

# Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Tuberkulosis di Jawa Timur Menggunakan Regresi Nonparametrik *Spline*

Firda Fahrur Nisa' dan I Nyoman Budiantara

Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: i\_nyoman\_b@statistika.its.ac.id dan firdafahrur@gmail.com

**Abstrak**—Tuberkulosis adalah penyakit menular langsung yang disebabkan oleh kuman TB (*Mycobacterium tuberculosis*). Sebagian besar kuman TB menyerang paru, tetapi dapat juga mengenai organ tubuh lainnya. 7,5% angka kematian di Indonesia disebabkan oleh penyakit TB. Provinsi Jawa Timur memiliki kasus TB terbanyak kedua setelah provinsi Jawa Barat. Data Dinas Kesehatan Provinsi Jawa Timur tahun 2014 menunjukkan kasus TB mencapai 42.458 kasus. Tindakan preventif yang dilakukan oleh pemerintah dengan mengendalikan faktor-faktor yang mempengaruhi penyakit TB. Pada penelitian ini, metode regresi nonparametrik spline digunakan untuk memodelkan jumlah kasus TB di Jawa Timur. Metode regresi nonparametrik *Spline* digunakan karena pola data antara jumlah kasus TB di Jawa Timur tidak menunjukkan kecenderungan pola tertentu dengan faktor-faktor yang diduga mempengaruhinya, yaitu persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat, persentase penduduk miskin, persentase gizi buruk, persentase tenaga kesehatan terlatih dan jumlah rumah tangga ber-PHBS. Berdasarkan hasil pemodelan didapatkan model spline terbaik yaitu spline dengan kombinasi knot (1,3,2,3,2). Nilai koefisien determinasi yang didapatkan adalah 83,42% dengan tiga variabel prediktor dalam model berpengaruh signifikan, yaitu persentase gizi buruk masyarakat, persentase tenaga kesehatan terlatih TB, dan persentase rumah tangga ber-PHBS.

**Kata Kunci**—Kasus Tuberkulosis, GCV, Titik Knot, Nonparametrik *Spline*.

## I. PENDAHULUAN

Tuberkulosis (TB) sampai saat ini masih menjadi isu kesehatan global di semua negara. Sebanyak 8,9 juta penderita tuberkulosis dengan proporsi 80% pada 22 negara berkembang dengan kematian 3 juta orang per tahun, dan satu orang dapat terinfeksi tuberkulosis setiap detik. Indonesia berada pada peringkat kelima negara yang berkategori negara beban tinggi terhadap TB [1]. TB merupakan salah satu penyakit menular dan langsung mengenai parenkim paru yang disebabkan oleh basil *Mycobacterium tuberculosis*.

Di Indonesia, provinsi Jawa Timur memiliki kasus TB terbanyak kedua setelah provinsi Jawa Barat. Data Dinas Kesehatan Provinsi Jawa Timur tahun 2013 menunjukkan kasus tuberkulosis mencapai 41.472 kasus dengan Basil Tahan Asam (BTA) positif baru sebanyak 25.618 kasus dan 1.233 penderita TB meninggal. Sedangkan pada tahun 2014 jumlah kasus tuberkulosis mencapai 42.458 kasus dengan BTA positif baru sebanyak 22.866 kasus. Salah satu faktor penyebab berkembangnya penyakit ini adalah kurangnya ventilasi

dan pencahayaan matahari pada rumah penduduk, serta kurangnya istirahat.

Penelitian sebelumnya mengenai TB pernah dilakukan Hasyim [2] tentang pemodelan angka kejadian penyakit infeksi tuberkulosis paru di Kabupaten Sorong, Papua Barat dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS). Hasil penelitian menunjukkan bahwa jenis pekerjaan, umur, kebiasaan merokok, status sosial ekonomi, konsumsi alkohol dan tingkat pendidikan berpengaruh signifikan. Penelitian mengenai faktor-faktor yang mempengaruhi kasus TB juga dilakukan oleh Faris Muaz [3] dengan studi kasus puskesmas wilayah Kecamatan Serang Kota Serang. Sedangkan berdasarkan penelitian Nurmala dalam Puspita [4], faktor-faktor TB pada anak adalah umur, status gizi, tingkat pendidikan orang tua yang rendah, tingkat pengetahuan orang tua yang rendah tentang penyakit TB dan kepadatan hunian. Penelitian Amrullah dalam Puspita [4] faktor-faktor resiko penderita TB yaitu pada kelompok produktif (15-50 tahun), jenis kelamin laki-laki, status gizi yang buruk, diabetes melitus, dan kondisi rumah. Penelitian terkait TB selanjutnya dilakukan oleh Lestari [5] yang meneliti faktor-faktor yang mempengaruhi jumlah kasus TB dengan menggunakan metode *Generalized Poisson Regression* (GPR) dan *Geographically Weighted Poisson Regression* (GWPR). Hasil penelitian tersebut menunjukkan bahwa variabel persentase penduduk usia produktif, persentase tenaga kesehatan terdidik, dan persentase TUPM sehat yang berpengaruh signifikan terhadap jumlah kasus TB di Jawa Timur.

*Spline* adalah suatu *piecewise polinomial*, yaitu polinomial yang memiliki sifat tersegmen [6]. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan pola perilaku data pada interval yang berbeda [7].

Berdasarkan uraian sebelumnya, perlu dilakukan analisis untuk mengetahui faktor-faktor yang mempengaruhi jumlah kasus TB di Jawa Timur. Berdasarkan pola data variabel prediktor terhadap variabel respon yang tidak mengikuti pola tertentu, maka akan dilakukan pemodelan menggunakan regresi nonparametrik spline. Keunggulan metode ini adalah dalam mengatasi pola data yang menunjukkan naik/turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus [8].

## II. TINJAUAN PUSTAKA

### A. Regresi Nonparametrik Spline

Regresi nonparametrik merupakan metode statistika yang digunakan apabila pola data yang tidak diketahui bentuk kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data [9]. Regresi nonparametrik spline merupakan regresi yang sangat fleksibel dalam memodelkan pola data. Model regresi nonparametrik spline secara umum adalah sebagai berikut.

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

dimana  $y$  adalah variabel respon dan  $f(x_i)$  adalah fungsi spline berorde  $m$  dengan titik knot  $k_1, k_2, \dots, k_r$  yang diberikan oleh persamaan

$$f(x_i) = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - k_k)_+^m \quad (2)$$

Apabila Persamaan 2 disubstitusikan ke dalam Persamaan 1 maka akan diperoleh model regresi nonparametrik spline sebagai berikut

$$y_i = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - k_k)_+^m + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3)$$

Fungsi *truncated* adalah sebagai berikut.

$$(x_i - k_k)_+^m = \begin{cases} (x_i - k_k)^m, & x_i \geq k_k \\ 0, & x_i < k_k \end{cases} \quad (4)$$

Persamaan 1 dapat ditulis dalam bentuk matriks sebagai berikut.

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

Estimasi parameter model regresi nonparametrik spline dilakukan dengan menggunakan metode *Ordinary Least Square* (OLS), yaitu dengan meminimumkan jumlah kuadrat error yang diberikan oleh persamaan berikut.

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

### B. Pemilihan Titik Knot Optimal

Model regresi spline terbaik diperoleh dari titik knot yang optimal. Metode yang sering digunakan untuk menentukan titik knot optimal adalah *Generalized Cross Validation* (GCV). Titik knot yang optimal diperoleh dari nilai GCV minimum. Fungsi GCV adalah sebagai berikut.

$$GCV(k_1, k_2, \dots, k_r) = \frac{MSE(k_1, k_2, \dots, k_r)}{(n^{-1} \text{tr}[I - A(k_1, k_2, \dots, k_r)])^2} \quad (7)$$

dengan  $I$  adalah matriks identitas dan  $n$  adalah jumlah pengamatan. Nilai dari  $MSE(k_1, k_2, \dots, k_r)$  adalah sebagai berikut

$$MSE(k_1, k_2, \dots, k_r) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

### C. Pengujian Estimasi Parameter

#### 1) Pengujian Parameter Model Secara Serentak

Uji serentak dilakukan untuk mengetahui signifikansi parameter model regresi secara bersama-sama. Berikut ini merupakan hipotesis untuk uji serentak.

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

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

dengan  $m + r$  adalah jumlah parameter dalam model regresi spline.

Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hitung} = \frac{MS_{regresi}}{MS_{residual}} \quad (9)$$

Kesimpulan Tolak  $H_0$  jika  $F_{hitung} > F_{\alpha; ((m+r), n-(m+r)-1)}$  atau  $p\text{-value} < \alpha$  yang menunjukkan bahwa paling sedikit terdapat satu parameter yang signifikan terhadap model.

#### 2) Pengujian Parameter Model Secara Individu

Pengujian parameter secara individu dilakukan apabila pengujian parameter model secara serentak diperoleh kesimpulan bahwa minimal terdapat satu parameter yang signifikan. Berikut ini merupakan hipotesis dari pengujian parameter secara individu.

$$H_0 : \beta_j = 0$$

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

Statistik uji disajikan sebagai berikut.

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (10)$$

dimana  $\hat{\beta}_j$  adalah penaksir untuk parameter  $\beta_j$  dan  $SE(\hat{\beta}_j)$  adalah akar-akar elemen diagonal ke- $j$  dari matriks  $\sigma^2(X'X)^{-1}$ .  $H_0$  ditolak jika  $|t_{hitung}| > t_{\frac{\alpha}{2}; (n-(m+r)-1)}$ .

Begitu juga dengan  $p\text{-value} < \alpha$  yang menunjukkan bahwa parameter signifikan terhadap model.  $R^2$  digunakan sebagai indikator kebaikan model, yang diberikan sebagai berikut.

$$R^2 = \frac{SS_{regresi}}{SS_{total}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad (11)$$

### D. Pengujian Asumsi Residual

#### 1) Asumsi Identik

Asumsi identik merupakan salah satu asumsi yang dipenuhi yaitu variansi residual adalah homogen atau tidak terjadi heteroskedastisitas. Kasus heteroskedastisitas dapat dideteksi menggunakan uji *glejser* dan dilakukan dengan cara meregresikan harga mutlak residual dengan variabel prediktor. Berikut ini merupakan hipotesis uji *glejser*.

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

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

Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hitung} = \frac{\left[ \sum_{i=1}^n (|\hat{e}_i| - |\bar{e}|)^2 \right] / ((m+r)-1)}{\left[ \sum_{i=1}^n (|e_i| - |\hat{e}_i|)^2 \right] / (n-(m+r))} \quad (12)$$

dimana  $n$  adalah banyak observasi dan  $s$  adalah banyak parameter model *Glejser* dengan daerah penolakan yaitu  $H_0$  ditolak jika  $F_{hitung} \geq F_{tabel}(F_{\alpha; ((m+r)-1, n-(m+r))})$ .

#### 2) Asumsi Independen

Asumsi independen digunakan untuk mengetahui adanya autokorelasi atau tidak. *Autocorrelation Function* atau plot ACF dapat digunakan untuk mengetahui ada tidaknya autokorelasi dalam residual. Apabila ada nilai autokorelasi yang keluar dari batas signifikansi maka dapat dikatakan terdapat kasus autokorelasi. Sebaliknya jika tidak terdapat nilai autokorelasi yang keluar batas signifikansi maka tidak terdapat kasus autokorelasi. Interval konfidensi  $(1-\alpha)$  100% untuk autokorelasi  $\rho_s$  diberikan oleh

$$-t_{n-1; \frac{\alpha}{2}} SE(\hat{\rho}_s) < \rho_s < t_{n-1; \frac{\alpha}{2}} SE(\hat{\rho}_s) \quad (13)$$

#### 3) Asumsi Distribusi Normal

Asumsi residual berdistribusi normal dilakukan untuk mengetahui apakah residual hasil regresi telah berdistribusi normal atau tidak. Cara lain dapat dilakukan untuk menguji asumsi residual berdistribusi normal

adalah berdasarkan visual menggunakan *normal probability plot residual* yang cenderung mengikuti garis atau menggunakan uji Kolmogorov-Smirnov. Apabila menggunakan uji Kolmogorov-Smirnov, berikut merupakan hipotesis yang digunakan.

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

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

Statistik uji yang digunakan adalah

$$D = \text{Sup} |S(x) - F_0(x)| \tag{14}$$

Daerah penolakan  $H_0$  jika  $|D| > q_{1-\alpha}$  dimana nilai  $q_{1-\alpha}$  adalah berdasarkan tabel *Kolmogorov-Smirnov* [10].

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Dinas Kesehatan Provinsi Jawa Timur tahun 2014 dan Badan Pusat Statistik Provinsi Jawa Timur tahun 2014. Data diambil pada 38 kabupaten/kota yang ada di Provinsi Jawa Timur.

#### B. Variabel Penelitian

Variabel respon yang digunakan dalam penelitian ini adalah jumlah kasus penyakit tuberkulosis (Y) di Provinsi Jawa Timur dengan 5 variabel yang prediktor yang diduga mempengaruhi jumlah kasus penyakit tuberkulosis di Jawa Timur. Variabel prediktor yang digunakan menggunakan satuan persen, sedangkan untuk variabel respon menggunakan satuan jumlah. Variabel-variabel yang digunakan dalam penelitian ini tercantum Tabel 1.

TABEL 1. VARIABEL PENELITIAN

Variabel	Nama Variabel
Y	Jumlah Kasus Baru Penyakit Tuberkulosis
X <sub>1</sub>	Persentase Keluarga dengan Kepemilikan Sarana Sanitasi Dasar Sehat
X <sub>2</sub>	Persentase Penduduk Miskin
X <sub>3</sub>	Persentase Gizi Buruk Masyarakat
X <sub>4</sub>	Persentase Tenaga Kesehatan Terlatih
X <sub>5</sub>	Persentase Rumah Tangga ber-PHBS

#### C. Langkah-Langkah Penelitian

Langkah-langkah analisis yang akan dilakukan dalam penelitian ini adalah sebagai berikut.

- Melakukan analisis statistika deskriptif untuk mengetahui karakteristik data meliputi rata-rata, deviasi standar, nilai maksimum dan nilai minimum pada variabel jumlah kasus TB di Jawa Timur dan faktor yang diduga mempengaruhinya.
- Membuat *scatterplot* antara jumlah kasus TB di Jawa Timur dengan masing-masing variabel prediktor yang diduga mempengaruhi untuk mengetahui pola hubungan data antara variabel respon dengan masing-masing variabel prediktornya.
- Melakukan pemodelan jumlah kasus TB dengan regresi nonparametrik menggunakan pendekatan *spline* dengan satu, dua, tiga, dan kombinasi knot.
- Memilih titik knot optimal dengan menggunakan karakteristik nilai GCV paling minimum.
- Melakukan pemodelan dengan menggunakan titik knot optimal.

- Menguji signifikansi parameter dari model yang terbentuk.
- Menguji asumsi residual dari model yang terbentuk.

### IV. ANALISIS DAN PEMBAHASAN

#### A. Karakteristik Jumlah Kasus Tuberkulosis di Jawa Timur

Terdapat beberapa faktor yang diduga mempengaruhi tingginya jumlah kasus tuberkulosis di Jawa Timur. Pada penelitian kali ini, terdapat lima faktor yang diduga mempengaruhi kasus tuberkulosis di Jawa Timur yaitu jumlah kasus penyakit tuberkulosis di Jawa Timur, variabel persentase kegiatan sarana sanitasi dasar keluarga, persentase gizi buruk masyarakat, persentase tenaga kesehatan terlatih TB, dan persentase rumah tangga ber-PHBS. Karakteristik kelima faktor tersebut disajikan dalam Tabel 2 sebagai berikut.

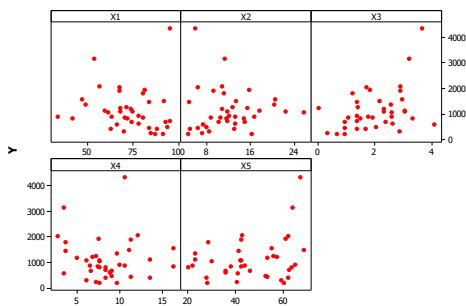
TABEL 2. STATISTIKA DESKRIPTIF FAKTOR-FAKTOR YANG DIDUGA MEMPENGARUHI PENGANGGURAN DI INDONESIA

Variabel	Mean	Varians	Minimum	Maksimum
Y	1.117	682.921	176	4.338
X <sub>1</sub>	72,81	235,45	33,70	95,55
X <sub>2</sub>	12,096	24,926	4,590	25,800
X <sub>3</sub>	2,019	0,915	0,000	4,097
X <sub>4</sub>	8,473	10,819	2,721	16,344
X <sub>5</sub>	46,29	209,51	20,06	68,67

Berdasarkan Tabel 2, diketahui bahwa persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat (X<sub>1</sub>) tertinggi terdapat di Kabupaten Magetan yaitu sebesar 95,55% sedangkan persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat terendah adalah di Kabupaten Situbondo yaitu sebesar 33,7%. Provinsi dengan persentase penduduk miskin (X<sub>2</sub>) tertinggi sebesar 25,8% yaitu pada Kabupaten Sampang dan terendah sebesar 4,59% yaitu pada Kota Batu. Persentase gizi buruk masyarakat (X<sub>3</sub>) paling tinggi adalah Kota Pasuruan dengan sekitar 4,10% namun berbeda dengan Kabupaten Bojonegoro yang dimana persentase gizi buruk masyarakatnya adalah 0,00%, artinya status gizi masyarakat di Kabupaten Bojonegoro terbilang baik. Untuk persentase tenaga kesehatan terlatih TB (X<sub>4</sub>), persentase tertinggi adalah di Kabupaten Sumenep yaitu sebesar 16,34% dan terendah adalah Kabupaten Sidoarjo sebesar 2,72%. Sedangkan pada persentase rumah tangga ber-PHBS (X<sub>5</sub>), kabupaten dengan persentase rumah tangga ber-PHBS terbanyak adalah Kabupaten Gresik dan kabupaten dengan persentase rumah tangga ber-PHBS terendah adalah Kabupaten Bondowoso.

#### B. Pola Data Antara Jumlah Jumlah Kasus Tuberkulosis dengan Variabel yang Diduga Mempengaruhi

Gambar 2 adalah *scatterplot* yang menunjukkan pola hubungan antara jumlah kasus tuberkulosis (Y) dengan lima variabel yang diduga mempengaruhinya. Berdasarkan hasil *scatterplot* dapat diketahui bahwa kelima pola data tidak ada yang membentuk suatu pola tertentu. Dengan demikian, dalam pemodelan regresi akan digunakan pendekatan regresi nonparametrik. Berikut *scatterplot* dari data.



Gambar 1 Scatterplot variabel prediktor terhadap respon

C. Pemilihan Titik Knot Optimal

Nilai GCV minimum pada pemilihan titik knot optimal dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot ditampilkan sebagai berikut.

TABEL 3. NILAI GCV MINIMUM SETIAP PEMILIHAN TITIK KNOT OPTIMAL

Banyak Titik Knot Optimal	Nilai GCV Minimum
Satu Titik Knot	638147,0798
Dua Titik Knot	489149,0305
Tiga Titik Knot	371212,9481
Kombinasi Knot (1,3,2,3,2)	276418,0578 (*)

Kombinasi knot terbaik adalah kombinasi (1,3,2,3,2) dimana nilai GCV yang dihasilkan adalah paling minimum dibandingkan lainnya. Sehingga, model regresi nonparametrik spline terbaik adalah ketika menggunakan kombinasi knot 1,3,2,3,2. Berdasarkan Tabel 3 dapat diketahui bahwa kombinasi titik knot 3,3,1,2,3 memiliki nilai GCV minimum terkecil yaitu sebesar 276418,0578.

D. Penaksiran Parameter Model Regresi Nonparametrik Spline

Berdasarkan nilai GCV terkecil, didapatkan model terbaik yaitu model regresi nonparametrik spline dengan kombinasi knot 1,3,2,3,2. Berikut ini adalah hasil estimasi parameter model regresi nonparametrik spline terbaik.

$$\hat{y} = 2184,86 - 2,564X_1 - 75,863(X_1 - 89,238)_+^1 + 141,526X_2 + 128,430(X_2 - 9,351)_+^1 - 21,847(X_2 - 24,934)_+^1 - 10,923(X_2 - 25,367)_+^1 + 176,203X_3 - 6129,817(X_3 - 3,848)_+^1 - 2043,264(X_3 - 4,0163) - 595,473X_4 + 649,170(X_4 - 5,7792)_+^1 - 368,3163(X_4 - 15,7879)_+^1 - 184,158(X_4 - 16,0659) + 12,7073X_5 + 2647,442(X_5 - 65,6938)_+^1 - 7652,740(X_5 - 67,6779)_+^1.$$

E. Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline

1) Uji Serentak

Pengujian secara serentak ini dilakukan untuk menguji estimasi parameter model secara bersamaan (simultan). Nilai  $\alpha$  yang digunakan sebesar 0,05. Berikut ini adalah hasil analisis ragam model regresi nonparametrik spline.

TABEL 4. ANALISIS RAGAM UJI SERENTAK

Sumber Variasi	Derajat Bebas (df)	Jumlah Kuadrat (SS)	Rataan Kuadrat (MS)	F hitung	p-value
Regresi	16	21078174	1317386		
Error	21	4189916	199519,8	6,602783	5,09448x10 <sup>-5</sup>
Total	37	25268090			

Berdasarkan Tabel 4 diketahui bahwa nilai p-value sebesar 5,09448x10<sup>-5</sup>. Nilai ini adalah lebih kecil daripada  $\alpha$  (0,05). Disamping itu, nilai  $F_{hitung} >$

$F_{(0,05;16;21)}$  yaitu 6,602783 > 2,1562 yang menunjukkan bahwa  $H_0$  ditolak, artinya minimal terdapat satu parameter yang signifikan dalam model. Untuk mengetahui parameter manakah yang signifikan terhadap model, dilakukan uji individu.

2) Uji Individu

Pengujian secara individu dilakukan untuk mengetahui variabel yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Timur. Berikut ini adalah hasil pengujian signifikansi parameter model secara individu.

TABEL 5. HASIL UJI INDIVIDU

Variabel	Parameter	Estimasi Parameter	p-value	t-hitung	Keputusan
Constant	$\beta_0$	2184,8640	0,05915988	1,99518	Tidak signifikan
X <sub>1</sub>	$\beta_1$	-2,564361	0,7793626	-0,28377	Tidak signifikan
	$\beta_2$	-75,863216	0,2688836	-1,13568	Tidak signifikan
	$\beta_3$	141,5261	0,05166073	2,06333	Tidak signifikan
X <sub>2</sub>	$\beta_4$	-128,430	0,1209485	-1,61632	Tidak signifikan
	$\beta_5$	-21,847	0,9664244	-0,04259	Tidak signifikan
	$\gamma_6$	-10,923	0,9664244	-0,04259	Tidak signifikan
X <sub>3</sub>	$\beta_7$	176,2039	0,1426633	1,52305	Signifikan
	$\beta_8$	-6129,817	0,005388	-3,10278	Signifikan
	$\beta_9$	-2043,2642	0,0053883	-3,10278	Signifikan
X <sub>4</sub>	$\beta_{10}$	-595,473536	9,020 x10 <sup>-5</sup>	-4,82729	Signifikan
	$\beta_{11}$	649,170892	0,0003451	4,26453	Signifikan
	$\beta_{12}$	-368,316391	0,6015729	-0,53014	Tidak Signifikan
	$\beta_{13}$	-184,158195	0,6015729	-0,53014	Tidak Signifikan
X <sub>5</sub>	$\beta_{14}$	12,707361	0,06177	1,97324	Tidak Signifikan
	$\beta_{15}$	2647,442589	1,009 x10 <sup>-5</sup>	5,76871	Signifikan
	$\beta_{16}$	-7652,740754	1,498 x10 <sup>-5</sup>	-5,59322	Signifikan

Berdasarkan Tabel 5 diketahui terdapat tiga variabel yang berpengaruh secara signifikan terhadap model. Tiga variabel tersebut yaitu persentase gizi buruk masyarakat, persentase tenaga kesehatan terlatih, dan persentase rumah tangga ber-PHBS.

F. Pengujian Asumsi Residual

1) Asumsi Identik

Residual dari model regresi nonparametrik spline harus memenuhi asumsi identik yang artinya tidak terjadi kasus heteroskedastisitas pada residual. Berikut ini adalah hasil analisis ragam uji Glejser.

TABEL 6. ANALISIS RAGAM UJI GLEJSER

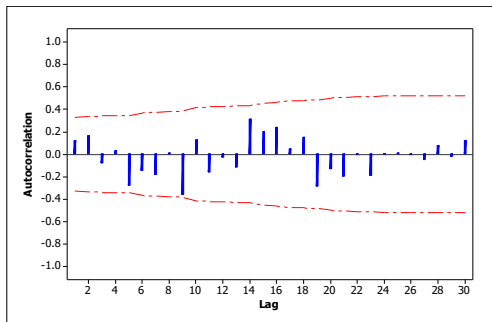
Sumber Variasi	Derajat Bebas (df)	Jumlah Kuadrat (SS)	Rataan Kuadrat (MS)	F hitung	p-value
Regresi	16	827030,4	51689,4		
Error	21	1104173	52579,64	0,9830687	0,5058166
Total	37	1931203			

Berdasarkan Tabel 6 diketahui bahwa nilai p-value adalah sebesar 0,5058166 yang lebih besar daripada  $\alpha$  yaitu 0,05 serta nilai  $F_{hitung} < F_{(0,05;16;21)}$  yaitu 0,9830687 < 2,1562 sehingga gagal tolak  $H_0$ , artinya variansi residual homogen atau tidak terjadi heterkodastisitas dalam model. Baik pengujian secara visual menggunakan scatterplot ataupun berdasarkan uji Glejser, diperoleh hasil yang sama yaitu residual model regresi nonparametrik spline memenuhi asumsi identik.

2) Asumsi Independen

Residual dari model regresi nonparametrik spline yang didapatkan harus memenuhi asumsi independen, artinya

tidak terdapat autokorelasi pada residual. Cara mendeteksi adanya autokorelasi pada residual adalah dengan menggunakan plot *Autocorrelation Function* (ACF). Ber-dasarkan Gambar 2 tidak terlihat adanya autokorelasi yang keluar batas toleransi.



Gambar 2. Plot ACF Residual

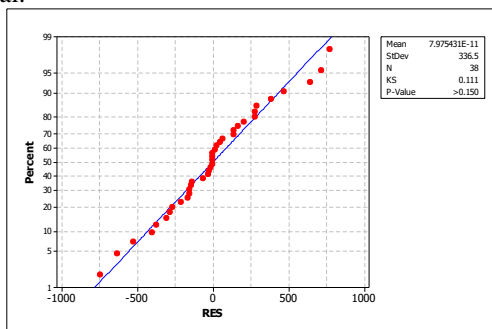
3) *Asumsi Distribusi Normal*

Untuk melakukan pengujian asumsi residual berdistribusi normal digunakan hipotesis sebagai berikut.

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

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

Pengujian asumsi residual yang ketiga adalah residual harus memenuhi asumsi distribusi normal. Pengujian ini dilakukan menggunakan uji Kolmogorov-Smirnov. Berdasarkan Gambar 3 diketahui bahwa nilai *Kolmogorov-Smirnov* sebesar 0,111 dimana nilai ini lebih kecil dibandingkan  $q_{(1-\alpha)}$  yaitu 0,215 dan *p-value* adalah  $> 0,150$  sehingga gagal tolak  $H_0$ . Hal ini menunjukkan bahwa residual model regresi nonparametrik spline telah memenuhi asumsi distribusi normal.



Gambar 3. Hasil Uji Kolmogorov-Smirnov

G. *Nilai Koefisien Determinasi (R<sup>2</sup>)*

Berdasarkan perhitungan didapatkan  $R^2$  sebesar 83,42%. Hal ini berarti model tersebut mampu menjelaskan keragaman jumlah kasus tuberkulosis di Jawa Timur sebesar 83,42%, sedangkan sisanya dijelaskan oleh variabel lain.

H. *Interpretasi Model Regresi Nonparametrik Spline*

Setelah dilakukan pengujian asumsi pada residual model regresi nonparametrik spline dan didapatkan bahwa semua asumsi terpenuhi, selanjutnya dilakukan interpretasi model. Interpretasi ini dibuat berdasarkan model untuk setiap variabel untuk memudahkan dalam interpretasi.

Model regresi nonparametrik spline terbaik adalah dengan titik knot kombinasi sebagai berikut.

$$\hat{y} = 2184,86 - 2,564X_1 - 75,863(X_1 - 89,238)_+^1 + 141,526X_2 + - 128,430(X_2 - 9,351)_+^1 - 21,847(X_2 - 24,934)_+^1 + - 10,923(X_2 - 25,367)_+^1 + 176,203X_3 - 6129,817(X_3 - 3,848)_+^1$$

$$- 2043,264(X_3 - 4,0163) - 595,473X_4 + + 649,170(X_4 - 5,7792)_+^1 - 368,3163(X_4 - 15,7879)_+^1 + - 184,158(X_4 - 16,0659) + 12,7073X_5 + + 2647,442(X_5 - 65,6938)_+^1 - 7652,740(X_5 - 67,6779)_+^1.$$

Model tersebut memiliki lima variabel yaitu persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat, persentase penduduk miskin, persentase gizi buruk masyarakat, persentase tenaga kesehatan terlatih, dan persentase rumah tangga ber-PHBS. Interpretasi model bertujuan untuk mengetahui seberapa besar pengaruh masing-masing variabel terhadap jumlah kasus tuberkulosis di Jawa Timur.

Dari model tersebut dapat diinterpretasikan sebagai berikut.

1. Apabila variabel  $X_2$ ,  $X_3$ ,  $X_4$  dan  $X_5$  diasumsikan konstan, maka pengaruh persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat ( $X_1$ ) terhadap jumlah kasus tuberkulosis di Jawa Timur adalah

$$\hat{y} = 2184,86 - 2,564X_1 - 75,863(X_1 - 89,238)_+^1 = \begin{cases} 2184,86 - 2,564X_1, & X_1 < 89,238 \\ 8954,722 - 754,427 X_1, & X_1 \geq 89,238 \end{cases}$$

Berdasarkan model di atas, dapat diinterpretasikan bahwa apabila persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat kurang dari 89,238 persen dan memiliki kenaikan sebesar 1 persen, maka kasus tuberkulosis akan turun sebesar 2,564 kasus. Dan apabila persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat lebih dari 89,238 persen, maka apabila persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat naik sebesar 1 persen, maka kasus tuberkulosis akan turun sebesar 754,427.

2. Apabila variabel  $X_1$ ,  $X_3$ ,  $X_4$  dan  $X_5$  dianggap konstan, maka pengaruh persentase penduduk miskin ( $X_2$ ) terhadap jumlah kasus tuberkulosis di Jawa Timur adalah

$$\hat{y} = 2184,86 + 141,526X_2 - 128,430(X_2 - 9,351)_+^1 - 21,847(X_2 - 24,934)_+^1 - 10,923(X_2 - 25,367)_+^1 = \begin{cases} 2184,86 + 141,526X_2, & X_2 < 9,351 \\ 3385,8089 + 13,096 X_2, & 9,351 \leq X_2 < 24,934 \\ 3930,5410 - 8,751 X_2, & 24,934 \leq X_2 < 25,367 \\ 4207,6247 - 19,674 X_2, & X_2 \geq 25,367 \end{cases}$$

Pada model tersebut memiliki interpretasi bahwa apabila persentase penduduk miskin kurang dari 9,351 persen, maka apabila persentase penduduk miskin naik sebesar 1 persen, maka kasus tuberkulosis akan naik sebesar 141,526 kasus. Apabila persentase penduduk miskin berkisar antara 9,351 dan 24,934 persen, maka apabila persentase penduduk miskin naik sebesar 1 persen, maka kasus tuberkulosis akan naik sebesar 13,096. Apabila persentase penduduk miskin adalah lebih besar dari 25,367 persen, maka kasus tuberkulosis akan turun sebesar 19,674.

3. Apabila variabel  $X_1$ ,  $X_2$ ,  $X_4$  dan  $X_5$  dianggap konstan, maka pengaruh persentase gizi buruk masyarakat ( $X_3$ ) terhadap jumlah kasus tuberkulosis di Jawa Timur adalah.

$$\hat{y} = 2184,86 + 176,203X_3 - 6129,817(X_3 - 3,848)_+^1 - 2043,264(X_3 - 4,0163)_+^1 = \begin{cases} 2184,86 + 176,203 X_3, & X_3 < 3,848 \\ 25772,395 - 5953,614 X_3, & 3,848 \leq X_3 < 4,0163 \\ 33978,7562 - 7996,878 X_3, & X_3 \geq 4,0163 \end{cases}$$



Pada model tersebut memiliki interpretasi bahwa jika persentase gizi buruk masyarakat kurang dari 3,848, maka apabila persentase gizi buruk masyarakat naik sebesar 1 persen, maka kasus tuberkulosis akan naik sebesar 176,203. Selanjutnya apabila persentase gizi buruk masyarakat lebih dari 3,848, maka apabila persentase gizi buruk masyarakat naik sebesar 1 persen, maka kasus tuberkulosis akan turun sebesar 7996,878.

4. Apabila variabel  $X_1$ ,  $X_2$ ,  $X_3$  dan  $X_5$  dianggap konstan, maka pengaruh persentase tenaga kesehatan terlatih TB ( $X_4$ ) terhadap jumlah kasus tuberkulosis di Jawa Timur adalah

$$\hat{y} = 2184,86 + 595,473X_4 + 649,170(X_4 - 5,7792)^{\frac{1}{4}} - 368,3163(X_4 - 15,7879)^{\frac{1}{4}} - 184,158(X_4 - 16,0659)^{\frac{1}{4}}$$

$$= \begin{cases} 2184,86 + 595,473X_4, & X_4 < 5,7792 \\ -1566,8232 + 1244,643X_4, & 5,7792 \leq X_4 < 15,7879 \\ 4248,1179 + 876,3267 X_4, & 15,7879 \leq X_4 < 16,0659 \\ 7206,7810 + 692,1687 X_4, & X_4 \geq 16,0659 \end{cases}$$

Pada model tersebut memiliki interpretasi bahwa jika persentase tenaga kesehatan terlatih TB dalam suatu provinsi kurang dari 5,7792 persen, maka apabila persentase tenaga kesehatan terlatih TB naik sebesar 1 persen, maka kasus tuberkulosis akan naik sebesar 595,473. Kemudian, apabila persentase tenaga kesehatan terlatih TB berkisar antara 5,7792 dan 15,7879 persen, maka apabila persentase tenaga kesehatan terlatih TB naik sebesar 1 persen, maka akan menaikkan kasus tuberkulosis sebesar 1244,643. Apabila persentase tenaga kesehatan terlatih TB adalah lebih dari 16,0659, maka apabila persentase tenaga kesehatan terlatih TB naik sebesar 1 persen, akan meningkatkan kasus tuberkulosis sebanyak 692,1687.

5. Apabila variabel  $X_1$ ,  $X_2$ ,  $X_3$  dan  $X_4$  dianggap konstan, maka pengaruh persentase rumah tangga ber-PHBS ( $X_5$ ) terhadap jumlah kasus tuberkulosis di Jawa Timur adalah

$$\hat{y} = 2184,86 + 12,7073X_5 + 2647,442(X_5 - 65,6938)^{\frac{1}{4}} - 7652,740(X_5 - 67,6779)^{\frac{1}{4}}$$

$$= \begin{cases} 2184,86 + 12,7073X_5, & X_5 < 65,6938 \\ -171735,66 + 2660,1493 X_5, & 65,6938 \leq X_5 < 67,6779 \\ 346185,712 - 4992,590 X_5, & X_5 \geq 67,6779 \end{cases}$$

Pada model tersebut memiliki interpretasi bahwa jika persentase rumah tangga ber-PHBS kurang dari 65,6938 persen, maka apabila persentase rumah tangga ber-PHBS naik sebesar 1 persen, maka kasus tuberkulosis akan naik sebesar 12,7073. Selanjutnya apabila persentase rumah tangga ber-PHBS lebih dari 67,6779, maka apabila persentase rumah tangga ber-PHBS naik sebesar 1 persen, maka kasus tuberkulosis akan turun sebesar 4992,590.

## V. KESIMPULAN DAN SARAN

Jumlah kasus tuberkulosis di Jawa Timur pada tahun 2014 sebanyak 42.458 kasus dengan kasus BTA positif baru sebanyak 22.866 kasus. Kota Surabaya merupakan kota yang memiliki jumlah kasus tuberkulosis terbanyak yaitu sebesar 4.338 kasus. Tingginya jumlah kasus tuberkulosis di Kota Surabaya dapat disebabkan karena

Kota Surabaya merupakan salah satu kota dengan tingkat kepadatan penduduk tertinggi di Jawa Timur yaitu sebesar 8683 jiwa/km<sup>2</sup>. Kabupaten/kota dengan jumlah kasus tuberkulosis terbanyak kedua adalah Kabupaten Jember yaitu sebanyak 3.139 kasus. Kota dengan jumlah kasus tuberkulosis terendah adalah Kota Batu yaitu sebanyak 176 kasus.

Model regresi nonparametrik *spline* terbaik untuk pemodelan jumlah kasus tuberkulosis di Jawa Timur adalah dengan menggunakan kombinasi knot 1,3,2,3,2. Model ini mempunyai nilai koefisien determinasi ( $R^2$ ) sebesar 83,42% dengan tiga variabel yang berpengaruh secara signifikan yaitu persentase gizi buruk masyarakat, persentase tenaga kesehatan terlatih TB, dan persentase rumah tangga ber-PHBS.

Terdapat beberapa saran yang dapat diberikan oleh penulis yaitu bagi peneliti selanjutnya diharapkan perlu adanya pengembangan menggunakan empat titik knot karena pada penelitian ini masih terbatas menggunakan regresi nonparametrik *spline* linier hanya sampai tiga knot. Sedangkan untuk pemerintah Jawa Timur, untuk mensosialisasikan mengenai pentingnya kebersihan lingkungan bagi kesehatan.

## DAFTAR PUSTAKA

- [1] Kementerian Kesehatan RI (2011). *Strategi Nasional Pengendalian Tuberkulosis di Indonesia 2010-2014*. Jakarta: Kementerian Kesehatan RI Direktorat Jendral Pengendalian Penyakit dan Penyehatan Lingkungan
- [2] Hasyim, M., 2010. *Pemodelan Angka Kejadian Penyakit Infeksi Tuberkulosis Paru (TB Paru) di Kabupaten Sorong Selatan (Provinsi Papua Barat) dengan Pendekatan Multivariate Adaptive Regression Spline (MARS)*. Surabaya: Tugas Akhir Jurusan Statistika FMIPA-ITS Surabaya.
- [3] Muaz, F. (2014). *Faktor-Faktor yang Mempengaruhi Kejadian Tuberkulosis Paru Basil Tahan Asam Positif di Puskesmas Wilayah Kecamatan Serang Kota Serang Tahun 2014*. Jakarta: Tugas Akhir Jurusan Pendidikan Dokter UIN Syarif Hidayatullah Jakarta.
- [4] Puspita, E. 2014. *Analisis Regresi Logistik Biner pada Faktor Resiko yang Mempengaruhi Penderita Penyakit TB Paru di RSU Haji Surabaya*. Tugas Akhir, Jurusan Statistika, ITS.
- [5] Lestari, R. D., 2014. *Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Penyakit Tuberkulosis di Jawa Timur dengan Pendekatan Generalized Poisson Regression (GPR) dan Geographically Weighted Poisson Regression (GWPR)*. Surabaya: Tugas Akhir Jurusan Statistika ITS Surabaya
- [6] Budiantara, I. N. (2006). Model Spline dengan Knot Optimal. *Jurnal Ilmu Dasar*, FMIPA, Universitas Jember, 7,77-85.
- [7] Budiantara, I. N. (2007). Inferensi Statistik untuk Model Spline. *Jurnal Ilmiah Matematika dan Statistika (Matstat)*, vol. 7, No. 1, Universitas Bina Nusantara, Jakarta, 1-14.
- [8] Hardle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press. New York.
- [9] Eubank, R. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker Inc. New York.
- [10] Daniel, W. 1990. *Applied Nonparametric Statistical Method (2nd ed)*. Boston: PWS-Kent Publishing Company.