

Pemodelan Jumlah Penderita Campak di Indonesia dengan Pendekatan Regresi Nonparametrik Spline

¹⁾Nawaafila, ²⁾Madu Ratna, dan ³⁾I Nyoman Budiantara
^{1,2,3)}Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)
 Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
 e-mail: ²⁾madu_r@statistika.its.ac.id, dan ³⁾i_nyoman_b@statistika.its.ac.id

Abstrak—Campak merupakan salah satu penyakit penyebab kematian di Indonesia. Pada tahun 2013, lebih dari 70 persen kematian di dunia disebabkan karena campak. Jumlah kasus penderita campak di Indonesia tahun 2013 yaitu sebanyak 11.521 kasus. Campak di Indonesia masih tergolong tinggi meskipun pemerintah sudah menerapkan imunisasi campak, hal ini disebabkan pelayanan di Indonesia masih menitikberatkan pada pelayanan kuratif dan rehabilitatif, sehingga permasalahan campak ini sangat kompleks dan krusial. Untuk itu perlu adanya pencegahan, salah satunya dengan mengetahui faktor-faktor yang berpengaruh terhadap jumlah kasus penderita campak di Indonesia. Tujuan penelitian ini adalah untuk mengetahui faktor-faktor yang berpengaruh terhadap jumlah kasus penderita campak dengan menggunakan metode Regresi Nonparametrik Spline. Dalam penelitian ini *Scatterplot* antara variabel respon dengan variabel prediktor tidak membentuk pola tertentu, sehingga metode yang digunakan adalah regresi nonparametrik spline. Pada regresi nonparametrik spline dilakukan pemilihan titik knot optimum berdasarkan nilai Generalized Cross Validation (GCV) terkecil untuk membentuk suatu model. Berdasarkan hasil analisis model terbaik adalah kombinasi knot 2,2,3 dengan variabel yang signifikan terhadap model regresi nonparametrik spline adalah persentase balita kekurangan gizi (x_3), kepadatan penduduk (x_4), dan jumlah tenaga sanitasi (x_5). Model regresi nonparametrik spline yang terbentuk menghasilkan nilai koefisien determinasi sebesar 97,82%.

Kata kunci—Generalized Cross Validation (GCV), Jumlah Penderita Campak, Titik Knot Optimum, Regresi Nonparametrik Spline

I. PENDAHULUAN

CAMPAN merupakan salah satu dari lima penyakit penyebab utama kematian anak di Indonesia [1], meskipun adanya vaksin telah dikembangkan lebih dari 30 tahun yang lalu, virus campak ini menyerang 50 juta orang setiap tahun dan menyebabkan lebih dari 1 juta kematian. Penyakit campak disebabkan oleh virus yaitu virus campak (Paramiksovirus). Campak adalah salah satu penyakit yang mudah menular, cara penularannya adalah melalui droplet atau kontak langsung dengan sekresi hidung atau tenggorokan dari orang yang terinfeksi[2]. Sebagian besar campak menyerang anak-anak dengan rentang usia pra-sekolah dan usia SD. Jika seseorang pernah menderita campak, maka dia akan mendapatkan kekebalan terhadap penyakit tersebut seumur hidup[3].

Indonesia pernah melakukan cakupan imunisasi campak tahun 2010 yaitu sebesar 77,9%, sementara di NTT sebesar 76,1% [4]. Berdasarkan data tersebut dapat disimpulkan bahwa upaya pelayanan di Indonesia masih menitikberatkan

pada pelayanan kuratif dan rehabilitatif dimana pemerintah tidak melakukan adanya pencegahan yang tepat sasaran bukan berlandaskan analisis penelitian yang dilakukan daerah setempat, sehingga variabel yang membutuhkan penanganan penting dalam upaya pencegahan tidak dilaksanakan, malah sebaliknya [5]. Pencegahan secara dini terhadap penyakit campak perlu dilakukan agar tidak terjadi penambahan angka kematian akibat penyakit campak. Dalam hal ini yang dapat dilakukan untuk mengurangi angka kematian akibat campak yaitu dengan mengetahui pola hubungan faktor-faktor yang mempengaruhi banyaknya penderita campak sehingga dapat menentukan langkah apa selanjutnya yang akan dilakukan.

Untuk mengetahui pola hubungan faktor-faktor yang mempengaruhi banyaknya penderita campak maka dapat dilakukan dengan menggunakan analisis regresi. Pada analisis regresi nonparametrik, data tidak membentuk pola tertentu, tetapi metode spline mampu mencari sendiri bentuk pola datanya melalui titik-titik knot optimum yang didapat sehingga subjektivitas dari peneliti dapat diminimalisir. Salah satu kelebihan dari pendekatan Spline adalah memiliki kemampuan yang sangat baik dalam menangani data dengan perilaku yang berubah-ubah pada sub-sub interval tertentu [6]. Metode ini dipilih karena berdasarkan plot pada variabel-variabel yang mempengaruhi jumlah penderita campak di Indonesia tidak diketahui bentuk polanya. Selanjutnya dari hasil penelitian ini diharapkan dapat memberikan informasi kepada menteri kesehatan maupun pemerintahan mengenai faktor-faktor yang mempengaruhi jumlah penderita campak yang ada di Indonesia agar dapat melakukan kebijakan terhadap pelayanan kesehatan yang semakin ditingkatkan lagi.

II. TINJAUAN PUSTAKA

A. Deteksi Outlier

Outlier merupakan pengamatan yang menyimpang jauh dibandingkan pengamatan lainnya. Cara yang dapat dilakukan untuk memutuskan yaitu suatu pengamatan outlier dihilangkan kemudian menganalisis kembali tanpa pengamatan tersebut. Namun dalam beberapa kasus, pengamatan outlier memberikan informasi yang tidak dapat diberikan oleh pengamatan lain dan bahkan perlu diselidiki lebih jauh. Untuk mendeteksi pengamatan outlier dapat dilakukan dengan DFFITS. Adapun persamaan untuk DFFITS sebagai berikut :

$$DFFITS = \frac{(\hat{Y}_i - \hat{Y}_{i-1})}{\sqrt{s_{-i}^2 h_{ii}}}, \text{ untuk } i = 1, 2, \dots, n \quad (1)$$

dimana Y_i merupakan nilai taksiran Y_i dengan menggunakan

pengamatan ke- i , $\hat{Y}_{i-1}$ nilai taksiran Y_i tanpa menggunakan pengamatan ke- i dan h_{ii} adalah elemen diagonal ke- i dari matriks $H = X(X'X)^{-1}X'$. Jika $|DFFITS| > 2 \cdot \sqrt{p/n}$ maka pengamatan *outlier*, dengan p merupakan banyaknya parameter dalam model dan n merupakan banyaknya pengamatan ke- i .

B. Regresi Nonparametrik Spline

Spline memiliki kemampuan yang sangat baik dalam menangani data atau kurva yang berubah-ubah pada sub-sub interval tertentu [8]. Metode regresi nonparametrik spline merupakan metode regresi yang digunakan ketika kurva regresi antara variabel respon dan prediktor tidak diketahui bentuk atau polanya. Model regresi nonparametrik spline dapat dijabarkan sebagai berikut :

$$y_j = f(x_j) + \varepsilon_j, j = 1, 2, 3, \dots, n \quad (2)$$

y_j merupakan variabel respon, sedangkan x_j merupakan variabel prediktor, dan $f(x_j)$ merupakan fungsi regresi yang tidak mengikuti pola tertentu dengan syarat $\varepsilon_j \sim \text{IIDN}(0, \sigma^2)$. Secara umum fungsi spline $f(x_j)$ berorde p dengan titik knot $K_1, K_2, \dots, K_m$ dapat disajikan sebagai berikut.

$$f(x_j) = \sum_{h=0}^p \beta_h x_j^h + \sum_{l=1}^m \beta_{p+l} (x_j - K_l)_+^p \quad (3)$$

$$\text{dengan } (x_j - K_l)_+^p = \begin{cases} (x_j - K_l)^p, & x_j - K_l \geq 0 \\ 0, & x_j - K_l < 0 \end{cases} \quad (4)$$

dimana $K_1, K_2, \dots, K_m$ merupakan titik-titik knot dari perubahan pola perilaku fungsi pada sub-sub interval yang berbeda. Sedangkan nilai p merupakan derajat polinomial. Untuk model regresi nonparametrik spline dapat dituliskan dalam bentuk matriks sebagai berikut.

$$y = X\beta + \varepsilon \quad (5)$$

dimana y merupakan vektor berukuran $n \times 1$, X merupakan matriks berukuran $n \times (p + m + 1)$ dan ε merupakan vektor *error random* berukuran $n \times 1$. Estimasi parameter pada model regresi nonparametrik spline didapat dengan menggunakan metode *Ordinary Least Square* (OLS), estimator β diberikan oleh :

$$\hat{\beta} = (X'X)^{-1} X'y \quad (6)$$

C. Pemilihan Titik Knot Optimal

Untuk menentukan titik knot optimal digunakan metode Generalized Cross Validation (GCV). Model spline yang terbaik dengan titik knot optimal didapat dari nilai GCV yang terkecil [8]. Metode GCV dapat dituliskan sebagai berikut [9].

$$GCV(K) = \frac{MSE(K)}{[n^{-1} \text{trace}(\mathbf{I} - \mathbf{A}(K))]^2} \quad (7)$$

dimana $K = (K_1, K_2, \dots, K_m)$ merupakan titik-titik knot, $\mathbf{I}$ merupakan matriks identitas, dan

$$MSE(K) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (8)$$

$$\text{dimana } \mathbf{A}(K) = X(X'X)^{-1}X', \quad (9)$$

D. Pengujian Parameter Model Regresi

Terdapat dua pengujian yang dilakukan untuk mengetahui signifikansi dari variabel bebas, diantaranya uji serentak dan uji individu.

1. Uji Serentak

Pengujian serentak digunakan untuk mengetahui apakah keserempakan tersebut mempunyai pengaruh yang signifikan terhadap model, dimana pengujian ini menggunakan uji F. Hipotesis dari pengujian ini adalah :

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

$$H_1 : \text{minimal adasatu } \beta_q \neq 0, q = 1, 2, \dots, p + m$$

dimana nilai $p + m$ adalah jumlah parameter dalam model regresi. Untuk model regresi pada persamaan 1, maka *Analysis of Variance* (ANOVA) dapat dinyatakan sebagai berikut [8].

Tabel 1
ANOVA

Sumber Variansi	Derajat Bebas	Jumlah Kuadrat (JK)	Rata-rata Kuadrat (RK)
Regresi	$p + m$	$\beta X'y - ny^2$	$\frac{JK \text{ Regresi}}{df \text{ Regresi}}$
Residual	$n - (p + m) - 1$	$y'y - \beta X'y$	$\frac{JK \text{ Residual}}{df \text{ Residual}}$
Total	$n - 1$	$y'y - n\bar{y}^2$	-

Statistik uji yang digunakan adalah :

$$F_{\text{hitung}} = \frac{RK \text{ Regresi}}{RK \text{ Regresi}} \quad (10)$$

Tolak H_0 jika $F_{\text{hitung}} > F_{\alpha, (p+m, n-(p+m)-1)}$ atau $p\text{-value} < \alpha$ sebesar 0,05, artinya paling sedikit ada satu parameter regresi yang signifikan dalam model.

2. Uji Individu

Pengujian individu digunakan untuk mengetahui apakah parameter secara individual mempunyai pengaruh yang signifikan terhadap variabel respon, dimana pengujian ini menggunakan uji t. Sedangkan hipotesisnya sebagai berikut.

$$H_0 : \beta_q = 0$$

$$H_1 : \beta_q \neq 0$$

dimana $q = 1, 2, \dots, p, p+1, p+2, \dots, p+m$

Statistik uji yang digunakan adalah:

$$t_{\text{hitung}} = \frac{\hat{\beta}_q}{SE(\hat{\beta}_q)} \quad (11)$$

dengan $SE(\hat{\beta}_q)$ diperoleh dari akar dari elemen diagonal dari matriks. Tolak H_0 jika $|t_{\text{hitung}}| > t_{\alpha/2, n-(p+m)-1}$ atau $p\text{-valu} < \alpha$

sebesar 0,05, artinya parameter telah signifikan terhadap variabel respon.

E. Uji Asumsi Residual

Terdapat beberapa asumsi yang harus dipenuhi dalam pengujian asumsi residual, yaitu identik, independen, dan berdistribusi normal.

1. Identik

Asumsi identik terpenuhi jika varians respon sama dengan varians error yaitu sama dengan σ^2 [10]. Hal tersebut merupakan asumsi homoskedastisitas (variens sama).

Hipotesis dari pengujian asumsi residual identik dengan

uji Glejser 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 yang digunakan adalah:

$$F_{hitung} = \frac{\sum_{i=1}^n (|\hat{\varepsilon}| - |\hat{\varepsilon}_i|)^2 / (s-1)}{\sum_{i=1}^n (|\varepsilon_i| - |\hat{\varepsilon}_i|)^2 / (n-s)} \quad (12)$$

dimana k merupakan banyaknya parameter. Tolak H_0 jika $F_{hitung} > F_{(\alpha; s-1, n-s)}$ atau $p\text{-value} < \alpha$ sebesar 0,05.

2. Independen

Asumsi residual independen yaitu tidak ada korelasi antar residual. Salah satu metode yang digunakan untuk mengetahui adanya residual yaitu dengan menggunakan ACF (Autocorrelation Function). Apabila pada plot ACF terdapat lag yang keluar dari batas signifikansi maka dapat dikatakan bahwa terjadi autokorelasi yang artinya asumsi independen tidak terpenuhi. Berikut ini adalah rumus untuk menghitung nilai ACF (rk) dan batas signifikasinya.

$$\rho_k = \frac{\text{cov}(e_t, e_{t+k})}{\sqrt{\text{Var}(e_t)}\sqrt{\text{Var}(e_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (13)$$

dimana:

ρ_k : korelasi antara e_t dan e_{t+k}

γ_k : kovarian antara e_t dan e_{t+k}

γ_k : $\text{var}(e_t) = \text{var}(e_{t+k})$

Interval konfidensi $(1-\alpha) \times 100\%$ korelasi ρ_k diberikan oleh :

$$-t_{(\alpha/2, n-1)} SE(\hat{\rho}_k) < \rho_k < t_{(\alpha/2, n-1)} SE(\hat{\rho}_k) \quad (14)$$

dimana $E(\hat{\rho}_k) = \sqrt{\frac{1}{n}(1 + \rho_1^2 + \dots + 2\rho_{k-1}^2)}$ dan $SE(\hat{\rho}_k) = \sqrt{\frac{1}{n}(1 + \rho_1^2 + \dots + 2\rho_{k-1}^2)}$

Jika terdapat $\hat{\rho}_k$ yang keluar dari batas signifikansi pada plot ACF, maka terdapat autokorelasi.

3. Distribusi Normal

Residual dari model regresi harus mengikuti distribusi Normal yaitu dengan mean nol dan varians σ^2 . Pengujian asumsi residual berdistribusi normal dilakukan dengan menggunakan uji Kolmogorov Smirnov. Hipotesis secara matematis adalah sebagai berikut [11].

H_0 : Residual mengikuti distribusi normal

H_1 : Residual tidak mengikuti distribusi normal

Statistik uji pengujian tersebut adalah sebagai berikut.

$$Z_{hit} = \frac{\sup_x |F_n(x) - F_0(x)|}{\sqrt{\frac{1}{n}}} \quad (15)$$

Tolak H_0 jika $p\text{-value} < \alpha$ sebesar 0,05 atau $Z_{hitung} > Z_\alpha$.

III. METODOLOGI PENELITIAN

A. Data dan Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi internet dan Laporan Tahunan Dinas Kesehatan Jawa Timur Tahun 2013 mengenai data jumlah penderita campak di Indonesia serta beberapa faktor-faktor yang diduga mempengaruhi. Data yang

digunakan merupakan data tahun 2013 dengan unit observasi terdiri dari 32 provinsi (tanpa DKI Jakarta) di Indonesia, karena DKI Jakarta merupakan data *outlier*.

Tabel 2

Variabel Penelitian

Variabel	Nama Variabel
Y	Jumlah penderita campak
x_1	Persentase Cakupan Pemberian Kapsul Vitamin A
x_2	Persentase Cakupan Imunisasi Campak
x_3	Persentase Balita kekurangan gizi
x_4	Kepadatan Penduduk
x_5	Banyak Tenaga Sanitasi

B. Langkah Analisis

Langkah analisis yang dilakukan dalam penelitian adalah sebagai berikut

- Mengidentifikasi karakteristik jumlah penderita campak di Indonesia dan faktor-faktor yang mempengaruhinya.
- Memodelkan Jumlah penderita campak di Indonesia dengan pendekatan *spline* dengan langkah sebagai berikut.
 - Memodelkan data dengan berbagai titik knot yang berbeda menggunakan metode Regresi Nonparametrik *Spline*.
 - Menentukan titik knot optimal dengan menggunakan pendekatan *Generalized Cross Validation* (GCV) dengan memilih nilai GCV terkecil.
 - Menentukan model terbaik dari titik knot optimal.
- Menentukan variabel-variabel yang berpengaruh terhadap jumlah penderita campak di Indonesia dengan tahapan sebagai berikut.
 - Melakukan uji signifikansi dengan uji serentak dan individu terhadap parameter.
 - Melakukan pengujian asumsi residual, seperti asumsi identik, independen dan berdistribusi normal (IIDN). Apabila asumsi tidak terpenuhi, maka dilakukan transformasi terhadap data dan dilakukan analisis kembali yaitu pada langkah 1a.
 - Menginterpretasikan model yang telah terbentuk dan melakukan prediksi.
 - Menarik kesimpulan dari hasil analisis yang didapat.

IV. ANALISIS DAN PEMBAHASAN

A. Deskripsi Karakteristik Jumlah Penderita Campak dan Faktor-Faktor yang diduga mempengaruhi.

Berikut akan disajikan statistika deskriptif dari data penelitian.

Tabel 3

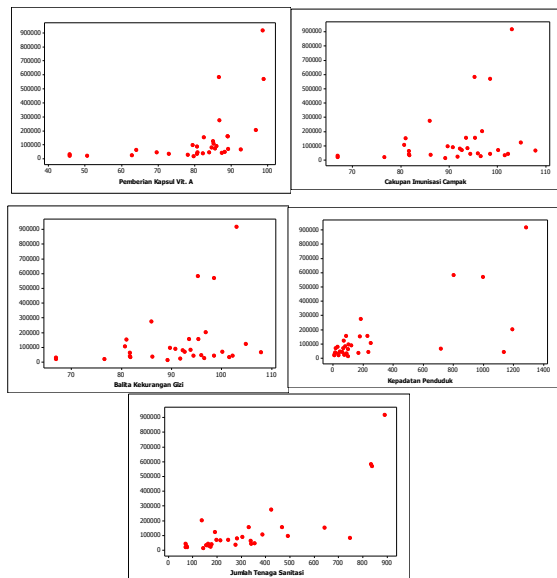
Deskripsi Jumlah Penderita Campak dan Faktor yang diduga Mempengaruhi

Variabel	Rata-rata	Varians	Minimum	Maksimum
Y	158997	56×10^9	15197	916870
X_1	80,32	182,83	45,9	98,9
X_2	91,06	99,97	66,93	100
X_3	21,881	25,788	13,2	33
X_4	268,9	147095,1	9	1285
X_5	324	55353,6	70	890

Berdasarkan Tabel 3 didapatkan bahwa rata-rata jumlah

penderita campak di Indonesia tahun 2013 adalah sebesar 158.997 jiwa dengan varians sebesar 56×10^9 jiwa. Jumlah penderita campak terendah di Indonesia sebesar 15.197 jiwa yang terdapat di provinsi Gorontalo dan jumlah kasus penderita campak tertinggi yaitu terdapat di provinsi Jawa Barat dengan jumlah penderita campak sebesar 916.780 jiwa. Untuk variabel persentase cakupan pemberian kapsul vitamin A (x_1) tertinggi di Jawa Tengah sebesar 98,9 persen, sedangkan terendah di Papua Barat sebesar 45,9 persen. Untuk variabel persentase cakupan imunisasi campak (x_2) tertinggi di Bali sebesar 100 persen, sedangkan terendah di Papua Barat sebesar 66,93 persen. Untuk variabel persentase balita kekurangan gizi (x_3) tertinggi di NTT sebesar 33 persen, sedangkan terendah di Bali sebesar 13,2 persen. Untuk variabel kepadatan penduduk (x_4) tertinggi di Jawa Barat sebesar 1285 jiwa/km², sedangkan terendah di Papua Barat sebesar 9 jiwa/km². Untuk variabel jumlah tenaga sanitasi (x_5) tertinggi di Jawa Barat sebesar 890 tenaga sanitasi, sedangkan terendah di Maluku Utara sebesar 70 tenaga sanitasi.

B. Hubungan Jumlah Penderita Campak dan Faktor-Faktor yang diduga Mempengaruhi



Gambar 1. Scatterplot Jumlah Penderita Campak dan Faktor-Faktor yang diduga mempengaruhi

Berdasarkan Gambar 1 dapat dilihat bahwa antara jumlah penderita campak dan faktor-faktor yang diduga mempengaruhi tidak membentuk pola tertentu. Hal ini menunjukkan bahwa pola hubungan yang tidak membentuk pola tertentu menggunakan metode regresi nonparametrik spline. Regresi nonparametrik Spline digunakan karena mampu mengestimasi data yang tidak memiliki pola tertentu dan memiliki kecenderungan mencari sendiri estimasi data dari pola yang terbentuk.

C. Deteksi Outlier

Berikut akan disajikan tabel mengenai deteksi outlier dengan $DFITS$, jika $|DFITS| > 2 \cdot \sqrt{n/p} = 0,778$, maka data dari pengamatan merupakan data outlier.

Tabel 4
Deteksi Outlier

Provinsi	Dffits	Keterangan
Nanggroe Aceh Darussalam	-1,516	Tidak Outlier
Sumatera Utara	0,205	Tidak Outlier
Sumatera Barat	-0,080	Tidak Outlier
Riau	0,338	Tidak Outlier
Jambi	-0,071	Tidak Outlier
.	.	.
.	.	.
.	.	.
DKI Jakarta	2,781	Outlier
Jawa Barat	-29,773	Tidak Outlier
Jawa Tengah	0,217	Tidak Outlier
Sumatera Barat	-0,080	Tidak Outlier
D I Yogyakarta	-0,146	Tidak Outlier
Jawa Timur	0,429	Tidak Outlier
.	.	.
.	.	.
.	.	.
Maluku Utara	0,420	Tidak Outlier
Papua Barat	0,518	Tidak Outlier
Papua	-0,152	Tidak Outlier

Berdasarkan Tabel 4 didapatkan bahwa DKI Jakarta merupakan data outlier, sehingga selanjutnya dilakukan analisis data tanpa mengikutsertakan DKI Jakarta dalam penelitian.

D. Pemilihan Titik Knot Optimal

Nilai GCV minimum pada pemilihan titik knot optimal dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot dapat disajikan pada Tabel 5.

Tabel 5
Perbandingan Nilai GCV Minimum

Knot	GCV
Satu Knot	4.005.671.556,145
Dua Knot	4.005.671.556,121
Tiga Knot	2.933.699.102,222
Kombinasi Knot	2.468.826.965

Berdasarkan Tabel 5 dapat dilihat bahwa model spline terbaik adalah dengan menggunakan kombinasi knot (1,1,3,3,3), sehingga nantinya model terbaik yang digunakan adalah dengan kombinasi knot (1,1,3,3,3).

E. Penaksiran Parameter Model Regresi Nonparametrik Spline

Berdasarkan nilai GCV terkecil yang diperoleh, didapatkan model terbaik jumlah penderita campak di Indonesia dan faktor-faktor yang mempengaruhinya yang terbentuk dari hasil estimasi parameter regresi nonparametrik spline. Berikut merupakan model terbaik yang telah terbentuk.

$$\hat{y} = -309.183,29 + 1.149,4293x_1 + 8.282,4301(x_1 - 90,247)_+ - 614,2568x_2 - 1.532,0071(x_2 - 94,601)_+ + 1.770,4422x_3 - 86.633,6(x_3 - 20,878)_+ + 63.735,412(x_3 - 21,282)_+ - 5.316,8429(x_3 - 29,363)_+ + 277,53361x_4 + 485,1466(x_4 - 503,776)_+ - 1.085,3170(x_4 - 529,816)_+ + 1.051,9563(x_4 - 1.050,633)_+ - 27,9845x_5 + 8.327,8976(x_5 - 387,959)_+ - 8.701,1125(x_5 - 404,694)_+ + 4.944,7664(x_5 - 739,388)_+$$

F. Uji Parameter

1. Uji Serentak

Uji serentak dilakukan untuk mengetahui apakah variabel prediktor berpengaruh signifikan terhadap variabel respon secara keseluruhan.

Tabel 6
ANOVA Model Regresi Spline Secara Serentak

Sumber Variansi	df	SS	MS	Fhit	P-value
Regresi	16	1,18x10 ¹²	73.899.222.903		
Error	15	17.359.884.219	1.157.325.615	63,8534	0,0000
Total	31	1,2x10 ¹²	-		

Berdasarkan Tabel 6 didapatkan bahwa *p-value* adalah sebesar 0,000 dimana lebih kecil daripada nilai α sebesar 0,05 sehingga Tolak H_0 , artinya minimal terdapat satu parameter yang signifikan terhadap jumlah penderita campak.

2. Uji Individu

Uji individu dilakukan untuk mengetahui variabel mana yang berpengaruh signifikan terhadap jumlah penderita campak di Indonesia. Jika *p-value* < α yaitu sebesar 0,05, maka estimasi parameter signifikan. Berikut akan disajikan hasil uji signifikansi parameter secara individu.

Tabel 7
Pengujian Parameter Secara Individu

Variabel	Parameter	Estimasi	p-value	Keputusan
-	β_0	-309.183,2931	0,0644	Tidak Signifikan
x_1	β_1	1.149,4293	0,223	Tidak Signifikan
	β_2	8.282,4301	0,1452	Tidak Signifikan
x_2	β_3	-6,142,568	0,6611	Tidak Signifikan
	β_4	-1.532,0071	0,47546	Tidak Signifikan
x_3	β_5	17.704,4422	0,0185	Signifikan
	β_6	-86.633,7604	0,3172	Tidak Signifikan
	β_7	63.735,412	0,4467	Tidak Signifikan
	β_8	-5.316,8429	0,7336	Tidak Signifikan
x_4	β_9	27.753,361	0,0312	Signifikan
	β_{10}	4.851,466	0,8723	Tidak Signifikan
	β_{11}	-1.085,3170	0,7249	Tidak Signifikan
	β_{12}	1.051,9563	0,0156	Signifikan
x_5	β_{13}	-279,845	0,7905	Tidak Signifikan
	β_{14}	8.327,8976	0,0008	Signifikan
	β_{15}	-8.701,1125	0,0008	Signifikan
	β_{16}	4.944,7664	0,0000	Signifikan

Berdasarkan Tabel 7 didapatkan bahwa variabel yang berpengaruh signifikan terhadap jumlah penderita campak yaitu variabel persentase balita kekurangan gizi (x_3), kepadatan penduduk (x_4), dan jumlah tenaga sanitasi (x_5). Sedangkan variabel persentase cakupan pemberian kapsul vitamin A (x_1) dan variabel persentase cakupan imunisasi campak (x_2) tidak berpengaruh signifikan sehingga variabel tersebut dihilangkan dari model kemudian dimodelkan kembali dengan menggunakan variabel x_3 , x_4 dan x_5 .

G. Pemilihan Titik Knot Optimal tanpa Variabel x_1 dan x_2

Nilai GCV minimum pada pemilihan titik knot optimal dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot dapat disajikan pada Tabel 8.

Tabel 8
Perbandingan Nilai GCV Minimum Tanpa Variabel x_1 dan x_2

Knot	GCV
Satu Knot	3.262.716.151,193
Dua Knot	2.916.234.764
Tiga Knot	2.125.447.529
Kombinasi Knot	1.896.721.975,976

Berdasarkan Tabel 8 dapat dilihat bahwa model spline terbaik adalah dengan menggunakan kombinasi knot 2,2,3 dengan nilai GCV minimum sebesar 1.896.721.975,976, sehingga nantinya model terbaik yang digunakan adalah dengan kombinasi knot.

H. Penaksiran Parameter Model Regresi Nonparametrik Spline Tanpa Variabel x_1 dan x_2

Hasil estimasi parameter dengan kombinasi knot 2,3,3 dapat disajikan sebagai berikut.

$$\hat{y} = -149.051,444 + 9.207,862x_3 - 17.096,743(x_3 - 22,09)_+ - 4.998,181(x_3 - 29,767)_+ + 216,904x_4 - 379,612(x_4 - 581,898)_+ + 868,862(x_4 - 1.076,673)_+ + 68,812x_5 + 7.951,573(x_5 - 387,959)_+ - 8.336,454(x_5 - 404,694)_+ + 5.509,001(x_5 - 756,122)_+$$

I. Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline Tanpa Variabel x_1 dan x_2

1. Uji Serentak

Berikut ini hasil uji signifikansi secara individu.

Tabel 9
ANOVA Model Regresi Spline Secara Serentak Tanpa Variabel x_1 dan x_2

Sumber Variansi	df	SS	MS	Fhit	P-value
Regresi	10	1,174x10 ¹²	117.360.836.112		
Error	21	26.139.089.549	1.244.718.550	94.287	0,0000
Total	31	1,2x10 ¹²	-		

Berdasarkan Tabel 9 diketahui bahwa *p-value* sebesar 0,000 lebih kecil daripada α sebesar 0,05 sehingga Tolak H_0 , artinya minimal terdapat satu parameter yang signifikan terhadap jumlah penderita campak.

2. Uji Individu

Berdasarkan Tabel 10 didapatkan bahwa jika *p-value* kurang dari nilai α yaitu sebesar 0,05 maka estimasi parameter akan signifikan. Sehingga didapatkan kesimpulan bahwa variabel tersebut nantinya akan dimodelkan, dimana parameter

yang signifikan akan dimodelkan dan diinterpretasikan, sedangkan parameter yang tidak signifikan hanya dimodelkan saja tanpa dilakukan interpretasi model. Berikut ini hasil uji signifikansi parameter secara individu.

Tabel 10

Pengujian Parameter Secara Individu Tanpa Tanpa Variabel x_1 dan x_2

Variabel	Parameter	Estimasi	p-value	Keputusan
x_3	Konstan β_0	-149.051,444	0,0661	Tidak Signifikan
	β_1	9.207,862	0,035	Signifikan
	β_2	-17.096,743	0,0297	Signifikan
	β_3	-4.998,181	0,782	Tidak Signifikan
x_4	β_4	216,904	0,0201	Signifikan
	β_5	-379,612	0,0642	Tidak Signifikan
	β_6	868,862	0,0269	Signifikan
x_5	β_7	68,812	0,4418	Tidak Signifikan
	β_8	7.951,573	0,0008	Signifikan
	β_9	-8.336,454	0,0007	Signifikan
	β_{10}	5.509,001	0,0007	Signifikan

J. Uji Asumsi Residual

1. Uji Identik

Model regresi nonparametrik spline harus memenuhi asumsi identik, yaitu tidak terjadi kasus heteroskedastisitas pada residualnya atau variansi residual harus homogen. Salah satu uji yang digunakan untuk uji identik adalah uji Glejser. Berikut ini akan disajikan hasil uji identik dengan uji Glejser berdasarkan tabel ANOVA.

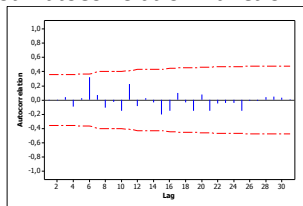
Tabel 11
ANOVA Uji Glejser

Sumber Variansi	df	SS	MS	Fhit	P-value
Regresi	10	4.013.506.742	401.350.674		
Error	21	5.322.739.013	253.463.763	1,5835	0,1798
Total	31	9.336.245.755	-		

Berdasarkan Tabel 11 dapat diketahui bahwa p -value adalah sebesar 0,1798 dimana lebih besar dari nilai α sebesar 0,05 sehingga gagal tolak H_0 , artinya tidak terjadi adanya kasus heteroskedastisitas, atau variansi residual homogen. Sehingga model regresi nonparametrik spline memenuhi asumsi identik

2. Independen

Asumsi independen, artinya tidak terdapat adanya autokorelasi pada residual. Cara mendeteksinya adalah dengan menggunakan plot Autocorrelation Function (ACF).



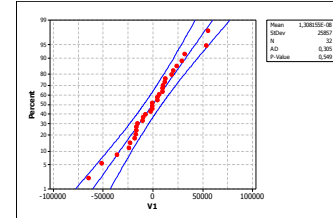
Gambar 2. Plot ACF Residual

Berdasarkan Gambar 2 tidak terlihat adanya autokorelasi yang keluar batas toleransi. Hal ini berarti asumsi

independen pada residual telah terpenuhi.

3. Uji Distribusi Normal

Pengujian distribusi normal dilakukan dengan menggunakan uji *Kolmogorov-Smirnov*. Berdasarkan Gambar 3 didapatkan bahwa p -value adalah 0,549 lebih besar dari nilai α sebesar 0,05 sehingga gagal tolak H_0 , artinya bahwa residual dari model regresi nonparametrik spline telah memenuhi asumsi berdistribusi normal.



Gambar 3. Uji Kolmogorov-Smirnov

K. Interpretasi Model Regresi Nonparametrik Spline

1. Pengaruh persentase balita kekurangan gizi (x_3) apabila variabel x_4 dan x_5 konstan adalah sebagai berikut.

$$\hat{y} = 9207,862x_3 - 17096,743(x_3 - 22,09)_+ - 4998,181(x_3 - 29,767)_+ \\ = \begin{cases} 9207,862x_3 & ; x_3 < 22,09 \\ -7.887,881x_3 + 377.667,053 & ; x_3 \geq 22,09 \end{cases}$$

Apabila persentase balita kekurangan gizi (x_3) berada pada nilai kurang dari 22,09, maka apabila persentase balita kekurangan gizi naik satu persen maka jumlah penderita campak akan cenderung naik sebesar 9.208 kasus. Selanjutnya Sedangkan apabila persentase balita kekurangan gizi (x_3) berada pada nilai lebih dari 22,09, maka apabila naik satu persen, maka jumlah penderita campak akan cenderung turun sebesar 7.888 kasus

2. Pengaruh kepadatan penduduk (x_4) apabila variabel dan x_5 konstan adalah sebagai berikut.

$$\hat{y} = 216.904x_4 - 379,612(x_4 - 581,898)_+ + 868,862(x_4 - 1.076,673)_+ \\ = \begin{cases} -216,904x_4 & ; x_4 < 581,898 \\ 1.085,766x_4 - 935.480,256 & ; x_4 \geq 1.076,673 \end{cases}$$

Pada kondisi pertama, provinsi dengan kepadatan penduduk kurang dari 581,898 jiwa per kilometer persegi, jika kepadatan penduduk berkurang 1 jiwa per kilometer persegi, maka jumlah penderita campak akan berkurang sebesar 217 kasus. Pada kondisi kedua, provinsi dengan kepadatan penduduk lebih dari 1.076,673 jiwa per kilometer persegi, jika kepadatan penduduk bertambah 1 jiwa per kilometer persegi, maka jumlah penderita campak akan cenderung bertambah sebesar 1.086 kasus dan provinsi yang termasuk adalah provinsi DIY, Banten dan Jawa Barat. Untuk provinsi-provinsi lainnya masuk dalam kategori kondisi pertama.

3. Pengaruh jumlah tenaga sanitasi (x_5) apabila variabel x_3 dan x_4 konstan adalah sebagai berikut.

$$\hat{y} = 68,812x_5 + 7.951,573(x_5 - 387,959)_+ - 8.336,454(x_5 - 404,694)_+ + 5.509,001(x_5 - 756,122)_+ \\ = \begin{cases} 8.020,385x_5 - 3.084.884,31 & ; x_5 < 404,694 \\ -8.267,642x_5 - 3.373.712,915 & ; 404,694 \leq x_5 < 756,122 \\ 5.577,813x_5 - 4.165.476,854 & ; x_5 \geq 756,122 \end{cases}$$

Apabila provinsi dengan jumlah tenaga sanitasi kurang dari 405 tenaga sanitasi, jika jumlah tenaga sanitasi bertambah sebanyak tenaga sanitasi, maka jumlah penderita campak akan cenderung bertambah sebesar 8.020 kasus. Pada kondisi kedua, provinsi dengan dengan jumlah tenaga sanitasi antara 405 tenaga sanitasi hingga 756 tenaga sanitasi, jika jumlah tenaga sanitasi bertambah sebanyak satu tenaga sanitasi, maka jumlah penderita campak akan cenderung berkurang sebesar 8.268 kasus dan provinsi yang termasuk adalah provinsi Sumatra Utara, Sumatra Selatan, NTT, Sulawesi Selatan, dan NAD. Sedangkan kondisi ketiga, provinsi dengan jumlah tenaga sanitasi lebih dari 756 tenaga sanitasi, jika jumlah tenaga sanitasi bertambah sebanyak satu tenaga sanitasi, maka jumlah penderita campak akan cenderung bertambah sebesar 5.578 kasus dan provinsi yang termasuk adalah provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat.

V. KESIMPULAN DAN SARAN

Jumlah penderita campak tertinggi terdapat di Jawa Barat, sedangkan terendah di Gorontalo. Model regresi yang terbentuk adalah model regresi nonparametrik spline dengan kombinasi knot (2,2,3) dan R^2 sebesar 97,82%. Berdasarkan pengujian parameter didapatkan bahwa semua variabel signifikan terhadap model, diantaranya adalah persentase balita kekurangan gizi (x_3), kepadatan penduduk (x_4), dan jumlah tenaga sanitasi (x_5).

Saran yang diharapkan adalah bagi pemerintah dapat memperhatikan provinsi-provinsi yang memiliki kasus campak tinggikan variabel-variabel yang signifikan terhadap jumlah penderita campak di Indonesia. Selain itu agar terus memberikan penyuluhan kepada masyarakat terutama mengenai pentingnya pelaksanaan imunisasi campak dan pemberian vitamin A untuk balita. Untuk provinsi-provinsi terpencil dalam kasus penanganan masalah kesehatan terkadang susah terjangkau, sehingga diperlukan transportasi dan SDM lebih dalam memudahkan pelayanan kesehatan.

DAFTAR PUSTAKA

- [1] Gujarati, D. N. (2004). *Basic Econometric 4th Edition*. New York: McGraw Hill Companies, Inc.
- [2] WHO. (2007). *Retrieved from Manual For The Laboratory Diagnosis of Measles and Rubella Virus*: <http://www.who.int/vaccines-documents/>
- [3] Riskesdas. (2013). *Riset Kesehatan Dasar*. Kementerian Kesehatan RI, Badan Penelitian dan Pengembangan Kesehatan. Retrieved from <http://www.depkes.go.id/resources/download/general/Hasil%20Riset%20Riset%20Kesehatan%202013.pdf>
- [4] Kemenkes RI, (2010). *Profil Kesehatan Indonesia 2010*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2010.
- [5] Kemenkes RI. 2013. *Data dan Informasi Kesehatan Nusa Tenggara Timur*. Jakarta.
- [6] Budiantara, I. N. (2005). *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik, Berkala MIPA*. Surabaya: Institut Teknologi Sepuluh Nopember.
- [7] Budiantara, I. N. (2009). Spline dalam Regresi Nonparametrik dan Semiparametrik : Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang. *Institut Teknologi Sepuluh Nopember, Surabaya Vol 15*, 49-59.
- [8] Budiantara, I. N. (2005). *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik, Berkala MIPA*. Surabaya: Institut Teknologi Sepuluh Nopember.
- [9] Wahba, G. (1990). *Spline Models for Observation Data*. Pennsylvania: University of Wisconsin at Madison.
- [10] Draper, N. R., & Smith, H. (1992). *Analisis Regresi Terapan*. Jakarta: PT Gramedia Pustaka Utama.
- [11] Daniel, W. W. (1989). *Statistika Nonparametrik Terapan*. Jakarta: PT Gramedia Pustaka Utama.