model secara serentak merupakan uji parameter kurva regresi secara simultan dengan menggunakan uji F .

Hipotesis pada uji serentak sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{m+p} = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0; k = 1, 2, \dots, m + p$$

Nilai $m + p$ merupakan banyak parameter dalam regresi nonparametrik spline kecuali β_0 .
Statistik uji:

$$F_{hitung} = \frac{MSR}{MSE}$$

dengan

$$MSR = \frac{SSR}{df_{reg}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{m+p} \quad \text{dan} \quad MSE = \frac{MSE}{df_{error}} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-(m+p)-1}$$

H_0 ditolak apabila $F_{hitung} > F_{n-(m+p)-1}$ atau $p\text{-value} < \alpha$.

2. Uji Individu

Pengujian secara individu dilakukan untuk mengetahui apakah parameter secara individual mempunyai pengaruh yang signifikan terhadap variabel respon.

Hipotesis pada uji individu sebagai berikut:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0, k = 1, 2, \dots, m + p$$

Pengujian secara individu dilakukan dengan menggunakan uji t (Draper & Smith, 1992).
Statistik uji:

$$t_{hitung} = \frac{\hat{\beta}_k}{\sqrt{\text{var}(\hat{\beta}_k)}}$$

dengan

$$\text{var}(\hat{\beta}_k) = \text{diag}[(\mathbf{X}'\mathbf{X})^{-1}\hat{\sigma}^2]$$

dimana $\hat{\sigma}^2$ merupakan MSE . H_0 ditolak apabila $|t_{hitung}| > t_{(n(m+p)-1)}$ atau $p\text{-value} < \alpha$.

Uji Asumsi Residual

Data yang akan dianalisis dengan menggunakan regresi nonparametrik spline, harus memenuhi asumsi residual identik dan berdistribusi normal.

1. Uji Asumsi Identik

Asumsi identik terpenuhi apabila varians antar residual sama, atau tidak terjadi heteroskedastisitas (Gujarati, 2003).

Uji identik dapat menggunakan uji Glejser. Hipotesis yang digunakan adalah sebagai berikut:

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

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2; i = 1, 2, \dots, n$$

Statistik uji:

$$F_{hitung} = \frac{\frac{\sum_{i=1}^n (|\varepsilon_i| - |\bar{\varepsilon}|)}{v-1}}{\frac{\sum_{i=1}^n (|\varepsilon_i| - |\hat{\varepsilon}_i|)}{n-v}}$$

H_0 ditolak apabila $F_{hitung} > F_{(v-1, n-v)}$ atau $p\text{-value} < \alpha$, dimana nilai v adalah banyaknya parameter model Glejser.

2. Uji Distribusi Normal

Pengujian asumsi distribusi normal dilakukan untuk mengetahui apakah residual berdistribusi normal. Pengujian ini dilakukan dengan menggunakan uji *Kolmogorov-Smirnov*.

Hipotesis pada uji distribusi normal sebagai berikut:

$$H_0 : F_0(x) = F(x) \quad (\text{Residual berdistribusi normal})$$

$$H_1 : F_0(x) \neq F(x) \quad (\text{Residual tidak berdistribusi normal})$$

Statistik uji:

$$D = \max |F_O - S_N(x)|$$

Daerah penolakan H_0 jika $|D| > q_{(1-\alpha)}$ dimana nilai $q_{(1-\alpha)}$ berdasarkan tabel *Kolmogorov-Smirnov*

Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) adalah alat untuk mengukur proporsi keragaman atau variansi total disekitar nilai tengah y yang dapat dijelaskan oleh model regresi. Secara umum semakin besar nilai R^2 , maka semakin baik pula model yang didapatkan karena mampu menjelaskan lebih banyak data (Draper & Smith, 1992).

Rumus R^2 dapat ditulis sebagai berikut:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \times 100\%$$

dengan

\hat{y} : nilai estimasi peubah respon ke- i

\bar{y} : rata-rata peubah respon

y_i : nilai peubah respon ke- i

DATA PENELITIAN

Penelitian ini merupakan penelitian terapan menggunakan metode analisis regresi nonparametrik dengan pendekatan Spline. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Rumah Sakit Ibu dan Anak Siti Fatimah Makassar tahun 2017. Penelitian ini terdapat 52 data yang digunakan. Variabel yang digunakan dalam penelitian ini terdiri dari 2, variabel dependen yaitu berat badan lahir rendah sedangkan variabel independen yaitu usia ibu, usia kehamilan dan jarak kehamilan.

HASIL PENELITIAN

Estimasi Regresi Nonparametrik dengan Pendekatan Spline pada Metode Maximum Likelihood Estimation

Berdasarkan model persamaan regresi nonparametrik adalah sebagaimana persamaan (5)

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, 3, \dots, n \quad (5)$$

dimana

y_i : variabel respon

$f(x_i)$: fungsi *smooth* yang tidak diketahui ke- i

x_i : variabel prediktor

ε_i : error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varians σ^2

Secara umum fungsi spline berorde m dengan titik knot k_1, k_2, \dots, k_n dapat dinyatakan sebagaimana persamaan (6)

$$f(x_i) = \sum_{h=0}^m \beta_j x_i^j + \sum_{l=1}^p \beta_{m+l} (x_i - k_l)^m \quad (6)$$

Regresi nonparametrik spline adalah suatu pendekatan dalam analisis regresi dengan pola hubungan (fungsi regresinya) yang diasumsikan tidak diketahui dan didekati dengan fungsi spline. Model regresi nonparametrik spline orde ke- m adalah sebagaimana persamaan (7)

$$y_i = \sum_{h=0}^m \beta_j x_i^j + \sum_{l=1}^p \beta_{m+l} (x_i - k_l)^m + \varepsilon_i \quad (7)$$

Dari persamaan (7) diketahui bahwa $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ merupakan variabel acak berdistribusi normal dengan $\varepsilon \sim N(X\beta, \sigma^2 I)$ dengan $X = (x_{i1}, x_{i2}, \dots, x_{nk})$ dan $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ dimana

$i = 1, 2, 3, \dots, n$ dan I menyatakan matriks ukuran $n \times n$. Sehingga fungsi kepadatan peluang bersama dari error adalah sebagaimana persamaan (8)

$$L(\beta, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \right) \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ = \left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} \right) \exp\left(-\frac{1}{2\sigma^2}(Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta)\right) \quad (8)$$

Dari fungsi *likelihood* pada persamaan (8), selanjutnya untuk mendapatkan estimasi parameter β dan σ^2 dengan metode *maximum likelihood estimation* dengan cara melogaritmakan persamaan (8) tersebut sebagaimana persamaan (9)

$$\ln L(\beta, \sigma^2) = \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}(Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta)\right)\right) \\ = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}Y^T Y + \frac{1}{2\sigma^2}2\beta^T X^T Y - \frac{1}{2\sigma^2}\beta^T X^T X \beta \quad (9)$$

Untuk mengestimasi parameter β yang dinotasikan dengan $\hat{\beta}$ yaitu dengan memaksimalkan persamaan (9) terhadap β artinya mendiferensialkan persamaan (9) terhadap β sebagaimana persamaan (10)

$$\frac{\partial \ln L(\beta, \sigma^2)}{\partial \beta} = \frac{\partial\left(\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}Y^T Y + \frac{1}{2\sigma^2}2\beta^T X^T Y - \frac{1}{2\sigma^2}\beta^T X^T X \beta\right)}{\partial \beta} \\ = \frac{1}{\sigma^2}X^T Y - \frac{1}{\sigma^2}(X^T X \beta) \quad (10)$$

Maka estimasi model regresi nonparametrik dengan pendekatan spline yang diperoleh adalah:

$$\hat{y} = X\hat{\beta} \\ = X(X^T X)^{-1}X^T Y$$

Analisis Karakteristik Variabel

Langkah awal sebelum dilakukan proses pengolahan data adalah melakukan deskriptif statistik. Karakteristik dari masing-masing variabel prediktor dapat diinformasikan melalui deskriptif statistik yaitu berat badan lahir rendah dan semua variabel bebas yang diduga berpengaruh signifikan terhadap berat badan lahir rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar tahun 2017. Deskriptif statistic yang digunakan berupa nilai maksimum, minimum, dan rata-rata dari setiap variabel seperti yang ditampilkan dalam Tabel 1.

Tabel 1. Hasil Analisis Statistika Deskriptif

Variabel	Mean	Varians	Minimum	Maksimum
Y	2.115673	0.057497	1.5	2.48
X_1	39.17308	5.440045	36	47
X_2	0.698317	0.000638	0.666667	0.75
X_3	1.445513	0.014402	1.25	1.75

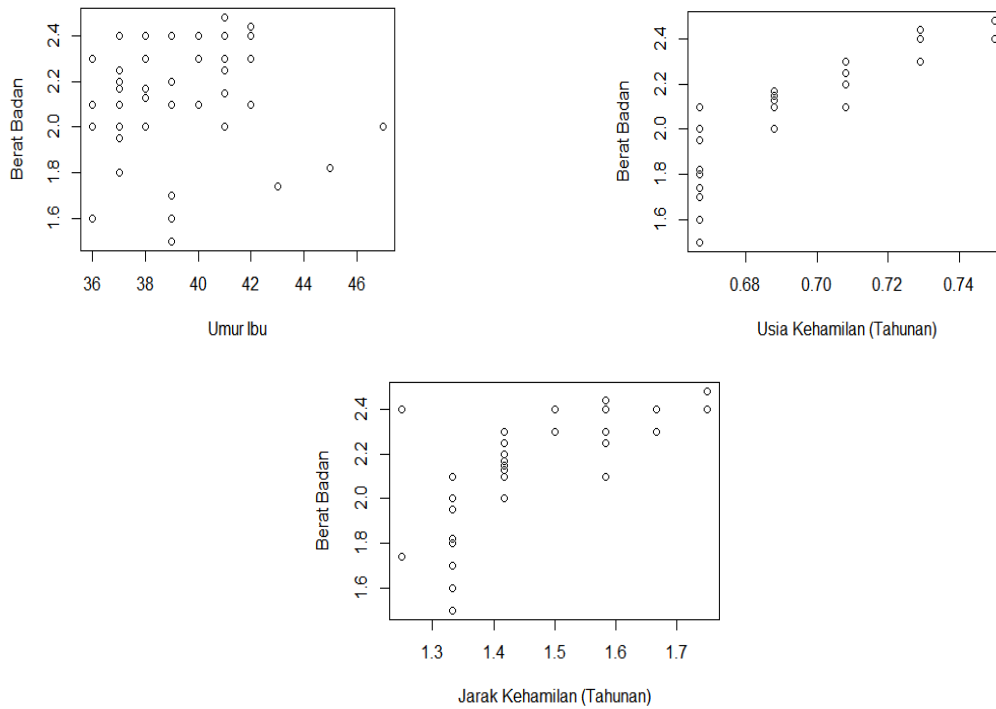
Pemilihan Titik Knot Terbaik

TABEL 2. Perbandingan Nilai GCV dan MSE

Model	GCV	MSE
1 Titik Knot	0.01543455	0.0130541
2 Titik Knot	0.01361829	0.0130541
3 Titik Knot	0.01172147	0.01229278

Dari Tabel 2 dapat diketahui bahwa nilai GCV dan MSE paling minimum dihasilkan oleh model regresi nonparametrik spline dengan tiga titik knot yaitu sebesar 0.01172147.

Pola Hubungan Faktor-faktor yang diduga Mempengaruhi Berat Badan Lahir Rendah



GAMBAR 1. Scatterplot antara BBLR dengan 3 variabel yang diduga berpengaruh

Dari Gambar 1 diketahui bahwa pola hubungan antara variabel umur ibu, usia kehamilan dan jarak kehamilan terhadap berat badan lahir rendah, menunjukkan pola hubungan yang tidak membentuk suatu pola tertentu, sehingga estimasi model menggunakan regresi nonparametrik.

Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline Tiga Variabel Prediktor

1. Pengujian serentak tiga variabel prediktor

TABEL 3. Analisis Ragam Uji Serentak Model Tiga Variabel

Sumber Variasi	Derajat Bebas (db)	Jumlah Kuadrat (JK)	Rataan Jumlah Kuadrat (RJK)	F Hitung	<i>p-value</i>
Regresi	39	145,1992	3,7230	8,19	0,001
Residual	12	5,4517	0,4543		
Total	51	78,6510			

Dari Tabel 3 dapat diketahui bahwa statistik uji menggunakan F_{hitung} sebesar 8,19 dengan *p-value* sebesar 0,00138. Pada tingkat signifikan (α) 5%, *p-value* bernilai kurang dari α , sehingga tolak H_0 .

2. Pengujian individu tiga variabel prediktor

TABEL 4. Estimasi Parameter Regresi Tiga Variabel

Variabel	Parameter	Estimasi	T	p-value	Ket.
	β_0	8,556	5,195	0,000	Signifikan
X_1	β_1	0,001	0,341	0,739	Tidak Signifikan
	β_2	0,229	4,253	0,001	Signifikan
	β_3	-0,354	-4,304	0,001	Signifikan
	β_4	0,126	3,967	0,002	Signifikan
X_2	β_5	-1,687	-2,683	0,019	Signifikan
	β_6	-0,001	-2,681	0,010	Signifikan
	β_7	-0,001	-2,681	0,010	Signifikan
	β_8	-0,001	-2,681	0,010	Signifikan
X_3	β_9	0,086	0,319	0,751	Tidak Signifikan
	β_{10}	0,000	0,122	0,903	Tidak Signifikan
	β_{11}	0,000	0,122	0,903	Tidak Signifikan
	β_{12}	0,000	0,122	0,903	Tidak Signifikan

Dari Tabel 4 dapat diketahui bahwa dari ketiga variabel prediktor, dua variabel prediktor mempunyai parameter yang signifikan terhadap model, sehingga variabel umur ibu (X_1) dan usia kehamilan (X_2) berpengaruh secara signifikan terhadap berat badan lahir rendah di Rumah Sakit Siti Fatimah Makassar.

Pengujian Asumsi Residual

1. Pengujian asumsi identik

TABEL 5. Analisis Ragam Uji Glejser

Sumber Variasi	Derajat Bebas (db)	Jumlah Kuadrat (JK)	Rataan Jumlah Kuadrat (RJK)	F Hitung	p-value
Regresi	39	2,333883	0,059843	2,776	0,087
Residual	12	0,258609	0,021551		
Total	51	1,449222			

Dari Tabel 5 dapat diketahui bahwa nilai F_{hitung} uji *Glejser* sebesar 2,776 dengan p-value sebesar 0,087, yaitu lebih besar dari α (0,05). Sehingga dapat diputuskan bahwa gagal tolak H_0 . Jadi dapat diartikan bahwa tidak terjadi heteroskedastisitas.

Pemilihan model terbaik

Berikut merupakan model regresi nonparametrik spline terbaik untuk dilakukan estimasi parameter menggunakan *Maksimum Likelihood Estimation* (MLE).

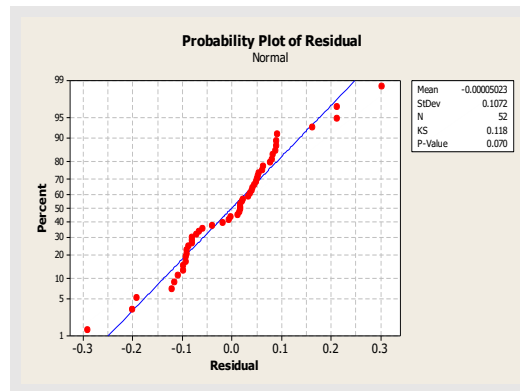
$$\hat{y} = 8,556 + 0,001x_1 + 0,229(x_1 - 36,673)^1 - 0,354(x_1 - 38,918)^1 + 0,126(x_1 - 42,061)^1 - 1,687x_2 - 0,001(x_2 - 0,672)^1 - 0,001(x_2 - 0,689)^1 - 0,001(x_2 - 0,713)^1$$

Koefisien Determinasi

Koefisiensi determinan (R^2) digunakan untuk mengukur kebaikan atau kesesuaian (*goodness of fit*) suatu model persamaan regresi. Nilai R^2 menyatakan proporsi variasi dalam variabel dependen dapat dijelaskan oleh variabel independen. Berdasarkan perhitungan didapatkan nilai R^2 sebesar 78,19%. Hal ini berarti model regresi nonparametrik spline yang didapatkan mampu menjelaskan variabilitas

berat badan lahir rendah di Rumah Sakit Ibu dan Anak Siti Fatimah. Nilai tersebut mendekati 100%, sehingga model sudah cukup baik.

2. Pengujian Normalitas Residual



GAMBAR 2. Uji Normalitas Residual *Kolmogorov-Smirnov*

Dari Gambar 2 diketahui bahwa titik-titik merah atau nilai residual berada disekitar garis biru, sehingga secara visual residual berdistribusi normal. Nilai KS yang diperoleh sebesar 0,118 dengan *p-value* lebih besar dari 0,070.

PEMBAHASAN

Penelitian sebelumnya tentang model regresi Nonparametrik dan metode Spline, telah dibahas oleh beberapa peneliti sebelumnya. Raupong (2015) membahas tentang penerapan model *B-Spline* dalam menaksir kurva regresi nonparametrik. Metode yang digunakan dalam menaksir parameter *B-Spline* adalah Metode Kuadrat Terkecil (*Ordinary Least Square*), *B-Spline* linier dengan menggunakan satu titik knot merupakan model yang terbaik, kriteria pemilihan model terbaik yaitu R^2 (Koefisien Determinasi) dan MSE (*Means Square Error*). Selanjutnya Agustien (2016) membahas tentang Pemodelan Risiko Kejadian Bayi Berat Badan Lahir Rendah Berdasarkan Pendekatan Multivariate Adaptive Regression Spline (MARS) (Studi Kasus di RSUD Haji Surabaya), metode yang digunakan dalam mengestimasi parameter pendekatan Multivariate Adaptive Regression Spline (MARS) adalah Metode Kuadrat Terkecil (*Ordinary Least Square*) dan metode yang digunakan untuk memodelkan bayi dengan Berat Badan Lahir Rendah (BBLR) merupakan *Spline* linier dengan menggunakan tiga titik knot merupakan model yang terbaik, kriteria pemilihan model terbaik yaitu R^2 (Koefisien Determinasi) dan MSE (*Means Square Error*).

Sementara untuk penelitian ini dijelaskan mengenai model regresi nonparametrik dengan pendekatan Spline untuk mencari model terbaik dengan menggunakan data Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar. Kemudian, menentukan model umum serta estimasi parameternya. Metode yang digunakan dalam mengestimasi parameter Spline adalah Maximum Likelihood Estimation (MLE) dan merupakan Spline linier dengan menggunakan tiga titik knot merupakan model terbaik untuk data Berat Badan Lahir Rendah, nilai GCV (Generalized Cross Validation) terkecil dan adapun variabel yang memberikan pengaruh signifikan adalah dua variabel dengan koefisien determinasi yang menunjukkan bahwa model yang terbentuk layak digunakan untuk memodelkan pola data.

KESIMPULAN

1. Model regresi nonparametrik dengan pendekatan spline adalah sebagai berikut:

$$y_i = \sum_{j=0}^m \beta_j x_i^j + \sum_{l=1}^p \beta_{m+l} (x_i - k_l)^m + \varepsilon_1$$

Untuk memperoleh estimasi parameter model regresi nonparametrik dengan pendekatan spline digunakan metode *Maximum Likelihood* sehingga diperoleh hasil

$$\hat{y} = X(X^T X)^{-1} X^T Y$$

2. Model regresi nonparametrik spline terbaik untuk berat badan lahir rendah di Rumah Sakit Ibu dan Anak Siti Fatimah menggunakan tiga titik knot pada setiap variabel prediktor. Nilai kebaikan model atau R^2 yang diperoleh sebesar 78,19% dengan 2 variabel prediktor yang signifikan yaitu umur ibu dan usia kehamilan. Berikut merupakan model regresi yang didapatkan.

$$\hat{y} = 8,556 + 0,001x_1 + 0,229(x_1 - 36,673)^1 - 0,354(x_1 - 38,918)^1 + 0,126(x_1 - 42,061)^1 - 1,687x_2 - 0,001(x_2 - 0,672)^1 - 0,001(x_2 - 0,689)^1 - 0,001(x_2 - 0,713)^1$$

Adapun penelitian ini hanya mengkaji tentang regresi nonparametrik dengan pendekatan spline. Pada penelitian ini hanya menggunakan orde 1 dan penelitian ini hanya terdapat tiga variabel independen, maka diharapkan kepada pembaca atau peneliti selanjutnya untuk dapat mengkaji lebih jauh tentang regresi nonparametrik dengan pendekatan spline dan di uji menggunakan orde 1, 2, dan 3 serta mencari estimasi model regresi nonparametrik dengan menggunakan metode pendekatan lainnya seperti deret fourier, histogram, dll dengan lebih banyak variabel independen.

DAFTAR PUSTAKA

- Agustien, R. A. (2016). *Pemodelan Risiko Kejadian Bayi Berat Badan Lahir Rendah Berdasarkan Pendekatan Multivariate Adaptive Regression Spline (MARS) (Studi Kasus di RSUD Haji Surabaya)*. (Skripsi tidak dipublikasikan). Program Studi Statistika Universitas Airlangga.
- Alfiani, M. L. (2014). Model Regresi Nonparametrik Berdasarkan Estimator Polinomial Lokal Kernel Pada Kasus Pertumbuhan Balita. *Jurnal Statistika*, Vol. 2, No. 1, 56-80
- Budiantara, I. N. (2000). Metode U, GML, CV, dan GCV dalam Regresi Nonparametrik Spline. *Jurnal Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 6, No. 1, 285-290.
- Dewi, R. K & Budiantara, I. N. (2012). Faktor-Faktor Yang Mempengaruhi Angka Gizi Buruk di Jawa Timur Dengan Pendekatan Regresi Nonparametrik Spline. *Jurnal Sains dan Seni ITS*, Vol. 1, No. 1.
- Draper, N. R. & Smith, R. (1992). Perbandingan Model GWR dengan FIXED dan ADAPTIVE Bandwidth untuk Persentase Penduduk Miskin di Jawa Tengah. *Jurnal Gaussian*, Vol. 5, No. 3, 535-544.
- Eubank. (1998). Penentuan Generalized Cross Validation (GCV) Sebagai Kriteria dalam Pemilihan Model Regresi B-Spline Terbaik. *Jurnal Statistika*, Vol. 2, No. 2, 121-126
- Eubank. (1999). Faktor-Faktor Yang Mempengaruhi Angka Gizi Buruk di Jawa Timur Dengan Pendekatan Regresi Nonparametrik Spline. *Jurnal Sains dan Seni ITS*, Vol. 1, No. 1, 76-88.
- Gujarati. (2003). Analisis Faktor yang Berpengaruh Terhadap Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Menggunakan Regresi Data Panel. *Jurnal Sains dan Seni ITS*, Vol. 6, No.1, 98-103
- Ismail, M. (2016). *Model Regresi Nonparametrik dan aplikasinya (Studi Kasus: Berat Badan Bayi Lahir di Rumah Sakit Ibu dan Anak Pertiwi Makassar)*. (Skripsi tidak dipublikasikan). Program Studi Matematika Universitas Negeri Makassar.
- Kosim. (2012). Faktor-faktor yang Berhubungan dengan Kejadian Berat Badan Lahir Rendah (BBLR) di Kabupaten Kudus. *Jurnal Kesehatan Masyarakat*, Vol. 5, No. 1
- Litawati & Budiantara. (2013). Pendekatan Regresi Nonparametrik Spline Untuk Pemodelan Laju Pertumbuhan Ekonomi (LPE) di Jawa Timur. *Jurnal Sains dan Seni Pomits*, Vol. 2, No. 2, 2337-3520.
- Raupong. (2015). Model B-Spline dalam Menaksir Kurva Regresi Nonparametrik. *Jurnal Matematika, Statistika, dan Komputasi*, Vol. 6, No. 1, 29-43.

- Supranto. (1989). Perbandingan Estimasi Parameter pada Distribusi Eksponensial dengan Menggunakan Metode Maksimum Likelihood dan Metode Bayesian. *Jurnal Sains Matematika dan Statistika*, Vol. 1, No. 2, 145-150
- Wang, J. & Yang, L. (2009). Polinomial Spline Kepercayaan Band untuk Regresi Kurva. *Jurnal Statistica Sinica*, Vol. 2, No. 19, 325-342.
- Winarti & Sony, S. (2010). Pendekatan Regresi Semiparametrik Spline. (Pada data nilai Ujian Nasional siswa SMKN 1 Nguling Pasuruan). *Jurnal Sains dan Seni Pomits*, Vol. 3 No. 2, 194-199.