

Faktor-Faktor Yang Mempengaruhi Angka Kecelakaan Lalu Lintas di Provinsi Jawa Timur Menggunakan Regresi Nonparametrik Spline *Truncated*

Nym Cista Striratna Dewi dan I Nyoman Budiantara

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data (FMKSD)

Institut Teknologi Sepuluh Nopember (ITS)

e-mail: i_nyoman_b@statistika.its.ac.id

Abstrak—Jawa Timur merupakan salah satu provinsi yang memiliki kepadatan penduduk yang tinggi di Indonesia. Kepadatan penduduk yang tinggi menyebabkan timbulnya masalah kecelakaan lalu lintas. Selama tiga tahun berturut-turut (2012-2014), Jawa Timur menjadi provinsi dengan jumlah kecelakaan lalu lintas tertinggi di Indonesia. Oleh sebab itu, dilakukan penelitian untuk mengetahui faktor penyebab terjadinya kecelakaan lalu lintas di Jawa Timur. Pada penelitian ini, angka kecelakaan lalu lintas dan faktor-faktor yang mempengaruhinya tidak membentuk pola tertentu, sehingga digunakan metode regresi nonparametrik Spline. Model terbaik didapatkan dari titik knot optimum berdasarkan nilai GCV yang paling minimum. Model regresi nonparametrik Spline terbaik adalah dengan menggunakan kombinasi knot (1,3,3,3,3). Kelima variabel prediktor yang digunakan berpengaruh terhadap variabel respon, yaitu kepadatan penduduk, persentase pelaku usia remaja, persentase pendidikan terakhir pelaku lebih dari SMP, rasio jenis kelamin pelaku, dan persentase pelaku tidak memiliki SIM. Nilai koefisien determinasi yang dihasilkan dari model ini sebesar 86,73%.

Kata Kunci — Angka Kecelakaan Lalu Lintas, GCV, Regresi Nonparametrik, Spline *Truncated*, Titik Knot.

I. PENDAHULUAN

PROVINSI Jawa Timur merupakan salah satu provinsi yang memiliki penduduk dengan jumlah yang tinggi. Pada tahun 2015, Provinsi Jawa Timur memiliki jumlah penduduk sebanyak 38.847.561 jiwa [1]. Hal ini pula menyebabkan Jawa Timur menjadi salah satu provinsi yang memiliki kepadatan penduduk yang tinggi di Indonesia yakni sebesar 813 jiwa/km² pada tahun 2015. Kepadatan penduduk yang tinggi mengakibatkan Jawa Timur menghadapi berbagai masalah, salah satunya masalah transportasi. Ketergantungan masyarakat Jawa Timur terhadap sektor transportasi, ditambah dengan padatnya penduduk menyebabkan banyaknya masalah transportasi yang terjadi di provinsi ini, diantaranya adalah kecelakaan lalu lintas.

Kecelakaan lalu lintas merupakan suatu peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia maupun kerugian harta benda [2]. Data BPS Provinsi Jawa Timur mencatat bahwa selama tiga tahun berturut-turut, Jawa Timur menjadi provinsi dengan jumlah kecelakaan lalu lintas tertinggi di Indonesia. Pada tahun 2012, tercatat jumlah kecelakaan lalu

lintas di Jawa Timur sebanyak 24.985 kasus. Selanjutnya, pada tahun 2013 menjadi 21.362 kasus dan pada tahun 2014 sebanyak 19.068 kasus kecelakaan [3]. Hal ini menyebabkan kecelakaan lalu lintas menjadi salah satu obyek yang menarik untuk diteliti dan dikaji lebih lanjut, yaitu dengan mengetahui hubungan kecelakaan lalu lintas dengan faktor-faktor yang diduga mempengaruhinya.

Metode yang berfungsi untuk mengetahui model hubungan antar variabel respon dengan variabel prediktor adalah metode regresi. Ada tiga pendekatan dalam metode analisis regresi untuk mengestimasi kurva regresi yaitu pendekatan parametrik, semiparametrik, dan nonparametrik [4]. Regresi parametrik digunakan jika kurva regresi mengikuti pola tertentu atau membentuk pola data yang jelas seperti linier, kuadratik, dan kubik. Regresi nonparametrik digunakan apabila kurva regresi tidak diketahui polanya, sedangkan regresi semiparametrik digunakan apabila terdapat komponen parametrik dan nonparametrik [5]. Pada regresi nonparametrik, data dibiarkan mencari sendiri pola datanya sehingga subjektivitas dari peneliti dapat diminimalisir, dengan demikian pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi [6]. Terdapat beberapa pendekatan regresi nonparametrik diantaranya adalah Spline. Pendekatan Spline merupakan model yang mempunyai interpretasi statistik dan visual yang sangat khusus dan sangat baik, serta Spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub interval tertentu [7].

Pada penelitian ini, akan digunakan metode regresi nonparametrik Spline *Truncated* untuk mengetahui faktor-faktor yang mempengaruhi kecelakaan lalu lintas. Penelitian sebelumnya pernah dilakukan tentang kecelakaan lalu lintas di Provinsi Jawa Timur menggunakan metode GWR [8]. Sebelumnya juga pernah dilakukan penelitian mengenai regresi nonparametrik dengan kasus persentase penduduk miskin dan pengeluaran perkapita makanan di Jawa Timur [9].

Penelitian ini bertujuan untuk mendeskripsikan karakteristik angka kecelakaan lalu lintas berdasarkan variabel yang telah digunakan pada penelitian - penelitian sebelumnya serta memperoleh variabel yang mempengaruhi terjadinya angka kecelakaan lalu lintas berdasarkan pemodelan dengan regresi nonparametrik Spline. Batasan masalah pada penelitian ini antara lain: faktor penyebab angka kecelakaan lalu lintas hanya meneliti pada faktor

manusia, menggunakan metode *Generalized Cross Validation* (GCV) dalam pemilihan titik knot optimal pada Spline linear 1,2,3 dan kombinasi knot. Manfaat dari penelitian ini salah satunya dapat memberikan informasi untuk pemerintah maupun masyarakat di setiap kabupaten atau kota di Jawa Timur dalam mencegah terjadinya kecelakaan lalu lintas.

II. TINJAUAN PUSTAKA

A. Regresi Parametrik

Regresi parametrik adalah metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon dengan asumsi bentuk kurva regresinya diketahui. Bentuk paling dasar dari regresi parametrik adalah regresi linier. Model regresi linier multivariabel atau disebut juga regresi linier berganda dengan p variabel prediktor ditunjukkan pada persamaan sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, i=1,2,\dots,n \quad (1)$$

dimana,

y_i adalah variabel respon, sedangkan $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ merupakan parameter, $x_{1i}, x_{2i}, \dots, x_{pi}$ adalah variabel prediktor, dan ε_i adalah *error* random yang diasumsikan identik, independen, dan berdistribusi normal [10].

B. Regresi Nonparametrik Spline

Regresi nonparametrik merupakan sebuah metode regresi yang tidak terikat asumsi bentuk kurva regresi tertentu. Pendekatan regresi nonparametrik digunakan untuk menyelesaikan pola data antara variabel respon dengan variabel prediktor yang tidak membentuk pola tertentu [11][12]. Model regresi nonparametrik secara umum dapat disajikan sebagai berikut:

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

dimana y_i merupakan variabel respon, x_i merupakan variabel prediktor, $f(x_i)$ merupakan fungsi regresi yang tidak mengikuti pola tertentu, dan ε_i adalah *error* random ke- i yang diasumsikan IIDN.

Spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah pada sub interval tertentu, serta memiliki kelebihan untuk cenderung mencari sendiri estimasi data kemana pola data tersebut bergerak [13]. Kelebihan ini terjadi karena Spline memiliki titik-titik knot K_1, K_2, \dots, K_r . Spline merupakan potongan-potongan polinomial yang memiliki sifat tersegmen pada titik knot. Salah satu kelemahan fungsi polinomial adalah bersifat global, sehingga dikembangkan suatu fungsi yang dapat mengatasi kelemahan polinomial yaitu fungsi Spline *Truncated* yang mempertahankan fungsi polinomial [14]. Fungsi Spline *Truncated* univariabel dengan derajat m diberikan oleh persamaan sebagai berikut.

$$f(x_i) = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m \quad (3)$$

sehingga, model regresi nonparametrik Spline univariabel berdasarkan persamaan sebagai berikut.

$$y_i = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m + \varepsilon_i, i=1,2,\dots,n \quad (4)$$

fungsi $(x - K_k)_+^m$ adalah fungsi potongan (*Truncated*) yang diberikan oleh:

$$(x_i - K_k)_+^m = \begin{cases} (x_i - K_k)^m, & x_i \geq K_k \\ 0 & , x_i < K_k \end{cases} \quad (5)$$

dimana,

β_j : parameter model polinomial, $j = 1, 2, \dots, m$

x_i : variabel prediktor, $i = 1, 2, \dots, n$

β_{m+k} : parameter pada komponen *Truncated*, $k = 1, 2, \dots, r$

r : banyaknya knot

K_k : titik-titik knot

C. Pemilihan Titik Knot Optimal

Salah satu metode yang digunakan untuk pemilihan titik knot optimal adalah *Generalized Cross Validation* (GCV). Model spline dengan nilai GCV terkecil dari titik knot optimal merupakan model spline yang terbaik [15]. Metode GCV dapat dituliskan sebagai berikut.

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{(n^{-1} \text{trace}[\mathbf{I} - \mathbf{A}(K_1, K_2, \dots, K_r)])^2} \quad (6)$$

dimana,

K_1, K_2, \dots, K_r : titik knot pertama hingga ke- r

\mathbf{I} : matriks identitas

n : jumlah pengamatan

$\mathbf{A}(K_1, K_2, \dots, K_r) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

$$MSE(K_1, K_2, \dots, K_r) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

D. Pengujian Parameter Model

Pengujian parameter model regresi nonparametrik Spline dilakukan untuk mengetahui apakah parameter signifikan terhadap model [10].

1. Pengujian Parameter Model Secara Serentak

Uji serentak dilakukan untuk mengetahui signifikansi parameter model regresi secara bersama-sama.

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

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

dengan $m+r$ merupakan banyaknya parameter dalam model regresi nonparametrik Spline kecuali β_0 . Statistik uji menggunakan uji F sebagai berikut.

$$F_{hitung} = \frac{MSR}{MSE} \quad (8)$$

dengan,

$$MSR (\text{Mean Square Regresi}) = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{m+r}$$

$$MSE (\text{Mean Square Error}) = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (m+r) - 1}$$

Keputusan H_0 ditolak apabila $F_{hitung} > F_{\alpha, (m+r), (n-(m+r)-1)}$, sehingga dapat disimpulkan minimal terdapat satu parameter pada model regresi nonparametrik Spline yang signifikan terhadap model.

2. Pengujian Parameter Model Secara Individu

Dalam pengujian secara parsial, variabel prediktor dikatakan berpengaruh terhadap variabel respon apabila terdapat minimal satu parameter yang signifikan, dengan hipotesis sebagai berikut:

$$H_0 : \beta_g = 0$$

$$H_1 : \beta_g \neq 0, \quad g = 1, 2, \dots, m+r$$

Pengujian secara individu dilakukan dengan menggunakan uji *t*. Statistik uji yang digunakan adalah sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_g}{se(\hat{\beta}_g)} \tag{9}$$

dimana,

$\hat{\beta}_g$: penaksir parameter β_g

$se(\hat{\beta}_g)$: standart error dari $\hat{\beta}_g$

Nilai $se(\hat{\beta}_g)$ didapatkan dari $\sqrt{var(\hat{\beta}_g)}$, $var(\hat{\beta}_g)$ adalah elemen diagonal ke-*g* dari matriks sebagai berikut :

$$\begin{aligned} var(\hat{\beta}) &= var[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' var(\mathbf{y})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= \sigma^2(\mathbf{X}\mathbf{X})^{-1} \end{aligned} \tag{10}$$

Keputusan H_0 ditolak apabila $|t_{hitung}| > t_{\alpha/2; (n-(m+r)-1)}$ yang berarti bahwa parameter signifikan terhadap model.

E. Kriteria Pemilihan Model Terbaik

Salah satu kriteria yang digunakan untuk menentukan model regresi terbaik yaitu menggunakan koefisien determinasi (R^2). Semakin tinggi nilai R^2 yang dihasilkan dari suatu model, maka semakin baik pula variabel prediktor dalam menjelaskan variabilitas variabel respon [10]. Berikut merupakan rumus dari (R^2).

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{11}$$

F. Angka Kecelakaan Lalu Lintas

Angka kecelakaan lalu lintas digunakan sebagai alat penentu untuk menyatakan tingkat kecelakaan lalu lintas pada satu satuan panjang jalan [16]. Panjang jalan yang digunakan dalam penelitian ini adalah jalan kabupaten/kota. Berikut merupakan rumus angka kecelakaan lalu lintas yang digunakan.

$$AR = \frac{A}{L} \times 100 \tag{12}$$

Keterangan:

- AR : Angka kecelakaan total per 100 kilometer
- A : Total kecelakaan yang terjadi pada tahun tertentu
- L : Panjang dari bagian jalan yang di kontrol (km)

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan adalah data sekunder yang diambil dari Kepolisian Daerah (Polda) Jawa Timur dan Badan Pusat Statistik (BPS) Provinsi Jawa Timur. Data yang digunakan adalah tahun 2015 dengan unit penelitian sebanyak 38 kabupaten/kota di Provinsi Jawa Timur.

B. Variabel Penelitian

Variabel penelitian yang digunakan untuk memodelkan angka kecelakaan lalu lintas di Provinsi Jawa Timur disajikan dalam Tabel 1 sebagai berikut.

Tabel 1.
Variabel Penelitian

Variabel	Keterangan
y	Angka kecelakaan lalu lintas
x_1	Kepadatan penduduk
x_2	Persentase pelaku usia remaja
x_3	Persentase pendidikan terakhir pelaku lebih dari SMP
x_4	Rasio jenis kelamin pelaku
x_5	Persentase pelaku kecelakaan tidak memiliki SIM

C. Langkah Analisis

Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan karakteristik angka kecelakaan lalu lintas di Provinsi Jawa Timur beserta faktor-faktor yang diduga mempengaruhinya.
2. Melakukan pemodelan angka kecelakaan lalu lintas menggunakan regresi linier berganda.
3. Membuat *scatterplot* antara variabel respon dengan masing - masing variabel prediktor.
4. Memodelkan variabel respon angka kecelakaan lalu lintas menggunakan regresi nonparametrik Spline dengan satu titik knot, dua titik knot, tiga titik knot dan kombinasi titik knot.
5. Memilih titik knot optimal berdasarkan *Generalized Cross Validation (GCV)* minimum.
6. Mendapatkan model regresi nonparametrik Spline terbaik dengan titik knot optimal.
7. Melakukan pengujian signifikansi parameter regresi nonparametrik Spline secara serentak dan parsial.
8. Melakukan uji asumsi residual identik, independen, dan berdistribusi normal (IIDN).
9. Menguji kebaikan model dengan menghitung koefisien determinasi (R^2).
10. Menarik kesimpulan dari hasil analisis.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Data

Angka kecelakaan lalu lintas merupakan alat penentu untuk menyatakan tingkat kecelakaan lalu lintas pada satu satuan segmen jalan. Berikut ini adalah karakteristik dari angka kecelakaan lalu lintas di Jawa Timur beserta variabel prediktor yang diduga mempengaruhinya menggunakan statistika deskriptif yang disajikan dalam Tabel 2.

Tabel 2.
Statistika Deskriptif Variabel

Var.	Rata-rata	Standar Deviasi	Minimum	Maksimum
y	85,545	85,127	6,471	439,602
x_1	1829,658	2184,106	388	8606
x_2	8,387	6,065	1,980	25,676
x_3	75,057	17,615	16,304	98,390
x_4	14,651	6,720	7,224	39,316
x_5	57,949	15,015	11,782	83,005

Berdasarkan Tabel 2, variabel respon (y) merupakan angka kecelakaan di Provinsi Jawa Timur dengan rata-rata sebesar 85,545. Hal ini berarti pada tahun 2015, angka kecelakaan lalu lintas di Jawa Timur cukup tinggi yaitu di setiap 100 kilometer rata-rata terjadi 85,545 atau 86 kecelakaan lalu lintas. Angka kecelakaan terendah yakni sebesar 6,47 di Kabupaten Blitar dan angka kecelakaan tertinggi yaitu sebesar 439,6 terdapat di Kota Pasuruan. Kemudian diketahui bahwa variabel kepadatan penduduk memiliki rata-rata sebesar 1829,658. Hal ini menunjukkan bahwa dalam setiap 1 kilometer persegi, rata-rata daerah di

Jawa Timur ditempati oleh 1830 jiwa, kepadatan penduduk yang paling rendah sebesar 388 jiwa/km² terdapat di Kabupaten Pacitan, sedangkan kepadatan penduduk di Provinsi Jawa Timur yang paling tinggi sebesar 8606 jiwa/km² berada di Kota Surabaya. Pada Tabel 2 diketahui bahwa variabel persentase pelaku kecelakaan usia remaja memiliki rata-rata sebesar 8,387 persen, nilai persentase pelaku usia remaja tertinggi sebesar 25,68 persen terdapat di Kabupaten Tulungagung, sedangkan yang terendah berada di Kabupaten Sumenep sebesar 1,98 persen. Selanjutnya diketahui bahwa variabel pelaku kecelakaan berpendidikan terakhir di atas SMP memiliki persentase rata-rata sebesar 75,057 persen, persentase terendah pada pelaku kecelakaan berpendidikan terakhir di atas SMP sebesar 16,3 persen berada di Kabupaten Sampang, sedangkan persentase tertinggi sebesar 98,39 persen terdapat di Kabupaten Sidoarjo. Variabel rasio jenis kelamin pelaku di Provinsi Jawa Timur memiliki rata-rata sebesar 14,651 per 100 pelaku kecelakaan atau menunjukkan bahwa pelaku kecelakaan berjenis kelamin laki-laki lebih banyak daripada perempuan, rasio jenis kelamin pelaku terendah berada di Kabupaten Jember yakni sebesar 7,224 dan rasio tertinggi sebesar 39,316 terdapat di Kabupaten Madiun. Serta variabel persentase pelaku tidak memiliki SIM mempunyai rata-rata sebesar 57,949 persen. Persentase tertinggi terdapat di Kabupaten Jember, yaitu sebesar 83,005 persen, nilai terendah pada persentase pelaku yang tidak memiliki SIM sebesar 11,78 persen terdapat di Kabupaten Sidoarjo, sedangkan persentase terendah sebesar 11,782 persen terdapat di Kabupaten Sidoarjo. Berdasarkan Tabel 2, diketahui bahwa standar deviasi untuk variabel kepadatan penduduk lebih tinggi dibandingkan variabel lainnya. Hal ini menunjukkan bahwa kepadatan penduduk merupakan variabel yang paling beragam dalam penelitian angka kecelakaan lalu lintas di Jawa Timur beserta faktor-faktor yang mempengaruhinya.

B. Pemodelan Angka Kecelakaan Lalu Lintas di Jawa Timur Menggunakan Regresi Linier Berganda

Pemodelan ini dilakukan dengan menganggap bahwa kurva regresinya adalah linier. Pada regresi linier berganda akan dilakukan pengujian signifikansi parameter secara serentak dan parsial. Pengujian serentak dilakukan untuk mengetahui apakah seluruh parameter signifikan terhadap angka kecelakaan lalu lintas di Jawa Timur. Berikut merupakan hipotesis yang digunakan.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_5 = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_j \neq 0, \text{ dimana } j=1,2,\dots,5$$

Tabel 3.
Uji Serentak Regresi Linier Berganda

Sumber	df	SS	MS	F
Regresi	5	96843	19369	3,62
Error	32	171282	5353	
Total	37	268125		

Berdasarkan Tabel 3, menggunakan taraf signifikan sebesar 5% maka didapatkan keputusan tolak H₀ karena nilai F_{hitung}=3,62 > F_(0,05;5;32)= 2,512. Hal ini menunjukkan bahwa minimal terdapat satu parameter yang signifikan pada model atau secara bersama terdapat minimal satu variabel prediktor berpengaruh terhadap angka kecelakaan lalu lintas di Jawa Timur. Selanjutnya akan dilakukan pengujian parsial untuk mengetahui variabel mana saja

yang berpengaruh terhadap angka kecelakaan lalu lintas di Jawa Timur. Berikut merupakan hipotesis yang digunakan.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \text{ dimana } j=1,2,\dots,5$$

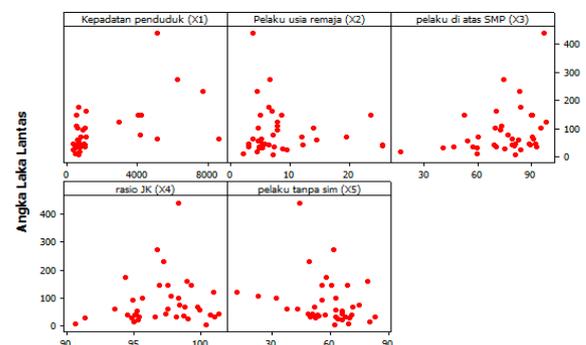
Tabel 4.
Uji Parsial Regresi Linier Berganda

Parameter	Estimasi	t	Keputusan
β_1	0,021	3,32	Signifikan
β_2	0,101	0,05	Tidak signifikan
β_3	0,922	1,2	Tidak signifikan
β_4	1,334	0,71	Tidak signifikan
β_5	0,064	0,07	Tidak signifikan

Hasil pada Tabel 4 menjelaskan bahwa dengan menggunakan taraf signifikan sebesar 5%, diperoleh nilai $t_{(0,05;32)}=2,037$ sehingga didapatkan kesimpulan variabel yang tidak berpengaruh terhadap angka kecelakaan lalu lintas adalah persentase pelaku usia remaja (x₂), persentase pelaku kecelakaan berpendidikan akhir lebih dari SMP (x₃), rasio jenis kelamin pelaku (x₄), dan persentase pelaku yang tidak memiliki SIM (x₅), sedangkan variabel kepadatan penduduk (x₁) berpengaruh terhadap angka kecelakaan lalu lintas. Nilai R-sq yang diperoleh dari model ini sebesar 36,1%. Berdasarkan hasil pengujian parsial, analisis menggunakan regresi ini memberikan kesimpulan yang berbeda dari penelitian yang pernah dilakukan sebelumnya, serta didasarkan pada nilai R-sq yang didapatkan model ini belum cukup baik dalam menjelaskan variabilitas angka kecelakaan lalu lintas di Jawa Timur. Dengan demikian, akan dilakukan analisis menggunakan metode regresi lain dengan mengamati pola data masing-masing variabel prediktor dengan variabel responnya terlebih dahulu.

C. Pola Data Antara Angka Kecelakaan Lalu Lintas dengan Variabel yang Diduga Mempengaruhi

Langkah pertama dalam melakukan analisis regresi adalah membuat *scatterplot* untuk mengetahui pola hubungan variabel prediktor terhadap variabel respon. Gambar 1 menunjukkan bahwa pola hubungan antara angka kecelakaan lalu lintas dengan seluruh variabel prediktor tidak membentuk pola tertentu, sehingga metode yang digunakan adalah regresi nonparametrik Spline.



Gambar 1. Scatterplot Variabel Prediktor dengan Respon.

D. Pemilihan Titik Knot Optimal

Model terbaik didapatkan dari titik knot yang optimum. Tabel 5 menunjukkan sepuluh nilai GCV yang berada di sekitar nilai GCV paling minimum untuk model regresi nonparametrik Spline satu titik knot.

Tabel 5.
Titik Knot dan GCV untuk Satu Knot

No	Knot					GCV
	x_1	x_2	x_3	x_4	x_5	
1	6928,86	20,84	81,64	32,77	68,47	6313,01
2	7096,57	21,32	83,31	33,42	69,92	6249,90
3	7264,29	21,81	84,99	34,08	71,38	6131,91
4	7432,00	22,29	86,66	34,73	72,83	5977,80
5	7599,71	22,77	88,34	35,39	74,28	5809,16
6	7767,43	23,26	90,01	36,04	75,74	5626,40
7	7935,14	23,74	91,69	36,70	77,19	5359,52
8	8102,86	24,23	93,36	37,35	78,64	5329,14
9	8270,57	24,71	95,04	38,01	80,10	5496,00
10	8438,29	25,19	96,71	38,66	81,55	6258,35

Tabel 6.
Nilai GCV Minimum Titik Knot Optimal

Jumlah Knot	GCV Minimum
Satu Titik Knot	5329,13551947391
Dua Titik Knot	3791,55173130658
Tiga Titik Knot	2794,3354549638
Kombinasi Knot (1,3,3,3,3)	2794,33545685926

Tabel 5 menunjukkan bahwa nilai GCV pada satu titik knot yang paling minimum adalah 5329,14 dengan titik knot optimum yaitu variabel kepadatan penduduk (x_1) berada pada titik knot 8102,86; variabel persentase pelaku usia remaja (x_2) berada pada titik knot 24,23; variabel persentase pendidikan terakhir pelaku adalah di atas SMP (x_3) berada pada titik knot 93,36; variabel rasio jenis kelamin pelaku (x_4) berada pada titik knot 37,35; serta variabel persentase pelaku kecelakaan tidak memiliki SIM (x_5) berada pada titik knot 78,64.

Selanjutnya untuk interpretasi pemilihan titik knot optimum pada dua titik knot, tiga titik knot, dan kombinasi titik knot dilakukan dengan cara yang serupa seperti pada satu titik knot. Langkah selanjutnya adalah pemilihan nilai GCV minimum dari masing-masing titik knot yang disajikan pada Tabel 6.

Berdasarkan Tabel 6 diperoleh informasi bahwa nilai GCV minimum pada tiga knot dan kombinasi knot adalah sama. Dalam penelitian ini dipilih titik knot yang memiliki jumlah parameter paling sedikit sesuai dengan teori *parsimony*, dimana suatu model yang baik adalah model yang memiliki jumlah parameter paling sederhana. Oleh karena itu, diputuskan titik knot optimum yang terpilih adalah kombinasi titik knot (1,3,3,3,3).

E. Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline

1. Uji Serentak

Pengujian secara serentak dilakukan pada parameter model regresi terhadap variabel respon secara bersama-sama. Hipotesis yang digunakan adalah sebagai berikut

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{18} = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_g \neq 0, g = 1, 2, \dots, 18$$

Hasil pengujian secara serentak ditampilkan pada Tabel 7.

Berdasarkan Tabel 7, menggunakan taraf signifikan sebesar 5% maka didapatkan keputusan tolak H_0 karena nilai $F_{hitung} = 5,863 > F_{(0,05;18,19)} = 2,182$. Hal ini menunjukkan bahwa minimal terdapat satu parameter yang signifikan pada model atau secara bersama terdapat minimal satu variabel prediktor berpengaruh secara signifikan terhadap angka kecelakaan lalu lintas di Jawa Timur.

2. Uji Parsial

Pada uji secara parsial, variabel prediktor dikatakan berpengaruh terhadap variabel respon apabila terdapat

Tabel 7.
Analisis Ragam Uji Serentak

Sumber	df	SS	MS	F
Regresi	18	232534,90	12918,61	6,897
Error	19	35590,87	1873,20	
Total	37	268125,80		

Tabel 8.
Hasil Uji Parsial

Variabel	Par.	Estimasi	t	Keputusan	Kesimpulan
x_1	β_1	0,029	6,723	Tolak	x_1 Berpengaruh
	β_2	-0,488	-4,533	Tolak	
	β_3	3,566	1,504	Gagal Tolak	
x_2	β_4	316,341	2,048	Gagal Tolak	x_2 Berpengaruh
	β_5	-532,815	-2,405	Tolak	
	β_6	-211,674	-2,464	Tolak	
	β_7	0,177	0,252	Gagal Tolak	
x_3	β_8	-118,672	-4,151	Tolak	x_3 Berpengaruh
	β_9	338,428	5,203	Tolak	
	β_{10}	-463,959	-5,502	Tolak	
x_4	β_{11}	-0,410	-0,257	Gagal Tolak	x_4 Berpengaruh
	β_{12}	68,385	3,164	Tolak	
	β_{13}	41,031	3,164	Tolak	
	β_{14}	13,677	3,164	Tolak	
	β_{15}	-1,589	-2,037	Gagal Tolak	
x_5	β_{16}	69,061	2,467	Tolak	x_5 Berpengaruh
	β_{17}	-173,130	-1,914	Gagal Tolak	
	β_{18}	458,183	2,853	Tolak	

minimal satu parameter yang signifikan. Adapun pengujian dilakukan dengan hipotesis sebagai berikut.

$$H_0 : \beta_g = 0$$

$$H_1 : \beta_g \neq 0, g = 1, 2, \dots, 18$$

Hasil pengujian secara serentak ditampilkan pada Tabel 8.

Pada Tabel 8 diketahui bahwa kelima variabel prediktor berpengaruh terhadap model karena terdapat minimal satu nilai $|t_{hitung}| > t_{(0,025;19)} = 2,093$, sehingga variabel kepadatan penduduk (x_1), variabel persentase pelaku usia remaja (x_2), variabel persentase pendidikan terakhir pelaku adalah di atas SMP (x_3), variabel rasio jenis kelamin pelaku (x_4), dan variabel persentase pelaku kecelakaan tidak memiliki SIM (x_5) berpengaruh terhadap angka kecelakaan lalu lintas di Jawa Timur. Setelah dilakukan pengujian signifikansi parameter secara parsial, selanjutnya dilakukan pengujian asumsi residual IIDN. Diperoleh hasil bahwa residual telah memenuhi asumsi identik, independen, dan berdistribusi normal.

F. Koefisien Determinasi (R^2)

Berdasarkan perhitungan didapatkan nilai R^2 sebesar 86,73 persen. Hal ini berarti model regresi nonparametrik Spline yang didapatkan mampu menjelaskan variabilitas angka kecelakaan lalu lintas di Jawa Timur sebesar 86,73 persen, sedangkan sisanya dijelaskan oleh variabel lain yang tidak masuk dalam model.

G. Interpretasi Model Regresi Nonparametrik Spline

Setelah melakukan pengujian parameter dan semua asumsi residual terpenuhi, selanjutnya menginterpretasikan model regresi yang telah didapatkan. Model yang terbentuk berdasarkan nilai GCV yang paling minimum yaitu menggunakan kombinasi titik knot (1,3,3,3,3) adalah sebagai berikut.

$$\hat{y} = 93,015 + 0,029x_1 - 0,488(x_1 - 8102,86)_+^1 + 3,566x_2 + 316,341(x_2 - 23,26)_+^1 - 532,815(x_2 - 24,23)_+^1 - 211,674(x_2 - 25,19)_+^1 + 0,177x_3 - 118,672(x_3 - 90,01)_+^1 + 338,428(x_3 - 93,36)_+^1 - 463,59(x_3 - 96,71)_+^1 - 0,41x_4 + 68,385(x_4 - 36,04)_+^1 + 41,031(x_4 - 37,35)_+^1 + 13,677(x_4 - 38,66)_+^1 - 1,589x_5 + 69,061(x_5 - 75,74)_+^1 - 173,13(x_5 - 78,64)_+^1 + 458,183(x_5 - 81,55)_+^1$$

Berdasarkan model tersebut, dapat diinterpretasikan dari masing-masing variabel yang berpengaruh terhadap angka kecelakaan lalu lintas di Jawa Timur sebagai berikut.

1. Dengan mengasumsikan variabel x_2 , x_3 , x_4 , dan x_5 konstan, maka pengaruh variabel kepadatan penduduk (x_1) terhadap angka kecelakaan lalu lintas di Jawa Timur dapat ditulis sebagai berikut.

$$\hat{y} = 0,029x_1 - 0,488(x_1 - 8102,86)_+^1$$

Apabila kepadatan penduduk kurang dari 8102,86 jiwa/km² naik sebesar 1 jiwa/km² maka angka kecelakaan lalu lintas di Jawa Timur akan meningkat sebesar 0,029 kecelakaan per 100 kilometer jalan. Kemudian apabila kepadatan penduduk lebih dari atau sama dengan 8102,86 jiwa/km² mengalami kenaikan sebesar 1 jiwa/km² maka angka kecelakaan lalu lintas di Jawa Timur akan menurun sebesar 0,459 kecelakaan per 100 kilometer jalan. Wilayah yang termasuk dalam interval ini hanya Kota Surabaya. Pengelompokan kabupaten/kota secara visual berdasarkan variabel kepadatan penduduk yang mempengaruhi angka kecelakaan lalu lintas di Jawa Timur disajikan pada Gambar 4 sebagai berikut.



Gambar 4. Peta Kepadatan Penduduk di Jawa Timur

2. Dengan mengasumsikan variabel x_1 , x_3 , x_4 , dan x_5 konstan, maka pengaruh variabel persentase pelaku usia remaja (x_2) terhadap angka kecelakaan lalu lintas di Jawa Timur dapat ditulis sebagai berikut.

$$\hat{y} = 3,566x_2 + 316,341(x_2 - 23,26)_+^1 - 532,815(x_2 - 24,23)_+^1 - 211,674(x_2 - 25,19)_+^1$$

Apabila persentase pelaku usia remaja yang kurang dari 23,26 persen mengalami kenaikan sebesar satu persen, maka akan terjadi kenaikan pada angka kecelakaan lalu lintas sebesar 3,566 kecelakaan per 100 kilometer jalan. Selanjutnya, apabila persentase pelaku usia remaja yang berada di antara 23,26 dan 24,23 naik sebesar satu persen, maka akan terjadi peningkatan angka kecelakaan lalu lintas sebesar 319,907 kecelakaan per 100 kilometer jalan. Wilayah yang masuk dalam interval ini hanya Kabupaten Tuban. Kemudian untuk pelaku usia remaja yang berada lebih dari 25,19 persen, apabila terjadi kenaikan sebesar satu persen maka akan terjadi penurunan pada angka

kecelakaan lalu lintas di Jawa Timur sebesar 185,64 kecelakaan per 100 kilometer jalan. Wilayah yang masuk dalam interval ini adalah Kabupaten Trenggalek dan Kabupaten Tulungagung. Secara visual, pengelompokan kabupaten/kota berdasarkan variabel persentase pelaku usia remaja disajikan pada Gambar 5 sebagai berikut.



Gambar 5. Peta Pelaku Usia Remaja di Jawa Timur

3. Dengan mengasumsikan variabel x_1 , x_2 , x_4 , dan x_5 konstan, maka pengaruh variabel persentase pelaku berpendidikan terakhir di atas SMP (x_3) terhadap angka kecelakaan lalu lintas di Jawa Timur dapat ditulis sebagai berikut.

$$\hat{y} = 0,177x_3 - 118,672(x_3 - 90,01)_+^1 + 338,428(x_3 - 93,36)_+^1 - 463,59(x_3 - 96,71)_+^1$$

Apabila persentase pelaku berpendidikan terakhir di atas SMP kurang dari 90,01 persen mengalami kenaikan sebesar satu persen, maka angka kecelakaan lalu lintas akan meningkat sebesar 0,177 kecelakaan per 100 kilometer. Kemudian apabila persentase pelaku berpendidikan terakhir di atas SMP berada di antara 90,01 dan 93,36 persen naik sebesar satu persen, maka angka kecelakaan lalu lintas akan menurun sebesar 118,495 kecelakaan per 100 kilometer jalan. Kabupaten Tuban, Kabupaten Mojokerto, Kota Madiun, dan Kabupaten Madiun tergolong dalam interval ini. Selanjutnya, apabila persentase pelaku berpendidikan terakhir di atas SMP yang berada di antara 93,36 dan 96,71 persen naik sebesar satu persen, maka angka kecelakaan lalu lintas akan meningkat sebesar 219,933 kejadian kecelakaan per 100 kilometer jalan. Terdapat dua kabupaten yang masuk dalam interval ini yaitu Kabupaten Jember dan Kabupaten Gresik. Serta jika persentase pelaku pendidikan terakhir di atas SMP lebih dari 96,71 persen naik sebesar satu persen, maka angka kecelakaan lalu lintas akan menurun sebesar 244,026 kejadian kecelakaan per 100 kilometer jalan. Kabupaten/kota yang masuk dalam interval ini adalah Kota Pasuruan dan Kabupaten Sidoarjo. Secara visual, pengelompokan kabupaten/kota di Jawa Timur berdasarkan variabel persentase pelaku berpendidikan terakhir lebih dari SMP disajikan pada Gambar 6 berikut.



Gambar 6. Peta Pelaku Berpendidikan Akhir Lebih dari SMP.

4. Dengan menganggap variabel $x_1, x_2, x_3,$ dan x_5 konstan, maka pengaruh variabel rasio jenis kelamin pelaku (x_4) terhadap angka kecelakaan lalu lintas di Jawa Timur dapat ditulis sebagai berikut.

$$\hat{y} = -0,41x_4 + 68,385(x_4 - 36,04)_+^1 + 41,031(x_4 - 37,35)_+^1 + 13,677(x_4 - 38,66)_+^1$$

Apabila rasio jenis kelamin pelaku kecelakaan kurang dari 36,04 dan mengalami pertambahan sebesar satu satuan, maka angka kecelakaan lalu lintas akan turun sebesar 0,41 kejadian kecelakaan per 100 kilometer jalan. Kemudian jika rasio jenis kelamin pelaku yang berada diatas 100,45 mengalami kenaikan sebesar satu satuan, maka angka kecelakaan lalu lintas akan naik sebesar 122,683 kejadian kecelakaan per 100 kilometer jalan. Kabupaten yang masuk dalam interval ini adalah hanya Kabupaten Madiun. Pengelompokan kabupaten/kota di Jawa Timur berdasarkan variabel rasio jenis kelamin pelaku secara visual disajikan pada Gambar 7.



Gambar 7. Peta Rasio Jenis Kelamin Pelaku di Jawa Timur.

5. Dengan menganggap variabel $x_1, x_2, x_3,$ dan x_4 konstan, maka pengaruh variabel persentase pelaku tidak memiliki SIM (x_5) terhadap angka kecelakaan lalu lintas di Jawa Timur dapat ditulis sebagai berikut.

$$\hat{y} = -1,589x_5 + 69,061(x_5 - 75,74)_+^1 - 173,13(x_5 - 78,64)_+^1 + 458,183(x_5 - 81,55)_+^1$$

Apabila persentase pelaku tidak memiliki SIM yang kurang dari 75,74 persen, apabila terjadi kenaikan pada persentase pelaku kecelakaan tidak memiliki SIM sebesar satu persen, maka akan terjadi penurunan angka kecelakaan lalu lintas di Jawa Timur sebesar 1,589 kejadian kecelakaan per 100 kilometer. Selanjutnya, apabila persentase pelaku kecelakaan tidak memiliki SIM yang berada pada interval 78,64 hingga 81,55 persen mengalami kenaikan sebesar satu persen, maka akan terjadi penurunan angka kecelakaan lalu lintas sebesar 105,658 kejadian kecelakaan per 100 kilometer jalan. Terdapat 2 kabupaten/kota yang masuk dalam interval ini yaitu Kabupaten Jombang dan Kabupaten Sampang. Serta, apabila persentase pelaku kecelakaan tidak memiliki SIM yang berada lebih dari 81,55 persen mengalami kenaikan sebesar satu persen maka akan terjadi kenaikan angka kecelakaan lalu lintas di Jawa Timur sebesar 352,525 kejadian kecelakaan per 100 kilometer jalan. Hanya terdapat satu kabupaten yang masuk dalam interval ini yaitu Kabupaten Jember. Secara visual disajikan pada Gambar 8 berikut.



Gambar 8. Peta Pelaku Tidak Memiliki SIM di Jawa Timur

V. KESIMPULAN DAN SARAN

Angka kecelakaan lalu lintas terendah di Provinsi Jawa Timur berada di Kabupaten Blitar dengan 7 kecelakaan lalu lintas, sedangkan angka kecelakaan tertinggi terdapat di Kota Pasuruan 86 kecelakaan lalu lintas. Model regresi nonparametrik Spline terbaik diperoleh menggunakan kombinasi titik knot (1,3,3,3,3) dengan nilai koefisien determinasi (R^2) sebesar 86,73%, dimana seluruh variabel berpengaruh terhadap respon yaitu kepadatan penduduk, persentase pelaku usia remaja, persentase pelaku kecelakaan berpendidikan akhir lebih dari SMP, rasio jenis kelamin pelaku, dan persentase pelaku kecelakaan yang tidak memiliki SIM.

Saran untuk penelitian selanjutnya diharapkan menambah faktor lain yang berkaitan, sehingga dapat lebih menggambarkan kondisi yang mempengaruhi angka kecelakaan lalu lintas di Provinsi Jawa Timur. Serta upaya yang perlu dilakukan pemerintah daerah Provinsi Jawa Timur, salah satunya dengan mengadakan sosialisasi yang berkaitan dengan pemahaman dasar tentang berlalu-lintas yang tertib, mengingat angka kecelakaan lalu lintas yang terjadi di Provinsi Jawa Timur sangat tinggi.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Kependudukan: Proyeksi Penduduk Provinsi Jawa Timur," Jawa Timur, 2015.
- [2] "Undang-Undang Nomor 22 Tahun 2009 tentang Lalu Lintas Jalan dan Angkutan Jalan," 2009.
- [3] Badan Pusat Statistik, "Publikasi Statistik Transportasi 2014," 2014.
- [4] A. Tripena and I. Budiantara, "Fourier Estimator in Nonparametric Regression," in *International Conference on Natural and Applied Natural Science*, 2006, pp. 2-4.
- [5] I. Wulandari and I. N. Budiantara, "Analisis Faktor-faktor yang Mempengaruhi Persentase Penduduk Miskin dan Pengeluaran Perkapita Makanan di Jawa Timur Menggunakan Regresi Nonparametrik Birespon Spline," *J. Sains dan Seni*, pp. D30-D35, 2014.
- [6] B. Lestari, I. N. Budiantara, S. Sunaryo, and M. Mashuri, "Spline Smoothing for Multiresponse Nonparametric Regression Model in Case of Heteroscedasticity of Variance," *J. Math. Stat.*, vol. 8, no. 3, pp. 377-384, 2012.
- [7] I. Budiantara, "Estimator Spline Terbobot dalam Regresi Semiparametrik," *Maj. Ilmu Pengetah. dan Teknol.*, vol. 10, pp. 103-109, 1999.
- [8] P. Dewi, "Pemodelan Faktor Penyebab Kecelakaan Lalu Lintas Berdasarkan Metode Geographically Weighted Regression di Jawa Timur," Institut Teknologi Sepuluh Nopember, 2016.
- [9] A. P. Sugiantari and I. N. Budiantara, "Analisis Faktor-faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline," *J. Sains dan Seni ITS*, vol. 2, no. 1, pp. D37-D41, 2013.
- [10] D. N.R and S. H., *Analisis Regresi Terapan*. Jakarta: Gramedia Pustaka Utama, 1992.
- [11] A. A. Fernandes, I. N. Budiantara, B. W. Otok, and Suhartono., "Spline Estimator for Bi-Response and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data,"

- [12] *J. Math. Stat.*, vol. 11, no. 2, pp. 61–69, 2015.
- [12] A. A. Fernandes, I. N. Budiantara, B. W. Otok, and Suhartono., “Reproducing Kernel Hilbert Space for Penalized Regression Multi Predictors : Case in Longitudinal Data,” *Int. J. Math. Anal.*, vol. 8, no. 40, pp. 1951–1961, 2014.
- [13] I. N. Budiantara, “Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter,” in *Seminar Nasional FMIPA*, 2011.
- [14] N. Chamidah, I. N. Budiantara, S. Sunaryo, and I. Zain, “Disigning of Child Growth Based on Multirespon Local Polynomial Modeling,” *J. Math. Stat.*, vol. 8, no. 3, pp. 342–347, 2012.
- [15] S. F. Nisa’ and I. Budiantara, “Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Spline pada Kasus Demam Berdarah Dengue (DBD),” *J. Sains dan Seni ITS*, pp. D318–D323, 2012.
- [16] S. Rahmadi, “Kecelakaan Lalu Lintas di Kecamatan Duren Sawit Jakarta Timur,” Depok, 2011.