

PEMODELAN REGRESI NONPARAMETRIK *SPLINE TRUNCATED* KASUS KEJADIAN DIARE DI PROVINSI BALI

Nadiya Yuvita Rizki^{1§}, I Gusti Ayu Made Srinadi², I Komang Gde Sukarsa³

¹Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: nadiyayuvitarizki@gmail.com]

²Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: srinadi@unud.ac.id]

³Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: gedesukarsa@unud.ac.id]

§Corresponding Author

ABSTRACT

Nonparametric regression is a flexible approach used to determine the relationship between the predictor variable and the response variable is unknown. One method that can be used to estimate nonparametric regression models is the truncated spline. The truncated spline is an effective method to estimate nonparametric regression models due to its ability to adapt to the data's characteristics through knots. Truncated spline estimates its parameters with the maximum likelihood estimator (MLE) method and finds the optimal knot points with the minimum generalized cross validation (GCV) value. This study used the truncated spline to model diarrhea cases in Bali Province, examining five variables that could affect incidence. The optimal knot points were 2-1-3-3-2 with a minimum GCV value of 67572,38. The study found that the number of clean drinking water facilities, food management places that meet health requirements, public places that meet health requirements, population density, and access to proper sanitation facilities had a significant effect on diarrhea incidence. The coefficient of determination for this model is 98,87%.

Keywords: Diarrhea, GCV, Knots, Truncated Spline Nonparametric Regression.

1. PENDAHULUAN

Pendekatan regresi nonparametrik digunakan untuk mengetahui hubungan antara kedua variabel yaitu variabel prediktor dan variabel respon yang tidak diketahui bentuk kurva regresinya. Pendekatan membebaskan data untuk menemukan bentuk penduganya sendiri (Eubank, 1988). Sehingga, kelebihan dari pendekatan ini adalah fleksibilitasnya yang tinggi dan tidak mengharuskan penggunaannya untuk melakukan pemenuhan asumsi-asumsi yang ketat. Beberapa metode yang populer digunakan oleh pendekatan regresi nonparametrik diantaranya histogram, kernel, *spline*, polinomial lokal, deret fourier, dan lain-lain (Budiantara, 2011).

Regresi *spline* merupakan salah satu pendekatan regresi nonparametrik yang dimodifikasi dari fungsi polinomial tersegmen (Rumlawang dkk., 2018). Polinomial yang bersegmen itu yang menyebabkan regresi nonparametrik *spline* termasuk kedalam salah satu jenis *piecewise* polinomial. Fleksibilitasnya melebihi polinomial biasa pada umumnya, sehingga sangat memungkinkan

dapat melakukan penyesuaian diri lebih efektif terhadap perilaku lokal suatu fungsi atau data. Kemampuannya dalam penyesuaian diri terhadap karakteristik data dinilai lebih efektif, dikarenakan estimator pada *spline* memiliki sifat *truncated* atau berupa potongan-potongan polinomial tersegmen yang fleksibel (Hidayat dkk., 2017). Potongan-potongan polinomial inilah yang menjadi kelebihan dari regresi *spline* atau dinamakan dengan titik knot. Titik knot berisi informasi mengenai perubahan pola perilaku data pada setiap interval berbeda, yang ditunjukkan dengan titik belok pada kurva fungsi. Sehingga bentuk estimasi kurva regresinya menyesuaikan data tanpa dipengaruhi oleh faktor subjektivitas peneliti (Eubank, 2004).

Kelemahan dari fungsi polinomial pada umumnya adalah memiliki sifat global. Sehingga diperlukan pengembangan suatu fungsi yang dapat mempertahankan fungsi polinomial secara lokal. Permasalahan tersebut dapat diatasi dengan fungsi *spline truncated* (Rositawati dkk., 2019). Dalam memodelkannya, regresi nonparametrik *spline*

truncated ini akan menentukan orde dan titik knot yang optimal pada suatu fungsi.

Pada penelitian ini, dilakukan penelitian mengenai faktor-faktor yang berpengaruh signifikan terhadap kasus kejadian diare dengan metode regresi nonparametrik *spline truncated* sekaligus memodelkan hubungan antara kedua variabel tersebut. Penyakit diare tergolong masih relatif tinggi ditemukan di Provinsi Bali. Informasi ini didapatkan dari data Dinas Kesehatan Provinsi Bali dari tahun 2020 hingga 2022. Diperkirakan penemuan jumlah target kasus diare di tahun 2020 sekitar 114.725 kasus, selanjutnya tahun 2021 mencapai angka 118.174, dan pada tahun 2022 mencapai 119.207 kasus. Menurut angka peningkatan yang ditemukan setiap tahun menunjukkan bahwa Provinsi Bali masih belum sempurna dalam penanganan masalah kesehatan, terutama diare. Tujuan dari penelitian ini untuk memodelkan kasus kejadian diare dan menentukan faktor-faktor yang berpengaruh signifikan terhadap kasus kejadian diare di provinsi Bali.

2. METODE PENELITIAN

2.1 Regresi Nonparametrik *Spline Truncated*

Secara umum, model regresi nonparametrik didefinisikan sebagai berikut (Eubank, 1999):

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

dimana y_i sebagai variabel respon pengamatan ke- i , $f(t_i)$ sebagai fungsi f dari variabel prediktor pada pengamatan t_1, \dots, t_n yang tidak diketahui, dan ε_i sebagai galat atau eror dari pengamatan ke- i yang diasumsikan tidak berkorelasi, dengan nilai *mean* adalah 0 dan nilai varians konstan yaitu σ^2 .

Berikut fungsi *spline truncated* f berorde m dengan titik knot $K_{j1}, K_{j2}, \dots, K_{jr}$ yang didefinisikan sebagai berikut (Eubank, 1999):

$$\begin{aligned} f(t_j) &= \beta_0 + \beta_{j1}t_j^1 + \beta_{j2}t_j^2 + \dots + \beta_{jm}t_j^m + \\ &\quad \sum_{k=1}^r \beta_{j(k+m)}(t_j - K_{jk})_+^m \\ &= \sum_{h=0}^m \beta_{jh}t_j^h + \sum_{k=1}^r \beta_{j(k+m)}(t_j - K_{jk})_+^m \end{aligned} \quad (2)$$

dimana,

j = variabel prediktor yang nilainya diketahui,

$j = 1, 2, \dots, p$

h = derajat polinomial, $h = 0, 1, 2, \dots, m$

k = titik-titik knot, $k = 1, 2, \dots, r$

$m = 1$ (karena fungsi regresi nonparametrik *spline truncated* yang digunakan pada penelitian ini adalah fungsi linier)

dan $(t_j - K_{jk})_+^m$ sebagai fungsi *truncated* yang dapat dijabarkan sebagai berikut:

$$(t_j - K_{jk})_+^m = \begin{cases} (t_j - K_{jk})_+^m, & t_j \geq K_{jk} \\ 0, & t_j < K_{jk} \end{cases} \quad (3)$$

Dari persamaan (1) akan disubstitusikan kedalam persamaan (2) untuk memperoleh model regresi nonparametrik *spline truncated*, maka secara umum dinyatakan sebagai berikut:

$$y_j = \sum_{h=0}^m \beta_{jh}t_j^h + \sum_{k=1}^r \beta_{j(k+m)}(t_j - K_{jk})_+^m + \varepsilon_j \quad (4)$$

Dalam pengestimasi parameter β dapat dilakukan dengan metode *Maximum Likelihood Estimator* (MLE). Metode MLE ini digunakan apabila distribusi populasinya telah diketahui. Berdasarkan asumsi t_j dan ε_j pada persamaan (1) yang telah dijelaskan sebelumnya, maka galat y_j juga diasumsikan tidak berkorelasi atau berdistribusi normal dengan nilai *mean* $f(t_i)$ dan variansi σ^2 . Sehingga fungsi densitas peluang y_i adalah sebagai berikut:

$$\begin{aligned} f(y; f(t), \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y-f(t))^2}{2\sigma^2} \right] \\ &\quad , f(t) > 0, \sigma^2 > 0 \end{aligned} \quad (5)$$

Selanjutnya fungsi *likelihood* dapat dinyatakan sebagai berikut:

$$\begin{aligned} L(y, f) &= \prod_{j=1}^p f(y_j; f(t_j), \sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}n} \exp \left[-\frac{1}{2\sigma^2} \sum_{j=1}^p (y_j - f(t_j))^2 \right] \end{aligned} \quad (6)$$

Setelah diperoleh fungsi *likelihood*, tahap selanjutnya adalah memaksimumkan fungsi *likelihood* $L(y, f)$ untuk mendapatkan estimasi titik fungsi f sebagai berikut:

$$\begin{aligned} \max_f \{L(y, f)\} &= \\ &\max_{\beta \in R^{m+h}} \left\{ (2\pi\sigma^2)^{-\frac{1}{2}n} \exp \left(-\frac{1}{2\sigma^2} \sum_{j=1}^p (y_j - \right. \right. \\ &\left. \left. \sum_{h=0}^m \beta_{jh}t_j^h + \sum_{k=1}^r \beta_{j(m+k)}(t_j - K_{jk})_+^m)^2 \right) \right\} \end{aligned} \quad (7)$$

Kemudian menerapkan transformasi logaritma pada persamaan (7) hingga diturunkan secara parsial terhadap β pada persamaan yang telah diperoleh serta di sisi kanannya disamakan dengan nol. Dengan demikian diperoleh:

$$\hat{\beta} = (t't)^{-1}t'y. \quad (8)$$

Sehingga pengestimasi \hat{y} dapat diperoleh dengan sebagai berikut:

$$\begin{aligned} \hat{y} &= t\hat{\beta} + \varepsilon \\ \hat{y} &= t((t't)^{-1}t'y) + \varepsilon \\ &= A(k)y + \varepsilon. \end{aligned} \quad (9)$$

Dimana $A(k)$ menyatakan matriks yang digunakan untuk menghitung nilai *generalized cross validation* (GCV) dalam menentukan titik knot yang optimal.

2.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah jenis data sekunder, yang diperoleh dari Dinas Kesehatan Provinsi Bali Tahun 2022. Unit amatan pada penelitian ini adalah 57 kecamatan pada 9 Kabupaten/Kota di Provinsi Bali.

2.3 Variabel Penelitian

Variabel penelitian yang digunakan dapat dilihat pada tabel di bawah ini:

Tabel 1. Variabel Penelitian dan Skala Pengukuran

| Kode | Nama Variabel |
|---------------------|--|
| Variabel respon: | |
| y | Kasus kejadian diare |
| Variabel prediktor: | |
| x_1 | Jumlah sarana air minum bersih |
| x_2 | Jumlah TPM (Tempat Pengelolaan Makanan) yang memenuhi syarat kesehatan |
| x_3 | Jumlah TTU (Tempat-Tempat Umum) yang memenuhi syarat kesehatan |
| x_4 | Kepadatan penduduk |
| x_5 | Jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak. |

2.4 Teknik Analisis Data

Berikut tahapan teknik dan analisis data:

- Menentukan variabel-variabel penelitian yang akan digunakan.
- Membuat gambaran umum kasus yang digunakan dimana meliputi statistika deskriptif, dan *scatterplot*.
- Memodelkan variabel respon dan variabel prediktor menggunakan metode regresi nonparametrik *spline truncated* dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot yang berorde linier.
- Menentukan titik knot optimal dengan menggunakan kriteria GCV yang paling minimum dengan langkah-langkah sebagai berikut:
 - Menghitung matriks $A(k)$
 - Menghitung nilai *Mean Square Error* (MSE).
 - Menentukan titik knot yang optimal dengan menggunakan nilai GCV yang paling minimum.
- Membuat model regresi nonparametrik *spline truncated* dengan orde dan titik knot yang optimal.
- Pengujian parameter model regresi nonparametrik *spline truncated* secara serentak maupun secara parsial.
- Pengujian asumsi residual dengan asumsi Identik, Independen, dan Berdistribusi Normal (IIDN) dari model regresi nonparametrik *spline truncated*.
- Menghitung nilai koefisien determinasi (R^2) model regresi nonparametrik *spline truncated*.
- Menginterpretasikan model yang telah diperoleh dari regresi nonparametrik *spline truncated*.
- Menarik kesimpulan dari hasil interpretasi model regresi nonparametrik *spline truncated*.

3. HASIL DAN PEMBAHASAN

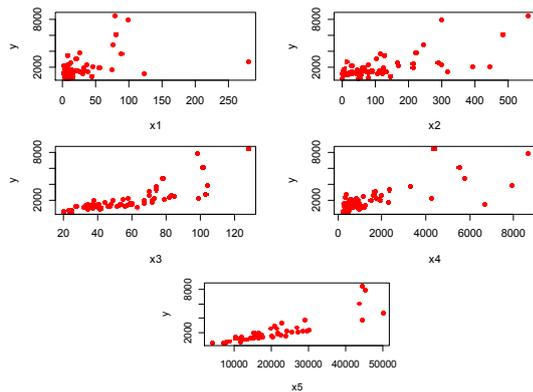
3.1 Gambaran Umum Kasus Kejadian Diare

Gambaran umum kasus kejadian diare di Provinsi Bali dengan kelima variabel yang diduga memengaruhinya dijelaskan melalui tabel statistika deskriptif. Berikut penjelasan tabel statistika deskriptif meliputi rata-rata, nilai minimum, dan nilai maksimum.

Tabel 2. Tabel Statistika Deskriptif

| Variabel | Mean | Min | Max |
|----------|----------|-------|--------|
| y | 2052,807 | 592 | 8530 |
| x_1 | 27,895 | 1 | 279 |
| x_2 | 118,754 | 0 | 555 |
| x_3 | 57,561 | 20 | 128 |
| x_4 | 1549,03 | 177,2 | 8619,5 |
| x_5 | 19305,7 | 4040 | 49828 |

Sebelum penelitian dilanjutkan ke tahap pemodelan, maka perlu diketahui dahulu pola hubungan antara kasus kejadian diare dengan kelima variabel yang diduga memengaruhinya. Pola hubungan kedua jenis variabel dapat diketahui menggunakan bantuan *scatterplot*. Berikut *scatterplot* hubungan antara variabel respon dengan masing-masing variabel prediktor.



Gambar 1. Scatterplot Hubungan antara Variabel Respon dengan Masing-Masing Variabel Prediktornya

Berdasarkan gambar tersebut, dapat diperhatikan bahwa *scatterplot* hubungan antara kasus kejadian diare dengan masing-masing variabel prediktor cenderung tidak membentuk suatu pola tertentu. Sehingga kedua jenis variabel ini dapat dilanjutkan ke tahap pemodelan dengan metode regresi nonparametrik *spline truncated*.

3.2 Pemodelan Kasus Kejadian Diare dengan Regresi Nonparametrik *Spline Truncated*

Pada tahap ini, langkah pertama yang dilakukan adalah menentukan orde dan titik knot yang optimal. Orde model yang digunakan adalah orde linier atau berorde satu. Titik knot yang dicobakan adalah satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot dengan menggunakan kriteria GCV.

3.2.1 Model *Spline Truncated* dengan Satu Titik Knot

Pada percobaan pertama, pemilihan titik knot optimal dilakukan dengan satu titik knot. Berikut model regresi nonparametrik *spline truncated* dan hasil 5 iterasi terbaik menggunakan orde satu dan satu titik knot pada Tabel 3.

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - K_{11})_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - K_{21})_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - K_{31})_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - K_{41})_+^1 + \hat{\beta}_{51}x_5 + \hat{\beta}_{52}(x_5 - K_{51})_+^1$$

Tabel 3. Hasil Iterasi Nilai GCV dengan Satu Titik Knot

| Banyak Kandidat Titik Knot | Knot | | | | | GCV |
|----------------------------|--------|--------|--------|---------|----------|----------|
| | x_1 | x_2 | x_3 | x_4 | x_5 | |
| 56 | 238,56 | 474,27 | 112,29 | 7391,53 | 43167,93 | 163782,6 |
| 49 | 238,46 | 474,06 | 112,25 | 7388,33 | 43150,58 | 163928,9 |
| 42 | 238,32 | 473,78 | 112,20 | 7384,04 | 43127,32 | 164124,5 |
| 35 | 238,1 | 473,40 | 112,1 | 7378 | 43094,5 | 164399,5 |
| 28 | 237,8 | 472,80 | 112 | 7368,80 | 43044,6 | 164814,7 |

Berdasarkan Tabel 3, terdapat hasil 5 iterasi nilai GCV dan nilai knot untuk masing-masing variabel prediktor. Nilai GCV minimum yang diperoleh pada satu titik knot adalah 163782,6 dengan nilai knot masing-masing variabel prediktor sebagai berikut:

- Variabel x_1 : 238,56 (K_{11})
- Variabel x_2 : 474,27 (K_{21})
- Variabel x_3 : 112,29 (K_{31})
- Variabel x_4 : 7391,53 (K_{41})
- Variabel x_5 : 43167,93 (K_{51})

3.2.2 Model *Spline Truncated* dengan Dua Titik Knot

Pada percobaan kedua, setelah mendapatkan nilai GCV minimum dengan satu titik knot, selanjutnya pemilihan titik knot optimal dilakukan dengan dua titik knot. Berikut model regresi nonparametrik *spline truncated* dan hasil 5 iterasi terbaik menggunakan orde satu dan dua titik knot pada Tabel 4.

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - K_{11})_+^1 + \hat{\beta}_{13}(x_1 - K_{12})_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - K_{21})_+^1 + \hat{\beta}_{23}(x_2 - K_{22})_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - K_{31})_+^1 + \hat{\beta}_{33}(x_3 - K_{32})_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - K_{41})_+^1 + \hat{\beta}_{43}(x_4 - K_{42})_+^1 + \hat{\beta}_{51}x_5 + \hat{\beta}_{52}(x_5 - K_{51})_+^1 + \hat{\beta}_{53}(x_5 - K_{52})_+^1$$

Tabel 4. Hasil Iterasi Nilai GCV dengan Dua Titik Knot

| Banyak Kandidat Titik Knot | Knot | | | | | GCV |
|----------------------------|--------|--------|--------|---------|----------|----------|
| | x_1 | x_2 | x_3 | x_4 | x_5 | |
| 56 | 162,75 | 322,91 | 82,84 | 5089,08 | 30680,29 | 83907,18 |
| | 223,4 | 444 | 106,4 | 6931,04 | 40670,4 | |
| 37 | 163,2 | 323,7 | 83 | 5101,9 | 30749,7 | 83915,6 |
| | 224,9 | 447,1 | 107 | 6977,9 | 40924,8 | |
| 42 | 163,73 | 324,88 | 83,22 | 5119,03 | 30842,73 | 83966,67 |
| | 224,76 | 446,71 | 106,93 | 6972,22 | 40893,76 | |
| 32 | 162,40 | 322,2 | 82,7 | 5079,2 | 30626,6 | 84018,3 |
| | 225,2 | 447,5 | 107,1 | 6985,5 | 40965,8 | |
| 51 | 162,24 | 321,9 | 82,64 | 5073,73 | 30597,04 | 84021,71 |
| | 223,4 | 444 | 106,4 | 6931,04 | 40670,4 | |

Berdasarkan Tabel 4, terdapat hasil 5 iterasi nilai GCV dan nilai knot untuk masing-masing variabel prediktor. Nilai GCV minimum yang diperoleh pada dua titik knot adalah 83907,18 dengan nilai knot masing-masing variabel prediktor sebagai berikut:

- Variabel x_1 : 162,75 (K_{11}) ; 223,4 (K_{12})
- Variabel x_2 : 322,91 (K_{21}) ; 444 (K_{22})
- Variabel x_3 : 82,84 (K_{31}) ; 106,4 (K_{32})
- Variabel x_4 : 5089,08 (K_{41}) ; 6931,04 (K_{42})
- Variabel x_5 : 30680,29 (K_{51}) ; 40670,4 (K_{52})

3.2.3 Model Spline Truncated dengan Tiga Titik Knot

Pada percobaan ketiga, setelah mendapatkan nilai GCV minimum dengan dua titik knot, selanjutnya pemilihan titik knot optimal dilakukan dengan tiga titik knot. Berikut model regresi nonparametrik *spline truncated* dan hasil 5 iterasi terbaik menggunakan orde satu dan tiga titik knot pada Tabel 5.

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - K_{11})_+^1 + \hat{\beta}_{13}(x_1 - K_{12})_+^1 + \hat{\beta}_{14}(x_1 - K_{13})_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - K_{21})_+^1 + \hat{\beta}_{23}(x_2 - K_{22})_+^1 + \hat{\beta}_{24}(x_2 - K_{23})_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - K_{31})_+^1 + \hat{\beta}_{33}(x_3 - K_{32})_+^1 + \hat{\beta}_{34}(x_3 - K_{33})_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - K_{41})_+^1 + \hat{\beta}_{43}(x_4 - K_{42})_+^1 + \hat{\beta}_{44}(x_4 - K_{43})_+^1 + \hat{\beta}_{51}x_5 + \hat{\beta}_{52}(x_5 - K_{51})_+^1 + \hat{\beta}_{53}(x_5 - K_{52})_+^1 + \hat{\beta}_{54}(x_5 - K_{53})_+^1$$

Tabel 5. Hasil Iterasi Nilai GCV dengan Tiga Titik Knot

| Banyak Kandidat Titik Knot | Knot | | | | | GCV |
|----------------------------|--------|--------|--------|---------|----------|----------|
| | x_1 | x_2 | x_3 | x_4 | x_5 | |
| 35 | 25,53 | 48,97 | 29,53 | 922,11 | 8080,12 | 7569,315 |
| | 197,24 | 391,76 | 96,24 | 6136,47 | 36360,94 | |
| | 205,41 | 408,09 | 99,41 | 6384,77 | 37707,65 | |
| 45 | 19,95 | 37,84 | 27,36 | 752,81 | 7161,91 | 7599,991 |
| | 165,27 | 327,95 | 83,82 | 5165,83 | 31096,55 | |
| | 222,14 | 441,48 | 105,91 | 6892,67 | 40462,27 | |
| 31 | 19,53 | 37 | 27,2 | 740,02 | 7092,53 | 7620,299 |
| | 195,6 | 388,5 | 95,6 | 6086,81 | 36091,6 | |
| | 204,87 | 407 | 99,2 | 6368,22 | 37617,87 | |
| 54 | 27,23 | 52,36 | 30,19 | 973,64 | 8359,62 | 7639,572 |
| | 195,08 | 387,45 | 95,4 | 6070,88 | 36005,21 | |
| | 205,57 | 408,40 | 99,47 | 6389,46 | 37733,06 | |
| 55 | 21,59 | 41,11 | 28,00 | 802,56 | 7431,7 | 7647,691 |
| | 165,74 | 328,89 | 84,00 | 5180,04 | 31173,63 | |
| | 222,37 | 441,94 | 106,00 | 6899,77 | 40500,81 | |

Berdasarkan Tabel 5, terdapat hasil 5 iterasi nilai GCV dan nilai knot untuk masing-masing variabel prediktor. Nilai GCV minimum yang diperoleh pada tiga titik knot adalah 75693,15 dengan nilai knot masing-masing variabel prediktor sebagai berikut:

- Variabel x_1 : 25,53 (K_{11}); 197,24 (K_{12}); 205,41 (K_{13})
 Variabel x_2 : 48,97 (K_{21}); 391,76 (K_{22}); 408,09 (K_{23})

- Variabel x_3 : 29,53 (K_{31}); 96,24 (K_{32}); 99,41 (K_{33})
 Variabel x_4 : 922,11 (K_{41}); 6136,47 (K_{42}); 6384,77 (K_{43})
 Variabel x_5 : 8080,12 (K_{51}); 36360,94 (K_{52}); 37707,65 (K_{53})

3.2.4 Model Spline Truncated dengan Kombinasi Titik Knot

Pada percobaan terakhir, setelah mendapatkan nilai GCV minimum dengan tiga titik knot, selanjutnya pemilihan titik knot optimal dilakukan dengan kombinasi titik knot. Berikut 5 iterasi terbaik yang dipilih berdasarkan nilai GCV minimum dari kombinasi titik knot pada Tabel 6 sebelum memperkirakan model regresi dengan kombinasi titik knot.

Tabel 6. Hasil Iterasi Nilai GCV dengan Kombinasi Titik Knot

| Banyak Kandidat Titik Knot | Variabel | Kombinasi Titik Knot | Titik-Titik Knot | | | GCV |
|----------------------------|----------|----------------------|------------------|----------|--------|----------|
| 43 | x_1 | 2 | 159,86 | 226,05 | | 67572,38 |
| | x_2 | 1 | 462,50 | | | |
| | x_3 | 3 | 27,71 | 84,29 | 104,86 | |
| | x_4 | 3 | 780,22 | 5202,38 | 6810,4 | |
| | x_5 | 2 | 30204,57 | 41106,48 | | |
| 53 | x_1 | 2 | 161,38 | 225,54 | | 67612,50 |
| | x_2 | 1 | 469,62 | | | |
| | x_3 | 3 | 28,31 | 84,38 | 105,15 | |
| | x_4 | 3 | 826,61 | 5210,11 | 6833,6 | |
| | x_5 | 2 | 30456,15 | 41022,62 | | |
| 23 | x_1 | 2 | 165,27 | 228,45 | | 67696,38 |
| | x_2 | 1 | 454,09 | | | |
| | x_3 | 3 | 29,82 | 88,73 | 103,45 | |
| | x_4 | 3 | 944,68 | 5549,57 | 6700,8 | |
| | x_5 | 2 | 31096,55 | 41502,91 | | |
| 38 | x_1 | 2 | 166,30 | 226,41 | | 67715,59 |
| | x_2 | 1 | 465,00 | | | |
| | x_3 | 2 | 84,22 | 107,57 | | |
| | x_4 | 3 | 861,71 | 6109,63 | 6337,8 | |
| | x_5 | 1 | 42402,92 | | | |
| 31 | x_1 | 2 | 167,80 | 223,4 | | 67947,10 |
| | x_2 | 1 | 462,50 | | | |
| | x_3 | 2 | 84,80 | 106,4 | | |
| | x_4 | 3 | 740,00 | 6086,8 | 6368,2 | |
| | x_5 | 1 | 42196,70 | | | |

Berdasarkan Tabel 6, terdapat hasil 5 iterasi nilai GCV dan nilai knot untuk masing-masing variabel prediktor. Kombinasi jumlah titik knot pada masing-masing variabel prediktor yang memiliki nilai GCV minimum diperoleh pada kombinasi 2-1-3-3-2. Nilai GCV minimumnya yaitu 67572,38 dengan nilai knot masing-masing variabel prediktor sebagai berikut:

- Variabel x_1 : 159,86 (K_{11}); 226,05 (K_{12})
 Variabel x_2 : 462,50 (K_{21})
 Variabel x_3 : 27,71 (K_{31}); 84,29 (K_{32}); 104,86 (K_{33})

Variabel x_4 : 780,22 (K_{41}); 5202,38 (K_{42});
6810,44 (K_{43})
Variabel x_5 : 30204,57 (K_{51}); 41106,48 (K_{52})

Model regresi nonparametrik *spline truncated* dari model kombinasi titik knot yang terbentuk yaitu sebagai berikut.

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - K_{11})_+^1 + \hat{\beta}_{13}(x_1 - K_{12})_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - K_{21})_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - K_{31})_+^1 + \hat{\beta}_{33}(x_3 - K_{32})_+^1 + \hat{\beta}_{34}(x_3 - K_{33})_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - K_{41})_+^1 + \hat{\beta}_{43}(x_4 - K_{42})_+^1 + \hat{\beta}_{44}(x_4 - K_{43})_+^1 + \hat{\beta}_{51}x_5 + \hat{\beta}_{52}(x_5 - K_{51})_+^1 + \hat{\beta}_{53}(x_5 - K_{52})_+^1$$

3.3 Pemilihan Titik Knot Optimal

Pada tahap ini, seluruh GCV minimum dari masing-masing percobaan titik knot telah diperoleh. Selanjutnya dibandingkan nilai GCV minimum dari satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot pada Tabel 7.

Tabel 7. Perbandingan Nilai GCV Minimum

| Model | GCV |
|---------------------------------|-----------------|
| 1 Knot | 163782,6 |
| 2 Knot | 83907,18 |
| 3 Knot | 75693,15 |
| Kombinasi Knot 2-1-3-3-2 | 67572,38 |

Berdasarkan Tabel 7, nilai GCV minimum diperoleh pada model regresi dari kombinasi titik knot dengan nilai sebesar 67572,38. Hal ini menunjukkan titik knot optimal berada pada model regresi dengan kombinasi titik knot 2-1-3-3-2.

3.4 Pengestimasi Parameter Model *Spline Truncated* Terbaik

Model regresi nonparametrik *spline truncated* terbaik dibentuk dari titik knot optimal sesuai kriteria nilai GCV yang telah diperoleh. Pada kasus kejadian diare di Provinsi Bali, pemilihan titik knot optimal untuk model regresi terbaik adalah dengan kombinasi titik knot 2-1-3-3-2. Estimasi parameter regresi model regresi disajikan pada Tabel 8.

Tabel 8. Estimasi Parameter Model

| Variabel | Parameter | Estimasi |
|----------|--------------|----------|
| | β_0 | -1,298 |
| x_1 | β_{11} | 0,968 |
| | β_{12} | 23,561 |
| | β_{13} | 10,471 |
| x_2 | β_{21} | 0,054 |
| | β_{22} | 222,612 |
| x_3 | β_{31} | 18,780 |
| | β_{32} | 3,956 |
| | β_{33} | -183,546 |
| | β_{34} | -405,231 |
| x_4 | β_{41} | -0,094 |
| | β_{42} | 0,804 |
| | β_{43} | -2,595 |
| | β_{44} | 6,274 |
| x_5 | β_{51} | 0,035 |
| | β_{52} | -0,096 |
| | β_{53} | 0,186 |

Hasil estimasi pada tabel tersebut dan hasil iterasi model kombinasi titik knot 2-1-3-3-2 akan disubstitusikan ke model regresi nonparametrik *spline truncated* terbaik yaitu sebagai berikut:

$$\hat{y}_j = -1,298 + 0,968x_1 + 23,561(x_1 - 159,86)_+^1 + 10,471(x_1 - 226,05)_+^1 + 0,054x_2 + 222,612(x_2 - 462,50)_+^1 + 18,780x_3 + 3,956(x_3 - 27,71)_+^1 - 183,546(x_3 - 84,29)_+^1 - 405,231(x_3 - 104,86)_+^1 - 0,094x_4 + 0,804(x_4 - 780,22)_+^1 - 2,595(x_4 - 5202,38)_+^1 + 6,274(x_4 - 6810,44)_+^1 + 0,035x_5 - 0,096(x_5 - 30204,57)_+^1 + 0,186(x_5 - 41106,48)_+^1$$

3.5 Pengujian Parameter Model

Pengujian parameter model digunakan ketika ingin mengetahui apakah variabel prediktor memberikan pengaruh yang signifikan atau tidak terhadap kasus kejadian diare di Provinsi Bali. Pengujian parameter model terdiri dari dua tahapan, yakni uji secara serentak terlebih dahulu, kemudian jika terbukti parameter berpengaruh signifikan dilanjutkan uji secara parsial.

Tabel 9. Tabel ANOVA Model Regresi Nonparametrik *Spline Truncated*

| Sumber | df | SS | MS | F_{hitung} | p -value |
|---------|----|-----------|------------|--------------|------------|
| Regresi | 16 | 135900030 | 8493751,85 | 162,47 | 3,274e-31 |
| Error | 40 | 2091187 | 52279,68 | | |
| Total | 56 | 137991217 | - | | |

Berdasarkan perhitungan *Analysis of Variance* (ANOVA) pada Tabel 9, dapat dilihat nilai $F_{hitung}(162,47) > F_{(0,05;16;40)}(1,90)$ dan $p\text{-value}(3,274e-31) < \alpha_{(0,05)}$, sehingga keputusan adalah tolak H_0 atau minimal ada satu parameter model yang berpengaruh signifikan. Selanjutnya untuk mengetahui model apa saja yang berpengaruh signifikan, maka pengujian parameter model dilanjutkan ke uji parsial.

Tabel 10. Hasil Pengujian Signifikansi Parameter Secara Individu

| Variabel | Parameter | Koefisien | t_{hitung} | $P\text{-value}$ |
|----------|--------------|-----------|--------------|------------------|
| x_1 | β_{11} | 0,968 | 0,592 | 0,556 |
| | β_{12} | 23,561 | 5,614 | 6,40e-07 |
| | β_{13} | 10,471 | 5,614 | 6,40e-07 |
| x_2 | β_{21} | 0,054 | 0,146 | 0,885 |
| | β_{22} | 222,612 | 8,570 | 9,01e-12 |
| x_3 | β_{31} | 18,780 | 3,587 | 7,04e-04 |
| | β_{32} | 3,956 | -0,710 | 0,481 |
| | β_{33} | -183,546 | -8,466 | 1,33e-11 |
| x_4 | β_{41} | -405,231 | -5,393 | 1,44e-06 |
| | β_{42} | -0,094 | -0,481 | 0,633 |
| | β_{43} | 0,804 | 3,191 | 2,32e-03 |
| | β_{44} | -2,595 | -7,083 | 2,53e-09 |
| x_5 | β_{51} | 6,274 | 10,153 | 2,63e-14 |
| | β_{52} | 0,035 | 3,360 | 1,41e-03 |
| | β_{53} | -0,096 | -2,170 | 0,034 |
| | | 0,186 | 1,936 | 0,058 |

Berdasarkan hasil pengujian parameter model dengan uji parsial pada Tabel 10, dapat diperhatikan bahwa pada parameter $\beta_{11}, \beta_{21}, \beta_{32}, \beta_{41}, \beta_{53}$ diperoleh keputusan gagal tolak H_0 dikarenakan nilai $p\text{-value} > \alpha_{(0,05)}$. Namun untuk parameter lainnya terbukti berpengaruh signifikan. Selanjutnya dapat dilihat pada tabel tersebut bahwa pada setiap variabel-variabel prediktornya masih terdapat parameter lainnya yang signifikan, maka secara umum parameter kelima variabel prediktor tersebut terbukti berpengaruh signifikan.

3.6 Pengujian Asumsi Residual

Pengujian asumsi residual diperlukan ketika model terbaik dari regresi nonparametrik *spline truncated* telah ditemukan. Tujuannya untuk mengetahui apakah residual dari model regresi yang dihasilkan telah memenuhi asumsi residual atau tidak. Adapun yang meliputi pengujian asumsi residual yaitu Identik, Independen, dan Berdistribusi Normal (IIDN).

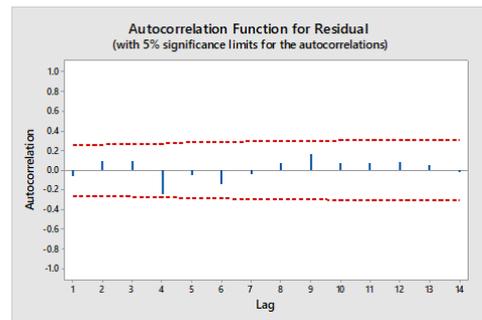
Tabel 11. Tabel ANOVA dari Uji *Glejser*

| Sumber | df | SS | MS | F_{hitung} | $P\text{-value}$ |
|---------|----|----------|-----------|--------------|------------------|
| Regresi | 16 | 320007,9 | 20000,497 | 0,887 | 0,5881 |
| Error | 40 | 902480,8 | 22562,02 | | |
| Total | 56 | 1222489 | - | | |

Berdasarkan perhitungan *Analysis of Variance* (ANOVA) dari uji *Glejser* pada Tabel 11, dapat dilihat nilai $F_{hitung}(0,887) <$

$F_{(0,05;16;40)}(1,90)$ dan $P\text{-value}(0,5881) > \alpha_{(0,05)}$, sehingga keputusan adalah gagal tolak H_0 . Kesimpulan yang didapatkan adalah varians residual tidak terjadi *heteroskedastisitas* atau residual memenuhi uji asumsi identik.

Pengujian asumsi residual selanjutnya adalah uji asumsi independen. Pengujian ini dilakukan dengan melihat plot *Autocorrelation Function* (ACF).



Gambar 2. Plot ACF dari Residual Model Regresi

Berdasarkan hasil Plot ACF pada Gambar 2, dapat diperhatikan bahwa tidak ada nilai autokorelasi pada lag yang keluar melewati batas signifikansi. Sehingga keputusan adalah gagal tolak H_0 atau tidak terdapat korelasi antar residual. Hal ini menunjukkan bahwa asumsi independen residual terpenuhi. Pengujian ini dilakukan dengan uji *Kolmogorov-Smirnov*.

Berdasarkan hasil uji *Kolmogorov-Smirnov* pada residual model regresi nonparametrik *spline truncated*, diperoleh nilai $D(0,1602) < D_{tabel}(0,179)$ dan $P\text{-value}(0,0959) > \alpha_{(0,05)}$, sehingga keputusan adalah gagal tolak H_0 . Kesimpulan yang didapatkan adalah residual model regresi memenuhi uji asumsi berdistribusi normal.

Seluruh pengujian asumsi residual yang meliputi uji Identik, uji Independen, dan uji Berdistribusi Normal (IIDN) terbukti dapat terpenuhi. Sehingga dapat disimpulkan bahwa model regresi nonparametrik *spline truncated* yang terbentuk layak untuk menjelaskan hubungan antara variabel respon dan variabel prediktor dari kasus kejadian diare di Provinsi Bali.

3.7 Koefisien Determinasi (R^2)

Salah satu indikasi model dikatakan baik jika nilai koefisien determinasi (R^2) yang dihasilkan tinggi. Berdasarkan perhitungan *Analysis of Variance* (ANOVA) dari Uji *Glejser*, diperoleh nilai *Sum of Square Regression* (SSR) sebesar 135900030 dan nilai

Sum of Square Total (SST) sebesar 137991217. Sehingga dari nilai SSR dan SST yang diperoleh, maka didapatkan nilai koefisien determinasi (R^2) sebesar 0,9848 atau sama dengan 98,48%. Nilai R^2 yang diperoleh ini menunjukkan bahwa model dapat menerangkan variasi-variabel respon sebesar 98,48%. Dengan nilai R^2 yang tergolong tinggi dapat dikatakan model regresi nonparametrik *spline truncated* yang terkontruksi merupakan model yang baik.

3.8 Interpretasi Model Regresi Nonparametrik *Spline Truncated*

Interpretasi model digunakan untuk mengetahui seberapa besar pengaruh masing-masing variabel prediktor yang terbukti signifikan dengan kasus kejadian diare di Provinsi Bali. Adapun variabel-variabel prediktor yang terbukti signifikan adalah jumlah sarana air minum bersih (x_1), jumlah TPM (Tempat Pengelolaan Makanan) yang memenuhi syarat kesehatan (x_2), jumlah TTU (Tempat-Tempat Umum) yang memenuhi syarat kesehatan (x_3), kepadatan penduduk (x_4), dan jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak (x_5). Berikut hasil interpretasi model terbaik pada masing-masing variabel prediktor:

1. Hubungan antara jumlah sarana air minum bersih (x_1) dengan kasus kejadian diare (y) dengan mengasumsikan variabel lain nilainya tetap atau konstan adalah sebagai berikut:

$$\hat{y}_j = 0,968x_1 + 23,561(x_1 - 159,86)_+^1 + 10,471(x_1 - 226,05)_+^1$$

$$= \begin{cases} 0,968x_1 & ; & x_1 < 159,86 \\ -3766,46 + 24,53x_1 & ; & 159,86 \leq x_1 < 226,05 \\ -6133,43 + 35x_1 & ; & x_1 \geq 226,05 \end{cases}$$

Berdasarkan model tersebut dapat ditarik kesimpulan, jika pada suatu wilayah memiliki jumlah sarana air minum bersih kurang dari 159,86 maka setiap terjadi kenaikan jumlah sarana air minum bersih sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare naik sebanyak 1 kasus. Selanjutnya, jika pada suatu wilayah memiliki jumlah sarana air minum bersih berada antara rentang 159,86 dan 226,05 maka setiap terjadi kenaikan jumlah sarana air minum bersih sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare naik secara signifikan sebanyak 25 kasus. Kemudian, jika pada suatu wilayah memiliki jumlah sarana air minum bersih lebih dari 226,05 maka setiap terjadi kenaikan jumlah sarana air minum bersih sebanyak 1 satuan, maka kasus kejadian

diare naik secara signifikan sebanyak 35 kasus. Terjadi inkonsistensi dalam interpretasi model hubungan antara jumlah sarana air minum bersih dengan kasus kejadian diare karena ada kemungkinan terdapat pengaruh dari variabel-variabel lain di luar variabel penelitian, yakni variabel biologis.

2. Hubungan antara jumlah TPM (Tempat Pengelolaan Makanan) yang memenuhi syarat kesehatan (x_2) dengan kasus kejadian diare (y) dengan mengasumsikan variabel lain nilainya tetap atau konstan adalah sebagai berikut:

$$\hat{y}_j = 0,054x_2 + 222,612(x_2 - 462,50)_+^1$$

$$= \begin{cases} 0,054x_2 & ; & x_2 < 462,50 \\ -102958,05 + 222,67x_2 & ; & x_2 \geq 462,50 \end{cases}$$

Berdasarkan model tersebut dapat ditarik kesimpulan, jika pada suatu wilayah memiliki jumlah TPM yang memenuhi syarat kesehatan kurang dari 462,50 maka setiap terjadi kenaikan jumlah TPM yang memenuhi syarat kesehatan sebanyak 10 satuan, mengakibatkan kasus kejadian diare akan naik sebanyak 1 kasus. Selanjutnya, jika pada suatu wilayah memiliki jumlah TPM yang memenuhi syarat kesehatan lebih dari 462,50 maka setiap terjadi kenaikan jumlah TPM yang memenuhi syarat kesehatan sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare naik secara signifikan sebanyak 223 kasus. Terjadi inkonsistensi dalam interpretasi model hubungan antara jumlah TPM yang memenuhi syarat kesehatan dengan kasus kejadian diare karena ada kemungkinan terdapat pengaruh dari variabel-variabel lain di luar variabel penelitian, yakni variabel biologis.

3. Hubungan antara jumlah TTU (Tempat-Tempat Umum) yang memenuhi syarat kesehatan (x_3) dengan kasus kejadian diare (y) dengan mengasumsikan variabel lain nilainya tetap atau konstan adalah sebagai berikut:

$$\hat{y}_j = 18,780x_3 + 3,956(x_3 - 27,71)_+^1 - 183,546(x_3 - 84,29)_+^1 - 405,231(x_3 - 104,86)_+^1$$

$$= \begin{cases} 18,780x_3 & ; & x_3 < 27,71 \\ -109,62 + 22,74x_3 & ; & 27,71 \leq x_3 < 84,29 \\ 15361,47 - 160,81x_3 & ; & 84,29 \leq x_3 < 104,86 \\ 57853,99 - 566,041x_3 & ; & x_3 \geq 104,86 \end{cases}$$

Berdasarkan model tersebut dapat ditarik kesimpulan, jika pada suatu wilayah memiliki jumlah TTU yang memenuhi syarat kesehatan kurang dari 27,71 maka setiap terjadi kenaikan jumlah TTU yang memenuhi syarat kesehatan sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare naik sebanyak 19 kasus.

Kemudian, jika pada suatu wilayah memiliki jumlah TTU yang memenuhi syarat kesehatan berada antara rentang 27,71 dan 84,29 maka setiap terjadi kenaikan jumlah TTU yang memenuhi syarat kesehatan sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare naik secara signifikan sebanyak 23 kasus. Namun, jika pada suatu wilayah memiliki jumlah TTU yang memenuhi syarat kesehatan berada antara rentang 84,29 dan 104,86 maka setiap terjadi kenaikan jumlah TTU yang memenuhi syarat kesehatan sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare turun secara signifikan sebanyak 161 kasus. Selanjutnya, jika pada suatu wilayah memiliki jumlah TTU yang memenuhi syarat kesehatan lebih dari 104,86 maka setiap terjadi kenaikan jumlah TTU yang memenuhi syarat kesehatan sebanyak 1 satuan, akan mengakibatkan kasus kejadian diare turun secara signifikan sebanyak 566 kasus.

4. Hubungan antara kepadatan penduduk (x_4) dengan kasus kejadian diare (y) dengan mengasumsikan variabel lain nilainya tetap atau konstan adalah sebagai berikut:

$$\hat{y}_j = -0,094x_4 + 0,804(x_4 - 780,22)_+^1 - 2,595(x_4 - 5202,38)_+^1 + 6,274(x_4 - 6810,44)_+^1$$

$$= \begin{cases} -0,094x_4 & ; & x_4 < 780,22 \\ -627,30 + 0,898x_4 & ; & 780,22 \leq x_4 < 5202,38 \\ 12872,88 - 1,697x_4 & ; & 5202,38 \leq x_4 < 6810,44 \\ -29885,82 + 4,58x_4 & ; & x_4 \geq 6810,44 \end{cases}$$

Berdasarkan model tersebut dapat ditarik kesimpulan, jika pada suatu wilayah memiliki kepadatan penduduk kurang dari 780,22 maka setiap terjadi kenaikan angka kepadatan penduduk sebanyak 10 jiwa/km², akan mengakibatkan kasus kejadian diare turun sebanyak 1 kasus. Lalu, jika pada suatu wilayah memiliki kepadatan penduduk berada antara rentang 780,22 dan 5202,38 maka setiap terjadi kenaikan angka kepadatan penduduk sebanyak 10 jiwa/km², akan mengakibatkan kasus kejadian diare naik sebanyak 9 kasus. Selanjutnya, jika suatu wilayah memiliki kepadatan penduduk berada antara rentang 5202,38 dan 6810,44 maka setiap terjadi kenaikan angka kepadatan penduduk sebanyak 10 jiwa/km², akan mengakibatkan kasus kejadian diare mengalami penurunan sebanyak 17 kasus. Namun, jika pada suatu wilayah memiliki kepadatan penduduk lebih dari 6810,44 maka setiap terjadi kenaikan angka kepadatan penduduk sebanyak 1 jiwa/km², akan mengakibatkan kasus kejadian diare naik secara signifikan sebanyak 46 kasus. Terjadi

inkonsistensi dalam interpretasi model hubungan antara kepadatan penduduk dengan kasus kejadian diare karena ada kemungkinan terdapat pengaruh dari variabel-variabel lain di luar variabel penelitian, yakni variabel biologis.

5. Hubungan antara jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak (x_5) dengan kasus kejadian diare (y) dengan mengasumsikan variabel lain nilainya tetap atau konstan adalah sebagai berikut:

$$\hat{y}_j = 0,035x_5 - 0,096(x_5 - 30204,57)_+^1 + 0,186(x_5 - 41106,48)_+^1$$

$$= \begin{cases} 0,035x_5 & ; & x_5 < 30204,57 \\ 2899,64 - 0,061x_5 & ; & 30204,57 \leq x_5 < 41106,48 \\ -4746,17 + 0,125x_5 & ; & x_5 \geq 41106,48 \end{cases}$$

Berdasarkan model tersebut dapat ditarik kesimpulan, jika suatu wilayah memiliki jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak kurang dari 30204,57 maka setiap terjadi kenaikan jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak sebanyak 100 satuan, akan mengakibatkan kasus kejadian diare naik sebanyak 4 kasus. Namun, jika suatu wilayah memiliki jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak berada antara 30204,57 dan 41106,48 maka setiap terjadi kenaikan jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak sebanyak 100 satuan, akan mengakibatkan kasus kejadian diare turun sebanyak 6 kasus. Selanjutnya, jika suatu wilayah memiliki jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak lebih dari 41106,48 maka setiap terjadi kenaikan jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak sebanyak 100 satuan, akan mengakibatkan kasus kejadian diare mengalami kenaikan secara sebanyak 13 kasus. Terjadi inkonsistensi dalam interpretasi model hubungan antara jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak dengan kasus kejadian diare karena ada kemungkinan terdapat pengaruh dari variabel-variabel lain di luar variabel penelitian, yakni variabel biologis.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan rumusan masalah dan hasil pembahasan yang dilakukan, maka diperoleh kesimpulan bahwa titik knot optimal yang digunakan dalam model regresi nonparametrik *spline truncated* terbaik diperoleh dari

kombinasi titik knot sesuai kriteria nilai GCV. Berdasarkan hasil yang diperoleh pada pemodelan kasus kejadian diare di Provinsi Bali, pemilihan titik knot optimal untuk model regresi terbaik adalah dengan kombinasi titik knot 2-1-3-3-2 dengan nilai GCV minimum sebesar 67572,38. Berikut model *spline truncated* yang digunakan:

$$\hat{y}_j = -1,298 + 0,968x_1 + 23,561(x_1 - 159,86)_+^1 + 10,471(x_1 - 226,05)_+^1 + 0,054x_2 + 222,612(x_2 - 462,50)_+^1 + 18,780x_3 + 3,956(x_3 - 27,71)_+^1 - 183,546(x_3 - 84,29)_+^1 - 405,231(x_3 - 104,86)_+^1 - 0,094x_4 + 0,804(x_4 - 780,22)_+^1 - 2,595(x_4 - 5202,38)_+^1 + 6,274(x_4 - 6810,44)_+^1 + 0,035x_5 - 0,096(x_5 - 30204,57)_+^1 + 0,186(x_5 - 41106,48)_+^1$$

Variabel-variabel prediktor yang terbukti berpengaruh signifikan dari model tersebut adalah jumlah sarana air minum bersih (x_1), jumlah TPM (Tempat Pengelolaan Makanan) yang memenuhi syarat kesehatan (x_2), jumlah TTU (Tempat-Tempat Umum) yang memenuhi syarat kesehatan (x_3), kepadatan penduduk (x_4), dan jumlah penduduk dengan akses terhadap fasilitas sanitasi yang layak (x_5). Dapat disimpulkan seluruh variabel prediktor berpengaruh signifikan. Nilai koefisien determinasi (R^2) yang diperoleh dari model terbaik ini sebesar 0,9847 atau sama dengan 98,48%, dapat dikatakan model regresi nonparametrik *spline truncated* yang terkonstruksi merupakan model yang baik untuk memodelkan kasus kejadian diare di Provinsi Bali.

4.2 Saran

Adapun saran yang dapat diberikan penulis untuk penelitian selanjutnya adalah sebagai berikut.

1. Penelitian ini dibatasi pada tiga titik knot, selanjutnya diharapkan pada penelitian selanjutnya dapat menambahkan lebih dari tiga titik knot.
2. Penelitian ini dibatasi pada orde linier, selanjutnya diharapkan pada penelitian selanjutnya dapat menambahkan orde lainnya seperti orde kuadratik dan kubik dengan kombinasi titik knot sehingga memungkinkan memperoleh hasil yang lebih baik.
3. Pada penelitian ini hanya menggunakan variabel prediktor nonbiologis, selanjutnya diharapkan pada penelitian selanjutnya

menambahkan variabel prediktor biologis pada kasus kejadian diare di Provinsi Bali ini.

DAFTAR PUSTAKA

- Budiantara, I.N. 2011. *Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter*. Prosiding Seminar Nasional FMIPA Undiksha.
- Dinas Kesehatan Provinsi Bali. 2022. *Profil Kesehatan Provinsi Bali 2022*. Denpasar: Dinas Kesehatan Provinsi Bali.
- Eubank, R. 1988. *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Eubank, R.L. 1999. *Nonparametric Regression and Spline Smoothing* (2nd ed). New York: Marcel Dekker.
- Eubank, R.L. 2004. *A Simple Smoothing Spline, III, Comput. Stat.* Vol.19(2), pp.227-241.
- Hidayat, R., Yuliani, & Sam, M. 2017. *Model Regresi Nonparametrik dengan Pendekatan Spline Truncated*. Prosiding Seminar Nasional. Vol.3(01), pp.203-210.
- Rositawati, A.F.D., & Budiantara, I.N. 2019. *Pemodelan Indeks Kebahagiaan Provinsi di Indonesia Menggunakan Regresi Nonparametrik Spline Truncated*. Jurnal Sains dan Seni ITS. Vol.8(02).
- Rumlawang, F.Y., Aulele, S.N., & Kasim, N. 2018. *Penentuan Model Regresi Nonparametrik Spline pada Data Pertumbuhan Balita di Desa Nania Provinsi Maluku Tahun 2013-2014*. Barekeng: Jurnal Ilmu Matematika dan Terapan. Vol.12(01), pp.27-32.