

# Pemodelan Faktor yang Memengaruhi Angka Kasus *Tuberculosis* di Surabaya Menggunakan Pendekatan Regresi Nonparametrik *Spline Truncated*

I Gusti Putu Surya Darma, Madu Ratna, dan I Nyoman Budiantara  
Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data,  
Institut Teknologi Sepuluh Nopember (ITS)  
*e-mail*: i\_nyoman\_b@statistika.its.ac.id

**Abstrak**—*Tuberculosis* merupakan penyakit infeksi kronik dan menular yang erat kaitannya dengan keadaan lingkungan dan perilaku masyarakat. *Tuberculosis* merupakan penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penggunaan metode regresi nonparametrik *Spline Truncated* pada angka kasus *Tuberculosis* di Kota Surabaya belum pernah dilakukan sebelumnya. Oleh karena itu, penelitian ini bertujuan untuk memodelkan faktor-faktor yang memengaruhi angka kasus *Tuberculosis* di Kota Surabaya menggunakan pendekatan regresi nonparametrik *Spline Truncated*. Dengan menggunakan data yang bersumber dari Dinas Kesehatan Kota Surabaya yakni data publikasi Profil Kesehatan Kota Surabaya tahun 2017 dengan unit observasi yang digunakan sebanyak 31 kecamatan yang ada di Surabaya, didapatkan hasil penelitian bahwa model terbaik regresi nonparametrik *Spline Truncated* untuk angka kasus *Tuberculosis* di Kota Surabaya tahun 2017 merupakan model kombinasi knot (3,2,3) dengan nilai GCV minimum sebesar 1338,53. Sebanyak tiga variabel prediktor yang digunakan berpengaruh signifikan. Koefisien determinasi atau  $R^2$  yang menunjukkan ukuran kebaikan model adalah sebesar 74,3515%.

**Kata Kunci**— GCV, Knot, *Spline*, Surabaya, *Tuberculosis*.

## I. PENDAHULUAN

KESEHATAN adalah hal yang selalu didambakan oleh setiap orang. Apabila seseorang kesehatannya menurun akan menimbulkan berbagai penyakit yang pada akhirnya akan menghambat kegiatan yang rutin dilaksanakan. Menurut WHO, kesehatan adalah keadaan sejahtera fisik, mental, sosial tanpa ada keluhan sama sekali. Kesehatan juga dinyatakan dalam UU RI No. 23 tahun 1992 mengandung dimensi mental dan sosial, dimana diartikan sebagai berikut “Kesehatan adalah keadaan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup produktif secara sosial dan ekonomi”. Kesehatan apabila tidak terjaga dengan baik akan menimbulkan penyakit.

*Tuberculosis* merupakan penyakit infeksi kronik dan menular yang erat kaitannya dengan keadaan lingkungan dan perilaku masyarakat. *Tuberculosis* merupakan penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit ini ditularkan melalui udara yaitu lewat percikan ludah, bersin dan batuk. Penyakit *Tuberculosis* biasanya menyerang paru-paru dan dapat pula menyerang organ tubuh yang lain. Penyakit *Tuberculosis* paru dapat menyerang siapa saja dan kebanyakan berasal dari kelompok sosial ekonomi rendah dan tingkat pendidikan yang rendah [1].

Untuk penelitian ini, digunakan metode regresi nonparametrik *Spline Truncated* dikarenakan variasi datanya adalah kecamatan yang ada di Surabaya dan akan membentuk pola acak. *Spline* akan membagi kurva regresi berdasarkan titik knot optimal sehingga *error* yang dihasilkan pada model akan kecil dibandingkan dengan metode lain. Selain itu, penggunaan regresi nonparametrik *Spline Truncated* sangat baik digunakan untuk mengatasi pola data yang memiliki perubahan perilaku pada sub-sub interval tertentu. Untuk memilih model regresi *spline* yang terbaik, digunakan dengan memilih model dengan titik knot yang optimal. Titik knot merupakan titik yang terdapat pada perubahan pola perilaku fungsi. Salah satu metode yang dapat digunakan untuk menentukan titik knot optimal adalah *Generalized Cross Validation* (GCV). Titik knot optimal diperoleh dari nilai GCV minimum dan GCV ini mempunyai keunggulan dibandingkan metode pemilihan titik knot lainnya seperti *Cross Validation*, *Unbiased Risk* dan *Generalized Maximum Likelihood* diantaranya GCV memiliki sifat sederhana dan efisien dalam perhitungan, optimal secara asimtotik, dan invarian terhadap transformasi dan tidak memerlukan informasi terhadap  $\sigma^2$ .

Oleh karena itu, penelitian ini bertujuan untuk memodelkan faktor-faktor yang memengaruhi angka kasus *Tuberculosis* di Kota Surabaya menggunakan pendekatan regresi nonparametrik *Spline Truncated*.

## II. TINJAUAN PUSTAKA

### A. Statistika Deskriptif

Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian data sehingga memberikan informasi yang berguna [2]. Statistika deskriptif yang digunakan pada penelitian ini adalah mean dan varians.

Untuk visualisasi data yang digunakan pada penelitian ini adalah *scatterplot*. *Scatterplot* yang diterjemahkan menjadi diagram pencar adalah diagram yang digunakan untuk menjelaskan hubungan antara dua variabel yang berbentuk garis lurus yang disebut dengan hubungan linear [3].

### B. Regresi Nonparametrik *Spline Truncated*

Regresi nonparametrik adalah metode pendekatan regresi yang digunakan saat bentuk kurva regresinya tidak diketahui atau tidak terdapat informasi masa lalu yang lengkap tentang pola datanya [4]. Model regresinya secara umum adalah sebagai berikut.

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

dengan:

$y_i$  : variabel respon ke- $i$

$f$  : kurva regresi yang diasumsikan bentuknya tidak diketahui

$\varepsilon_i$  : *error random* ke- $i$  yang diasumsikan memenuhi IIDN  $(0, \sigma^2)$

*Spline* merupakan salah satu model regresi nonparametrik. Model *Spline* memiliki fleksibilitas yang tinggi dan memiliki kemampuan sangat baik untuk menangani data yang perilakunya berubah pada sub interval tertentu. Salah satu kelebihan pendekatan *Spline* adalah model ini cenderung mencari sendiri estimasi data ke mana pola data tersebut bergerak. Kelebihan ini terjadi karena dalam *Spline* terdapat titik-titik knot. Menurut [5], titik knot merupakan titik perpaduan bersama dimana terjadi pola perubahan perilaku dari suatu fungsi pada selang yang berbeda. Diberikan suatu basis dari ruang *Spline* berorde  $p$  sebagai berikut.

$$\{1, x, x^2, \dots, x^p, (x - \lambda_1)_+, (x - \lambda_2)_+, \dots, (x - \lambda_k)_+\} \quad (2)$$

Fungsi potongan (*Truncated*) diberikan oleh:

$$(x - \lambda_k)_+^p = \begin{cases} (x - \lambda_k)^p & , x \geq \lambda_k \\ 0 & , x < \lambda_k \end{cases} \quad (3)$$

Pada regresi *Spline*, fungsi regresi  $f$  dapat ditulis sebagai berikut.

$$f(x) = \sum_{j=0}^p \beta_j x^j + \sum_k^K \beta_{p+k} (x - \lambda_k)_+^p \quad (4)$$

Sehingga model regresi *Spline* dapat dituliskan sebagai berikut.

$$y = f(x) + \varepsilon \quad (5)$$

atau dapat juga ditulis sebagai berikut.

$$y = \sum_{j=0}^p \beta_j x^j + \sum_k^K \beta_{p+k} (x - \lambda_k)_+^p + \varepsilon \quad (6)$$

dimana:

- $\beta_j$  : parameter model polinomial
- $x$  : variabel predictor
- $\beta_{p+k}$  : parameter pada komponen *Truncated*
- $k = 1, 2, 3, \dots, K$
- $K$  : banyaknya knot
- $\lambda_k$  : titik-titik knot

### C. Pemilihan Titik Knot Optimal

Titik knot adalah titik perpaduan bersama dimana terdapat perubahan perilaku pola data pada interval yang berlainan [6]. Model regresi *Spline* terbaik tergantung pada titik knot yang optimal. Salah satu metode yang dapat digunakan untuk menentukan titik knot optimal adalah *Generalized Cross Validation* (GCV). Titik knot optimal diperoleh dari nilai GCV minimum. Secara umum, fungsi GCV adalah sebagai berikut.

$$GCV(\lambda_1, \lambda_2, \dots, \lambda_K) = \frac{MSE(\lambda_1, \lambda_2, \dots, \lambda_K)}{(n^{-1} \text{trace}[I - A(\lambda_1, \lambda_2, \dots, \lambda_K)])^2} \quad (7)$$

Dimana  $\mathbf{I}$  adalah matriks identitas,  $n$  adalah jumlah pengamatan,  $A(\lambda_1, \lambda_2, \dots, \lambda_K)$  merupakan matriks  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , serta  $MSE(\lambda_1, \lambda_2, \dots, \lambda_K)$  diberikan oleh:

$$MSE(\lambda_1, \lambda_2, \dots, \lambda_K) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

### D. Pengujian Parameter Model Regresi

Pengujian parameter model digunakan untuk mengetahui apakah variabel prediktor berpengaruh atau tidak terhadap variabel respon. Tahap dalam pengujian parameter adalah pengujian secara serentak dan pengujian secara parsial atau individu.

#### i. Pengujian Secara Serentak

Pengujian secara serentak dilakukan untuk mengetahui signifikansi parameter model regresi secara bersama-sama. Hipotesis untuk pengujian parameter secara serentak adalah sebagai berikut.

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

$$H_1 : \text{minimal terdapat satu } \beta_j \neq 0, \quad j = 1, 2, 3, \dots, p+r$$

Perhitungan nilai statistik uji F didapatkan dari Analisis Ragam (ANOVA) seperti yang ditunjukkan Tabel 1.

Sumber Variasi	Df	Sum of Square (SS)	Mean Square (MS)	$F_{hitung}$
Regresi	$p+r$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{SS_{regresi}}{df_{regresi}}$	
Error	$n-(p+r)-1$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\frac{SS_{error}}{df_{error}}$	$\frac{MS_{regresi}}{MS_{error}}$
Total	$n-1$	$\sum_{i=1}^n (y_i - \bar{y})^2$		

Daerah penolakan  $H_0$  adalah  $F_{hitung} > F_{(\alpha, (p+r), (n-(p+r)-1))}$  atau  $P\text{-value} < \alpha$ . Jika  $H_0$  ditolak maka dapat disimpulkan bahwa minimal terdapat satu parameter pada model regresi yang signifikan terhadap model.

#### ii. Pengujian Secara Parsial

Berikut merupakan hipotesis untuk pengujian secara parsial:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \quad j = 1, 2, \dots, p+r$$

Pengujian secara parsial dilakukan dengan menggunakan uji  $t$  [7]. Statistik uji yang digunakan adalah sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (9)$$

dimana  $SE(\hat{\beta}_j)$  adalah *standard error*  $\hat{\beta}_j$  yang diperoleh dari akar elemen diagonal ke- $j$ ,  $j = 1, 2, 3, \dots, p+r$  dari matriks yang dapat diurai seperti berikut.

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{var}\left[(X'X)^{-1}X'Y\right] \\ &= (X'X)^{-1}X'\text{var}(Y)\left[(X'X)^{-1}X'\right]' \\ &= (X'X)^{-1}X'(\sigma^2I)X(X'X)^{-1} \quad (10) \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Daerah penolakan yaitu, tolak  $H_0$  jika  $|t_{hitung}| > t_{\frac{\alpha}{2};(n-(p+r)-1)}$  atau  $p\text{-value} < \alpha$ . Sehingga menghasilkan kesimpulan bahwa parameter berpengaruh signifikan terhadap model [7].

**E. Kriteria Keباikannya Model Regresi**

Koefisien determinasi merupakan nilai proporsi keragaman total di sekitar nilai  $\bar{y}$  yang dapat dijelaskan oleh model regresi [7]. Semakin tinggi nilai  $R^2$  yang dihasilkan suatu model, maka semakin baik pula variabel-variabel prediktor dalam model dalam menjelaskan variabilitas variabel respon. Berikut merupakan rumus dari koefisien determinasi ( $R^2$ ).

$$R^2 = \frac{SS_{regresi}}{SS_{total}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

**F. Pengujian Asumsi Residual Model Regresi**

**i. Uji Asumsi Identik**

Pengujian asumsi identik bertujuan untuk mengetahui homogenitas variansi residual. Pengujian asumsi identik dilakukan dengan uji Glejser. Uji Glejser dilakukan dengan cara meregresikan absolut residual dengan variabel prediktornya [8].

**ii. Uji Asumsi Independen**

Asumsi independen merupakan asumsi dari model regresi yang mengharuskan tidak terdapat korelasi antar residual. Uji yang digunakan untuk mendeteksi kasus autokorelasi adalah uji *Durbin-Watson* [7].

**iii. Uji Asumsi Berdistribusi Normal**

Untuk mengetahui apakah residual berdistribusi normal, secara visual dapat dilakukan dengan menggunakan *normality plot*. Apabila *plot* berada di sekitar garis regresi, maka dapat dikatakan bahwa residual berdistribusi normal. Sebaliknya, jika *plot* tidak berada di sekitar garis regresi, maka dapat dikatakan residual tidak berdistribusi normal. Salah satu pengujian normalitas adalah uji *Shapiro-Wilk*. Uji *Shapiro-Wilk* digunakan apabila sampel yang digunakan kurang dari lima puluh agar menghasilkan keputusan yang akurat [9].

**G. Tuberculosis**

*Tuberculosis* merupakan penyakit infeksi kronik dan menular yang erat kaitannya dengan keadaan lingkungan dan perilaku masyarakat. *Tuberculosis* merupakan penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium Tuberculosis*. Penyakit ini ditularkan melalui udara yaitu lewat percikan ludah, bersin dan batuk. Penyakit *Tuberculosis* biasanya menyerang paru-paru dan dapat pula menyerang organ tubuh yang lain. Penyakit *Tuberculosis* paru dapat menyerang siapa saja dan kebanyakan berasal dari kelompok sosial ekonomi rendah dan tingkat pendidikan yang rendah [1].

**III. METODOLOGI PENELITIAN**

**A. Sumber Data**

Sumber data yang digunakan pada penelitian ini merupakan data sekunder yang diolah kembali menjadi data angka kasus dan persentase kasus yang terlampir pada Lampiran 1 dan bersumber dari Dinas Kesehatan Kota Surabaya yakni data publikasi Profil Kesehatan Kota Surabaya tahun 2017 [10] dengan unit observasi yang digunakan sebanyak 31 kecamatan yang ada di Surabaya.

**B. Variabel Penelitian**

Pada penelitian ini terdapat satu variabel respon (Y) dan tiga variabel prediktor (X) yaitu sebagai berikut.

Tabel 2.  
Variabel Penelitian

Variabel	Deskripsi	Skala
Y	Angka Kasus <i>Tuberculosis</i> per 100.000 kejadian	Rasio
X <sub>1</sub>	Persentase Rumah Tangga yang Memiliki Rumah Sehat	Rasio
X <sub>2</sub>	Persentase Rumah Tangga ber-PHBS	Rasio
X <sub>3</sub>	Persentase Akses Fasilitas Sanitasi yang Layak	Rasio

**C. Langkah Analisis**

Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut.

1. Melakukan deskripsi data mengenai karakteristik angka penderita *Tuberculosis* di kecamatan yang ada di Surabaya pada tahun 2017 dan faktor-faktor yang diduga memengaruhinya.
2. Melakukan pemodelan jumlah penderita *Tuberculosis* di kecamatan yang ada di Surabaya pada tahun 2017 menggunakan pendekatan *Spline Truncated* dengan langkah-langkah sebagai berikut.
  - a. Mengetahui pola hubungan antara variabel respon dengan variabel prediktor dengan membuat *scatterplot*.
  - b. Memodelkan variabel respon dengan menggunakan model regresi nonparametrik *Spline Truncated*.
  - c. Melakukan pemilihan titik knot optimal yang didapatkan dari nilai GCV yang paling minimum.
  - d. Mendapatkan model regresi nonparametrik *Spline* terbaik dengan titik knot yang optimal.
  - e. Melakukan uji signifikansi parameter serentak dan parsial.
  - f. Melakukan uji asumsi identik, independen, dan berdistribusi normal (IIDN) dari model regresi nonparametrik *Spline*.
  - g. Membuat interpretasi model yang didapatkan serta menarik kesimpulan akhir.

**IV. ANALISIS DAN PEMBAHASAN**

**A. Karakteristik Data**

*Tuberculosis* merupakan salah satu penyakit infeksi kronik dan menular yang erat kaitannya dengan keadaan lingkungan dan perilaku masyarakat. Penyakit *Tuberculosis* biasanya menyerang paru-paru dan dapat pula menyerang organ tubuh yang lain. Dengan menggunakan mean, variansi, nilai minimum dan maksimum akan dijelaskan karakteristik kasus *Tuberculosis* di Kota Surabaya dalam Tabel 3.

Tabel 3.

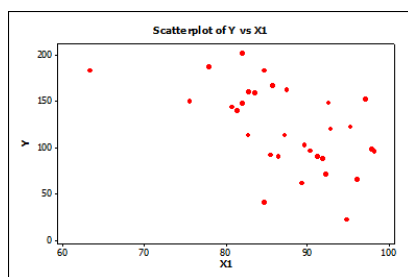
Karakteristik Data <i>Tuberculosis</i> di Kota Surabaya				
Variabel	Mean	Varians	Maksimum	Minimum
			Nilai	Nilai
Y	121,87	2025,78	201,73	22,59
X1	87,20	55,59	98,22	63,40
X2	72,47	210,26	96,96	30,47
X3	98,00	3,72	100,00	91,75

Berdasarkan tabel diatas, rata-rata dari angka kasus *Tuberculosis* di Kota Surabaya terdapat 122 kasus tiap 100.000 kejadian, dengan variansi atau persebaran data yang sangat besar yakni 2025,784, artinya di kecamatan tertentu angka kasus *Tuberculosis*nya ada yang besar dan di kecamatan lainnya angka kasus *Tuberculosis*nya ada yang kecil. Angka kasus *Tuberculosis* tertinggi terdapat di Kecamatan Tandes dan yang terendah di Kecamatan Gayungan.

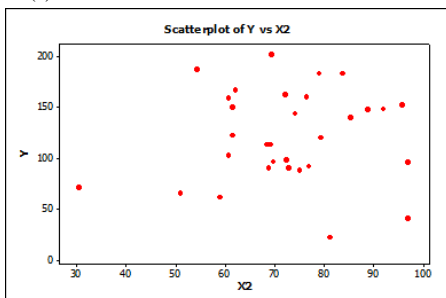
Rata-rata dari persentase rumah tangga yang memiliki rumah sehat adalah sebesar 87,20 dimana artinya sebagian besar rumah tangga di tiap kecamatan di Kota Surabaya memiliki rumah yang sehat. Begitupula dengan rumah tangga yang berperilaku hidup bersih dan sehat memiliki rata-rata persentase yang tinggi sebesar 72,47 yang artinya rumah tangga di tiap kecamatan di Kota Surabaya telah melakukan perilaku hidup bersih dan sehat, akan tetapi persebarannya masih belum merata dikarenakan variansi yang tinggi. Diikuti oleh akses fasilitas sanitasi yang layak, rumah tangga di tiap kecamatan di Kota Surabaya memiliki akses untuk fasilitas sanitasi layak seperti jamban yang sudah baik, dan dapat dikatakan aksesnya sama antara satu kecamatan dengan yang lainnya ditandai dengan nilai variansi yang rendah sebesar 3,72.

**B. Scatterplot Angka Kasus *Tuberculosis* di Surabaya dan Faktor-Faktor yang Diduga Memengaruhinya**

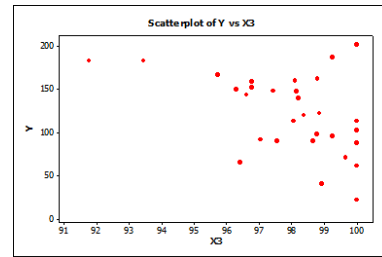
Identifikasi pola hubungan antar variabel prediktor terhadap variabel respon dilakukan untuk mengetahui bagaimana pola hubungan masing-masing variabel prediktor terhadap variabel respon. Dengan menggunakan *scatterplot* sebagai visualisasi data, berikut adalah hasil dari pola hubungan antar variabelnya ditunjukkan pada Gambar 1.



Gambar 1. Scatterplot Angka Kasus *Tuberculosis* dengan Faktor yang Memengaruhi (a)



Gambar 1. Scatterplot Angka Kasus *Tuberculosis* dengan Faktor yang Memengaruhi (b)



Gambar 1. Scatterplot Angka Kasus *Tuberculosis* dengan Faktor yang Memengaruhi (c).

Dikarenakan data tidak memiliki pola tertentu dan memiliki perubahan pola data pada sub interval tertentu maka metode regresi nonparametrik *Spline Truncated* dapat digunakan untuk melakukan pemodelan.

**C. Pemilihan Model Terbaik Regresi Nonparametrik *Spline Truncated***

Setelah dilakukan pemodelan regresi nonparametrik *Spline Truncated* menggunakan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot, maka akan dilihat perbandingan nilai GCV minimum untuk menentukan model terbaiknya. Berikut merupakan perbandingan nilai GCV satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot.

Tabel 4. Nilai GCV Masing-masing Knot

Jumlah Knot	GCV Minimum	Banyak Parameter
Satu Knot	1664,919	7
Dua Knot	1585,182	10
Tiga Knot	1496,895	13
Kombinasi Knot (3,2,3)	1338,538	11

Tabel 4 menunjukkan hasil nilai GCV minimum dari masing-masing knot. Pemodelan dilihat dari nilai GCV minimum. Dari tabel tersebut, dapat disimpulkan bahwa nilai GCV paling minimum yaitu pada kombinasi knot (3,2,3), sehingga kombinasi knot (3,2,3) digunakan sebagai model terbaik regresi nonparametrik *Spline Truncated* dengan jumlah parameter model sebanyak 11 termasuk konstanta  $\beta_0$ .

**D. Pengujian Signifikansi Parameter**

**i. Pengujian Serentak**

Hipotesis dari uji serentak adalah sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{11} = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0; j = 1, 2, \dots, 11$$

Tabel 5.

Tabel ANOVA Pengujian Parameter Secara Serentak				
Sumber Variasi	Derajat Bebas	SS	MS	F <sub>hitung</sub>
Regresi	11	45186,03	4107,821	5,007
Error	19	15587,5	820,3946	
Total	30	60773,53		

Diambil keputusan tolak  $H_0$  jika nilai  $F_{hitung}$  lebih besar dari nilai  $F_{tabel}$  yaitu  $F_{tabel(0.05,11,19)}$ . Diketahui bahwa nilai  $F_{hitung}$  sebesar 5,007 dan nilai  $F_{tabel}$  sebesar 2,31, sehingga dapat diambil keputusan tolak  $H_0$  yang artinya terdapat minimal satu parameter yang berpengaruh signifikan terhadap angka kasus *Tuberculosis*. Dari kesimpulan ini, maka dapat dilanjutkan uji individu atau parsial.

**ii. Pengujian Parsial**

Hipotesis yang digunakan dalam pengujian parameter secara individu adalah sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, j = 1, 2, \dots, 11$$

Tabel 6.  
Hasil Pengujian Parameter Secara Parsial

Variabel	Parameter	t <sub>hitung</sub>	P-value	Keputusan
	$\beta_0$	2,53	0,02	Tolak
X <sub>1</sub>	$\beta_1$	1,01	0,32	Gagal Tolak
	$\beta_2$	-2,8	0,01	Tolak
	$\beta_3$	-0,87	0,39	Gagal Tolak
	$\beta_4$	2,54	0,01	Tolak
	$\beta_5$	-1,81	0,08	Gagal Tolak
X <sub>2</sub>	$\beta_6$	3,99	0,07	Tolak
	$\beta_7$	-4,22	0,00	Tolak
	$\beta_8$	-2,21	0,03	Tolak
X <sub>3</sub>	$\beta_9$	2,15	0,04	Tolak
	$\beta_{10}$	-1,93	0,06	Gagal Tolak
	$\beta_{11}$	1,57	0,13	Gagal Tolak

Jika nilai jika |t<sub>hitung</sub>| lebih besar dari nilai t<sub>tabel</sub> yaitu t<sub>0,025;(19)</sub> sebesar 2,093 maka keputusannya adalah tolak H<sub>0</sub>. Jika P-value kurang dari  $\alpha=0,05$  maka keputusannya adalah tolak H<sub>0</sub>. Berdasarkan Tabel 6 diketahui bahwa terdapat 7 parameter keputusannya adalah tolak H<sub>0</sub> yaitu parameter tersebut signifikan terhadap model. Sedangkan ada 5 parameter lainnya yang memiliki P-value lebih dari  $\alpha$ , keputusannya adalah gagal tolak H<sub>0</sub> yaitu parameter tersebut tidak signifikan terhadap model. Namun, meski terdapat parameter yang tidak signifikan, variabel tersebut masih digunakan karena minimal dalam satu variabel terdapat satu parameter yang signifikan. Sehingga variabel prediktor X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, memberikan pengaruh yang signifikan terhadap angka kasus Tuberculosis.

E. Pengujian Asumsi Residual

i. Pengujian Asumsi Residual Identik

Metode yang digunakan dalam pengujian asumsi residual identik adalah menggunakan uji Glejser. Uji Glejser dilakukan dengan meregresikan antara harga mutlak dari residual dengan variabel prediktor. Hipotesis yang digunakan dalam uji Glejser adalah sebagai berikut.

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

(residual identik)

$$H_1 : \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, 11$$

(residual tidak identik)

Tabel 7.  
Tabel ANOVA Pengujian Asumsi Residual Identik

Sumber Variasi	df	SS	MS	F <sub>hitung</sub>	P-Value
Regresi	11	1920,889	174,6263	0,5244	0,8632
Error	19	6326,843	332,9917		
Total	30	8247,732			

Dari Tabel 7 dapat diambil keputusan gagal tolak H<sub>0</sub> karena nilai nilai F<sub>hitung</sub> lebih kecil dari F<sub>tabel</sub> yaitu 0,5244 atau P-Value yaitu 0,8632 lebih besar dari nilai  $\alpha=0,05$ . Sehingga dapat disimpulkan bahwa tidak terjadi kasus heteroskedastisitas atau asumsi residual identik telah terpenuhi.

ii. Pengujian Asumsi Residual Independen

Perumusan hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \rho = 0 \quad (\text{tidak terjadi autokorelasi})$$

$$H_1 : \rho \neq 0 \quad (\text{terjadi autokorelasi})$$

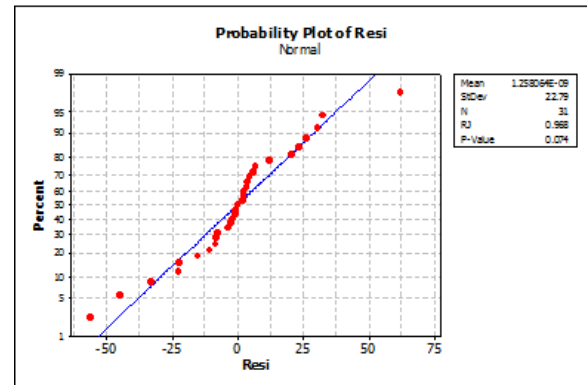
Dari hasil analisis uji Durbin-Watson yang diperoleh, nilai Durbin-Watson yang didapatkan adalah sebesar 1,726. Residual dapat dikatakan terdapat autokorelasi apabila nilai d (d = 1,726) lebih kecil dari nilai d<sub>L</sub> (d<sub>L</sub> = 1,229) atau nilai d lebih kecil dari nilai d<sub>U</sub> (d<sub>U</sub> = 1,650). Berdasarkan hasil yang diperoleh, nilai d lebih besar dari nilai d<sub>L</sub> dan nilai d juga lebih kecil dari nilai d<sub>U</sub>, sehingga keputusannya adalah gagal tolak H<sub>0</sub>. Kesimpulannya adalah telah terbukti bahwa residual independen terpenuhi dan tidak terdapat autokorelasi.

iii. Pengujian Asumsi Residual Berdistribusi Normal

Uji Shapiro-Wilk dilakukan untuk mengetahui apakah suatu data telah mengikuti suatu distribusi tertentu. Perumusan hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : F_n(\epsilon) = F_0(\epsilon) \quad (\text{residual berdistribusi Normal})$$

$$H_1 : F_n(\epsilon) \neq F_0(\epsilon) \quad (\text{residual tidak berdistribusi Normal})$$



Gambar 2. Probability Plot dengan Shapiro-Wilk.

Berdasarkan Gambar 2 di atas diperoleh nilai P-Value 0,074 dan RJ sebesar 0,968 yang memiliki nilai lebih besar dari  $\alpha=0,05$  sehingga keputusannya adalah gagal tolak H<sub>0</sub> yang artinya residual mengikuti distribusi normal atau asumsi residual berdistribusi normal terpenuhi.

F. Koefisien Determinasi (R<sup>2</sup>)

Model regresi nonparametrik Spline Truncated dengan kombinasi knot (3,2,3) menghasilkan nilai koefisien determinasi (R<sup>2</sup>) sebesar 74,3515% yang artinya bahwa variabel prediktor yang digunakan mampu menjelaskan model sebesar 74,3515%, dan sisanya sebesar 25,6485% dijelaskan oleh variabel lain yang tidak masuk ke dalam model.

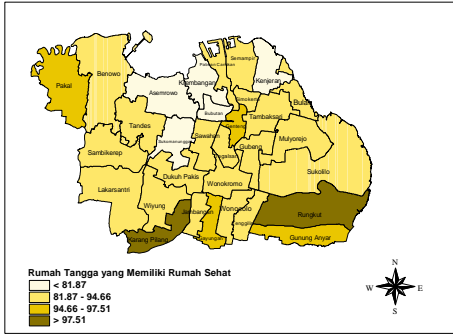
G. Interpretasi Model Regresi Nonparametrik Spline Truncated

Berikut merupakan model terbaik dengan kombinasi knot (3,2,3) yang telah didapatkan.

$$\hat{y} = 2308.13 + 2.50x_1 - 9.03(x_1 - 81.87)_+^1 - 14.78(x_1 - 94.66)_+^1 + 268.54(x_1 - 97.51)_+^1 - 1.37x_2 + 188.62(x_2 - 91.53)_+^1 - 406.28(x_2 - 94.24)_+^1 - 23.59x_3 + 34.05(x_3 - 96.12)_+^1 - 205.75(x_3 - 99.15)_+^1 + 737.12(x_3 - 99.83)_+^1$$

1. Dengan mengasumsikan variabel prediktor selain x<sub>1</sub> konstan, maka persamaan regresi dari variabel persentase rumah tangga yang memiliki rumah sehat dan variabel persentase rumah tangga yang memiliki rumah sehat termasuk ke dalam komponen nonparametrik, sehingga modelnya dapat ditulis sebagai berikut.

$$\hat{y} = \begin{cases} 2308,13 + 2,50x_1 & ; x_1 < 81,87 \\ 3047,4161 - 6,53x_1 & ; 81,87 \leq x_1 < 94,66 \\ 4446,4909 - 21,31x_1 & ; 94,66 \leq x_1 < 97,51 \\ -21738,8445 + 247,23x_1 & ; x_1 \geq 97,51 \end{cases}$$



Gambar 3. Peta Persebaran Rumah Tangga yang Memiliki Rumah Sehat.

Pada kondisi persentase rumah tangga yang memiliki rumah sehat yang kurang dari 81,87 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami kenaikan sebesar 2,50 per 100.000 kejadian.

Pada kondisi persentase rumah tangga yang memiliki rumah sehat di antara 81,87 dan 94,66 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami penurunan sebesar 6,53 per 100.000 kejadian.

Pada kondisi persentase rumah tangga yang memiliki rumah sehat di antara 94,66 dan 97,51 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami penurunan sebesar 21,31 per 100.000 kejadian.

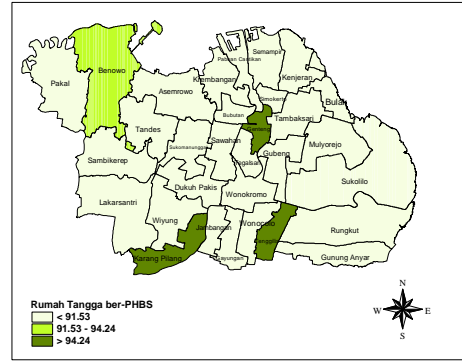
Pada kondisi persentase rumah tangga yang memiliki rumah sehat yang lebih besar dari 97,51 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami kenaikan sebesar 247,23 per 100.000 kejadian.

Tabel 8. Persebaran Kecamatan pada Variabel Persentase Rumah Tangga yang Memiliki Rumah Sehat

Interval	Kecamatan
< 81,87%	Asemrowo, Krembangan, Bubutan, Sukomanunggal, dan Kenjeran
81,87% – 94,66%	Benowo, Tandes, Sambikerep, Lakarsantri, Wiyung, Dukuh Pakis, Jambangan, Wonocolo, Tenggilis, Wonokromo, Sawahan, Tegalsari, Gubeng, Mulyorejo, Sukolilo, Tambaksari, Simokerto, Bulak, Pabean Cantikan, dan Semampir
94,66 – 97,51%	Pakal, Genteng, Gayungan, dan Gunung Anyar
≥ 97,51%	Karang Pilang dan Rungkut

2. Dengan mengasumsikan variabel prediktor selain  $x_2$  konstan, maka persamaan regresi dari variabel persentase rumah tangga ber-PHBS dan variabel persentase rumah tangga ber-PHBS termasuk ke dalam komponen nonparametrik, sehingga modelnya dapat ditulis sebagai berikut.

$$\hat{y} = \begin{cases} 2308,13 - 1,37x_2 & ; x_2 < 91,53 \\ -14956,2586 + 187,25x_2 & ; 91,53 \leq x_2 < 94,24 \\ 23331,5686 - 219,03x_2 & ; x_2 \geq 94,24 \end{cases}$$



Gambar 4. Peta Persebaran Rumah Tangga ber-PHBS.

Pada kondisi persentase rumah tangga yang ber-PHBS yang kurang dari 91,53 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami penurunan sebesar 1,37 per 100.000 kejadian.

Pada kondisi persentase rumah tangga yang ber-PHBS di antara 91,53 dan 94,24 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami kenaikan sebesar 187,25 per 100.000 kejadian.

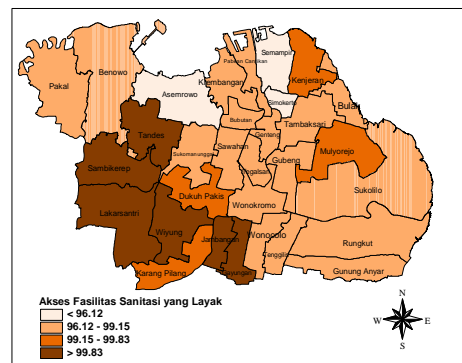
Pada kondisi persentase rumah tangga yang ber-PHBS yang lebih besar dari 94,24 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami penurunan sebesar 219,03 per 100.000 kejadian.

Tabel 9. Persebaran Kecamatan pada Variabel Persentase Rumah Tangga yang ber-PHBS

Interval	Kecamatan
< 91,53%	Asemrowo, Krembangan, Bubutan, Sukomanunggal, Kenjeran, Tandes, Sambikerep, Lakarsantri, Wiyung, Dukuh Pakis, Jambangan, Wonocolo, Wonokromo, Sawahan, Tegalsari, Gubeng, Mulyorejo, Sukolilo, Tambaksari, Simokerto, Bulak, Pabean Cantikan, Semampir, Pakal, Gayungan, Gunung Anyar, dan Rungkut
91,53% – 94,24%	Benowo
≥ 94,24%	Tenggilis, Genteng, Karang Pilang

3. Dengan mengasumsikan variabel prediktor selain  $x_3$  konstan, maka persamaan regresi dari variabel persentase akses fasilitas sanitasi layak dan variabel persentase akses fasilitas sanitasi layak termasuk ke dalam komponen nonparametrik, sehingga modelnya dapat ditulis sebagai berikut.

$$\hat{y} = \begin{cases} 2308,13 - 23,59x_3 & ; x_3 < 96,12 \\ -964,756 + 10,46x_3 & ; 96,12 \leq x_3 < 99,15 \\ 19435,3565 - 195,29x_3 & ; 99,15 \leq x_3 < 99,83 \\ -54151,3331 + 541,83x_3 & ; x_3 \geq 99,83 \end{cases}$$



Gambar 5. Peta Persebaran Akses Fasilitas Sanitasi yang Layak.

Pada kondisi persentase akses fasilitas sanitasi layak yang kurang dari 96,12 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami penurunan sebesar 23,59 per 100.000 kejadian. Pada kondisi persentase akses fasilitas sanitasi layak di antara 96,12 dan 99,15 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami kenaikan sebesar 10,46 per 100.000 kejadian.

Pada kondisi persentase akses fasilitas sanitasi layak di antara 99,15 dan 99,83 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami penurunan sebesar 195,29 per 100.000 kejadian. Pada kondisi persentase akses fasilitas sanitasi layak yang lebih besar dari 99,83 jika terjadi penambahan sebesar 1 satuan maka Angka Kasus *Tuberculosis* akan mengalami kenaikan sebesar 541,83 per 100.000 kejadian.

Tabel 10.

Persebaran Kecamatan pada Variabel Akses Sanitasi Layak

Interval	Kecamatan
< 96,12%	Asemrowo, Semampir, dan Simokerto Krempangan, Bubutan, Sukomanunggal, Benowo, Wonocolo, Tenggiling, Wonokromo, Sawahan, Tegalsari, Gubeng, Sukolilo,
96,12% – 99,15%	Tambaksari, Bulak, Pabean Cantikan, Pakal, Genteng, Gunung Anyar, dan Rungkut
99,15 – 99,83%	Kenjeran, Dukuh Pakis, Mulyorejo, dan Karang Pilang
≥ 99,83%	Tandes, Sambikerep, Lakarsantri, Wiyung, Jambangan, Gayungan

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, kesimpulan yang didapatkan adalah sebagai berikut.

1. Dari 31 Kecamatan yang ada di Kota Surabaya, angka kasus *Tuberculosis* tertinggi terdapat di Kecamatan Tandes dan yang terendah di Kecamatan Gayungan.
2. Model terbaik regresi nonparametrik *Spline Truncated* untuk angka kasus *Tuberculosis* di Kota Surabaya tahun

2017 merupakan model kombinasi knot (3,2,3) dengan nilai GCV minimum sebesar 1338,53. Sebanyak 3 variabel prediktor yang digunakan berpengaruh signifikan dengan ukuran kebaikan model sebesar 74,3515%.

### B. Saran

Saran yang dapat diberikan berdasarkan kesimpulan di atas adalah sebagai berikut.

1. Faktor-faktor yang berpengaruh signifikan diharapkan mampu menjadi pertimbangan pemerintah dalam mengambil kebijakan dalam pencegahan kasus *Tuberculosis* dan menurunkan angka kasus *Tuberculosis* di kecamatan di Kota Surabaya yang telah diteliti.
2. Untuk pemerintah dan dinas terkait diharapkan mampu melakukan sosialisasi mengenai penyakit *Tuberculosis* dan faktor-faktor yang dapat memengaruhi kasus *Tuberculosis* di Kota Surabaya.

## DAFTAR PUSTAKA

- [1] Aditama, T. (2002). Diagnosis dan Pengobatan *Tuberculosis* Terbaru. (<http://www.tbindonesia.or.id>).
- [2] Walpole, R. (1995). Pengantar Metode Statistika. Diterjemahkan oleh: Bambang Sumantri. Jakarta: PT. Gramedia Pustaka Utama.
- [3] Nasution, M. (2005). Manajemen Mutu Terpadu (Total Quality Management) Edisi Kedua. Bogor: Ghalia Indonesia.
- [4] Eubank, R. (1999). Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker Inc.
- [5] Hardle, W. (1994). Applied Nonparametric Regression. New York: Cambridge University Press.
- [6] Budiantara, I. (2006). Model Spline dengan Knots Optimal. Jurnal Ilmu Dasar, FMIPA Universitas Jember, Vol 7, Hal 77-85.
- [7] Drapper, N. R., & Smith, H. (1992). Analisis Regresi Terapan. Diterjemahkan oleh: Bambang Sumantri. Jakarta: Gramedia Pustaka Utama.
- [8] Gujarati, D., & Porter, D. (2009). Basic Econometrics, 5th Edition. New York: The McGraw-Hill Companies, Inc.
- [9] N.M. Razali, Y. B. (2011). Power Comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson Darling tests. Journal of Statistical modeling and Analytics Volume 2 No. 1, 21-33.
- [10] Dinas Kesehatan Pemerintah Kota Surabaya. (2018). Profil Kesehatan Kota Surabaya Tahun 2017. Surabaya: Dinas Kesehatan Pemerintah Kota Surabaya.