

Pemodelan Faktor Penyebab Kecelakaan Lalu Lintas Berdasarkan Metode *Geographically Weighted Regression* di Jawa Timur

Putu Laksmita Ari Dewi dan Ismaini Zain

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

Jl. Arief Rahman Hakim, Surabaya 60111

e-mail: mitaaridewi@gmail.com

Abstrak— Provinsi Jawa Timur menempati urutan pertama sebagai provinsi yang mengalami kecelakaan lalu lintas paling banyak di Indonesia pada tahun 2011, 2012, 2013. Berdasarkan fakta dan penelitian terdahulu, diketahui bahwa kecelakaan lalu lintas dipengaruhi oleh faktor yang bervariasi di setiap daerah, sehingga digunakan pemodelan menggunakan metode *Geographically Weighted Regression*. Variabel yang diduga mempengaruhi angka kecelakaan lalu lintas di 38 kabupaten/kota di Jawa Timur adalah kepadatan penduduk, persentase usia remaja, persentase kecelakaan terjadi di kawasan jalan kabupaten/kota, persentase pendidikan terakhir pelaku kecelakaan di atas SMP, rasio jenis kelamin dan persentase kecelakaan terjadi pada waktu gelap. Berdasarkan hasil analisis, data telah memenuhi asumsi heterogenitas spasial pada pengujian *Breusch Pagan* dengan *P-value* sebesar 0,009, dan terdapat dependensi spasial berdasarkan uji Moran's I dengan *P-value* sebesar 0,03. Fungsi pembobot yang digunakan pada penelitian ini adalah fungsi pembobot *Fixed Biquare* dengan *bandwidth* sebesar 1,67. Terjadi peningkatan nilai R^2 menjadi 52,80 persen pada model GWR dan penurunan nilai SSE menjadi 14,22. Sehingga, pada kasus ini model GWR lebih baik dalam memodelkan dibanding model OLS. Faktor kepadatan penduduk dan persentase pendidikan terakhir pelaku kecelakaan di atas SMP menjadi faktor yang signifikan mempengaruhi kecelakaan lalu lintas di setiap kabupaten/kota di Jawa Timur dengan taraf signifikansi sebesar 5 persen.

Kata kunci— Dependensi Spasial, Heterogenitas Spasial, GWR, Kecelakaan Lalu Lintas, R^2 , SSE

urutan pertama sebagai provinsi dengan jumlah kecelakaan lalu lintas tertinggi di Indonesia.

Faktor penyebab kecelakaan lalu lintas di Indonesia terbagi menjadi akibat manusia, mesin atau kendaraan, lingkungan dan keadaan jalan [1]. Secara administratif, Provinsi Jawa Timur terdiri dari 29 kabupaten dan 9 kota yang memiliki karakteristik berbeda secara geografis antara daerah satu dan daerah lainnya, karenanya permasalahan lalu lintas termasuk masalah kecelakaan lalu lintas dapat dikaji dari prespektif geografi [2].

Pada penelitian ini diharapkan diketahui hubungan antara kecelakaan lalu lintas di setiap kabupaten di Jawa Timur dengan faktor-faktor yang mempengaruhinya. Untuk mengetahui model hubungan antara variabel respon dan variabel prediktor digunakan metode regresi. Salah satu metode regresi yang menggunakan pertimbangan geografis adalah metode *Geographically Weighted Regression*. Penelitian tentang kecelakaan menggunakan GWR pernah dilakukan di Turki yang menghasilkan GWR lebih baik dibanding OLS berdasarkan peningkatan nilai R^2 [3].

Selain itu, pernah dilakukan penelitian terhadap kerugian yang terjadi akibat kecelakaan lalu lintas di Hampton, Virginia serta faktor-faktor yang signifikan mempengaruhi kecelakaan di daerah tersebut yang menghasilkan bahwa metode GWR lebih baik dibanding metode OLS karena hasil pemodelan lebih sesuai dengan data yang digunakan [4].

I. PENDAHULUAN

Jawa Timur merupakan salah satu provinsi yang memiliki kepadatan penduduk tertinggi di Indonesia. Kepadatan penduduk yang tinggi mengakibatkan Jawa Timur menghadapi berbagai masalah, salah satunya masalah transportasi. Transportasi merupakan sektor yang penting bagi seluruh lapisan masyarakat. Adanya transportasi yang baik akan berpengaruh terhadap pemerataan ekonomi dan pertumbuhan ekonomi suatu daerah. Namun, ketergantungan masyarakat Jawa Timur terhadap sektor transportasi, ditambah dengan padatnya penduduk di Provinsi Jawa Timur menyebabkan banyaknya masalah transportasi terjadi di provinsi ini. Salah satu masalah transportasi yang marak terjadi adalah kecelakaan lalu lintas. Data BPS menunjukkan bahwa selama tahun 2011-2013 Provinsi Jawa Timur menempati

II. TINJAUAN PUSTAKA

A. Statistika Deskriptif

Statistika deskriptif berfungsi untuk memberikan keterangan-keterangan mengenai suatu data atau kejadian atau fenomena seperti menerangkan keadaan, gejala, atau persoalan sehingga dapat memberikan informasi yang bermanfaat [5]. Statistika deskriptif digunakan untuk mendeskripsikan karakteristik data angka kecelakaan lalu lintas di Provinsi Jawa Timur dan faktor-faktor yang mempengaruhinya serta disajikan dalam bentuk peta tematik sehingga dapat memberikan informasi yang jelas dan mudah dimengerti.

B. Regresi Linier

Metode regresi adalah metode yang digunakan untuk menyatakan pola hubungan antara satu variabel respon dan satu atau lebih variabel prediktor. Regresi

linear merupakan metode yang memodelkan hubungan antara variabel respon (y) dan variabel prediktor ($x_1, x_2, x_3, \dots, x_p$). Model regresi linier untuk p variabel prediktor secara umum ditulis sebagai berikut [6].

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i$$

dengan

y_i : nilai observasi variabel respon ke- i

x_{ik} : nilai observasi variabel prediktor ke- k pada pengamatan ke- i

β_0 : nilai *intercept* model regresi

β_k : koefisien regresi variabel prediktor ke- k

ε_i : error pada pengamatan ke- i dengan asumsi independen, identik, dan berdistribusi normal dengan mean nol dan varians konstan σ^2

Untuk mengetahui variabel prediktor mana yang berpengaruh secara signifikan terhadap variabel respon dilakukan pengujian parameter serentak dan parsial.

1. Uji Serentak

Hipotesis:

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

H_1 : Minimal ada satu $\beta_k \neq 0 ; k = 1, 2, \dots, p$ (1)

Statistik uji:

$$F_{hit} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / p}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-p-1)}$$

Daerah penolakan:

Tolak H_0 jika $F_{hit} > F_{(\alpha; p, n-p-1)}$ atau jika $P_{value} < \alpha$

2. Uji Parsial

Hipotesis:

$H_0: \beta_k = 0$

H_1 : Minimal ada satu $\beta_k \neq 0 ; k = 1, 2, \dots, p$ (2)

Statistik uji:

$$t_{hitung} = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$

Daerah penolakan:

Tolak H_0 jika $|t_{hitung}| > t_{(\frac{\alpha}{2}; n-p-1)}$ atau $P_{value} < \alpha$

C. Pengujian Asumsi Residual Berdistribusi Normal Regresi Linier Berganda

Pengujian asumsi residual memiliki tujuan untuk mengetahui kelayakan suatu model. Salah satu syarat suatu data dapat dilakukan pemodelan dengan menggunakan metode GWR adalah memenuhi asumsi residual berdistribusi Normal. Hipotesis yang digunakan dalam pengujian ini sebagai berikut.

Hipotesis: (3)

$H_0: F_n(y) = F_0(y)$ (residual berdistribusi Normal)

$H_1: F_n(y) \neq F_0(y)$ (residual tidak berdistribusi Normal)

Statistik uji:

$$D = \text{Sup} |F_n(x) - F_0(x)|$$

Daerah Penolakan:

Tolak H_0 jika $D > D_\alpha$ atau jika $P_{value} < \alpha$

D. Heterogenitas Spasial

Heterogenitas spasial merujuk pada adanya keberagaman dalam hubungan secara kewilayahan. Pada hampir setiap kasus dianggap bahwa kita akan mendapatkan hubungan yang berbeda pada setiap lokasi pengamatan. Akibatnya, parameter global yang diduga dari data geografis tidak menggambarkan dengan baik fenomena geografis pada lokasi tertentu [7]. Heterogenitas spasial dapat diidentifikasi menggunakan pengujian *Breusch-Pagan*.

Hipotesis yang digunakan adalah sebagai berikut.

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2$ (tidak terdapat heterogenitas spasial)

H_1 : minimal ada satu $\sigma_i^2 \neq \sigma^2$ (terdapat heterogenitas spasial) (4)

Statistik uji:

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f}$$

dengan elemen vektor \mathbf{f} adalah:

$$f_i = \frac{e_i^2}{\sigma^2} - 1$$

dengan

$\mathbf{f}: (f_1, f_2, \dots, f_n)^T$ dengan $h_i = \frac{e_i^2}{\sigma^2} - 1$

$e_i: y_i - \hat{y}_i$, (\hat{y}_i diperoleh dari metode *Ordinary Least Square* (OLS))

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n}$$

\mathbf{Z} : Matriks berukuran $n \times (k + 1)$ yang berisi vektor yang sudah dinormal standarkan untuk setiap observasi

Daerah penolakan:

Tolak H_0 jika $BP > \chi^2(p)$ atau jika $P_{value} < \alpha$ dengan p adalah banyaknya prediktor.

E. Dependensi Spasial

Adanya dependensi spasial menunjukkan bahwa pengamatan pada lokasi yang satu dipengaruhi oleh pengamatan di lokasi yang lain. Salah satu pengujian yang dapat digunakan untuk mengidentifikasi adanya dependensi spasial yaitu dengan pengujian *Moran's I*.

Hipotesisnya adalah sebagai berikut.

$H_0: \lambda = 0$ (tidak ada dependensi spasial)

$H_1: \lambda \neq 0$ (terdapat dependensi spasial) (5)

Statistik uji:

$$Z_I = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{var}(\hat{I})}}$$

dengan

Z_I : nilai statistik uji indeks *Moran's I*

$E(\hat{I})$: nilai ekspektasi dari indeks *Moran's I*

$\text{var}(\hat{I})$: nilai varians dari indeks *Moran's I*

Daerah penolakan:

Tolak H_0 jika $|Z_I| > Z_{\alpha/2}$ atau $P_{value} < \alpha$

F. Geographically Weighted Regression (GWR)

Model *Geographically Weighted Regression* (GWR) adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai pengamatan regresi yang berbeda-beda. Variabel respon y dalam model GWR diprediksi dengan variabel prediktor yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati. Model GWR dapat ditulis sebagai berikut [6].

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i ; i = 1, 2, \dots, n$$

dengan

y_i : Nilai observasi variabel respon untuk lokasi ke- i

(u_i, v_i) : Menyatakan koordinat letak geografis (*longitude, latitude*) dari lokasi pengamatan ke- i

$\beta_0(u_i, v_i)$: Nilai *intercept* model regresi GWR

$\beta_k(u_i, v_i)$: Koefisien regresi variabel prediktor ke- k pada lokasi pengamatan ke- i

- x_{ik} : Nilai observasi variabel prediktor ke- k pada lokasi pengamatan ke- i
- ε_i : Error pengamatan ke- i yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian konstan σ^2

G. *Estimasi Parameter Model GWR*

Estimasi parameter model GWR dilakukan dengan metode *Weighted Least Square* (WLS) yaitu dengan memberikan penimbang/pembobot yang berbeda pada setiap lokasi pengamatan. Berikut adalah bentuk estimasi parameter dari model GWR untuk setiap lokasi.

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y$$

H. *Pembobotan Model GWR*

Peran pembobot pada model GWR sangat penting karena nilai pembobot ini mewakili letak data observasi satu dengan lainnya. Fungsi kernel digunakan untuk mengestimasi parameter dalam model GWR jika fungsi jarak (w_j) adalah fungsi yang kontinu. Fungsi Kernel yang digunakan dalam penelitian ini, yaitu fungsi *Fixed Gaussian*, fungsi *Fixed Bisquare*, fungsi *Fixed Tricube*, fungsi *Adaptive Gaussian*, fungsi *Adaptive Bisquare* dan fungsi *Adaptive Tricube*.

Pemilihan *bandwidth* optimum menjadi sangat penting karena akan mempengaruhi ketepatan model terhadap data, yaitu mengatur varians dan bias dari model. Salah satu metode yang digunakan untuk menentukan *bandwidth* optimum adalah metode *cross validation* (CV).

I. *Pengujian Model GWR*

Pengujian model GWR terdiri dari dua macam, yaitu uji kesesuaian antara model regresi linear dengan model GWR dan uji parsial model GWR.

1. Uji Kesesuaian Antara Model Regresi Linear dengan Model GWR

Hipotesis:

$$H_0: \beta_k(u_i, v_i) = \beta_k; i = 1, 2, \dots, n; k = 1, 2, \dots, p$$

$$H_1: \beta_k(u_i, v_i) \neq \beta_k$$

Statistik uji:

$$F_{hitung} = \frac{SSE(OLS) - SSE(GWR)/(v)}{SSE(GWR)/(\delta_1)} \quad (6)$$

Daerah penolakan:

$$\text{Tolak } H_0 \text{ jika } F_{hitung} > F_{(1-\alpha; df_1; df_2)}$$

2. Uji Parsial Model GWR

Hipotesis:

$$H_0: \beta_k(u_i, v_i) = 0$$

$$H_1: \beta_k(u_i, v_i) \neq 0; k = 1, 2, \dots, p$$

Statistik uji:

$$T_{hitung} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}_{\sqrt{c_{kk}}}} \quad (7)$$

Daerah penolakan:

$$\text{Tolak } H_0 \text{ jika } |T_{hitung}| > t_{\frac{\alpha}{2}; df} \text{ atau parameter}$$

$\beta_k(u_i, v_i)$ signifikan terhadap model

J. *Angka Kecelakaan Lalu Lintas dan Faktor-faktor yang Diduga Mempengaruhi*

Angka kecelakaan lalu lintas digunakan untuk mengukur tingkat kecelakaan pada satu satuan ruas jalan. Berdasarkan UU No. 22 Tahun 2009 kecelakaan lalu lintas didefinisikan sebagai suatu peristiwa di jalan

yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda. Kecelakaan lalu lintas dapat menyebabkan luka-luka hingga kematian pada manusia. Kecelakaan lalu lintas dapat terjadi karena beberapa faktor, yaitu manusia, kondisi lingkungan, keadaan jalan dan kondisi kendaraan.

Kepadatan penduduk adalah banyaknya jumlah penduduk per satuan luas. Penelitian mengenai kecelakaan lalu lintas dengan pendekatan spasial di kota besar di Indonesia pernah dilakukan dengan menggunakan metode berjenjang tertimbang dan menghasilkan bahwa kepadatan penduduk merupakan salah satu faktor yang memberikan kontribusi terhadap tingkat kecelakaan di kota besar [7].

Sepuluh kecelakaan lalu lintas yang terjadi berasal dari pengemudi yang berada pada rentang usia 18-24 tahun. Hal ini bisa terjadi karena pada usia dewasa muda terdapat sikap tergesa-gesa dan kecerobohan. Selain itu, kelompok umur tersebut merupakan pengemudi pemula dengan tingkat emosi yang belum stabil serta belum berhati-hati dalam mengendarai kendaraannya [8]. Penelitian yang bertujuan menjelaskan hubungan spasial titik rawan kecelakaan bagi para remaja di Houston, Texas pernah dilakukan menggunakan metode *Getis Ord Gi Star* yang menghasilkan pengelompokan daerah-daerah yang menjadi titik rawan kecelakaan bagi para pengendara remaja [9].

Terdapat perbedaan yang signifikan antara pelaku kecelakaan yang pendidikan terakhirnya SD, SMP, SMA atau Perguruan Tinggi yang dipengaruhi oleh sikap disiplin berlalu lintas yang berbeda. Selain itu tingkat pendidikan sangat berpengaruh terhadap program peningkatan pengetahuan secara langsung dan secara tidak langsung terhadap perilaku berkendara [10].

Menurut Dinas Perhubungan Darat, angka kematian akibat kecelakaan lalu lintas pada pria lebih tinggi daripada wanita. Hal ini dikarenakan jenis kelamin wanita sebagai pengguna sepeda motor jumlahnya lebih sedikit dibandingkan dengan jumlah pengguna sepeda motor pria.

Keadaan lingkungan saat terang maupun gelap dapat menjadi salah satu penyebab terjadinya kecelakaan lalu lintas. Waktu kecelakaan terbagi menjadi dua jenis yaitu pada saat terang (pukul 06.00-18.00) dan saat gelap (pukul 18.00-06.00). Hasil penelitian dengan menggunakan metode spasial elipsoidal di Inggris menyimpulkan bahwa keadaan lingkungan berpengaruh secara signifikan terhadap angka kecelakaan lalu lintas yang ditunjukkan dengan banyaknya angka kecelakaan lalu lintas pada waktu terang [11].

Status jalan terbagi menjadi jalan nasional, jalan provinsi, jalan kabupaten/kota, dan jalan desa. Berdasarkan data BPS, diketahui bahwa jalan kabupaten/kota merupakan bagian terbesar yaitu 415.788 kilometer atau 81,85% dari total panjang jalan di Indonesia, namun yang memiliki kondisi baik hanya 39,97%, kondisi sedang 22,08%, rusak 22,33% dan sisanya dalam kondisi rusak berat. Kondisi jalan yang rusak dapat menyebabkan kemacetan bahkan berujung pada kecelakaan lalu lintas [12].

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Kepolisian Daerah Jawa Timur, Dinas Pekerjaan Umum Binamarga Provinsi Jawa Timur dan Badan Pusat Statistik.

B. Variabel Penelitian

Dalam penelitian ini digunakan 6 variabel prediktor dan 1 variabel respon sebagai berikut.

Tabel 1. Sumber Data Penelitian

No	Variabel	Tipe Variabel
1.	Angka kecelakaan lalu lintas (Y)	Kuantitatif
2.	Kepadatan penduduk (X ₁)	Kuantitatif
3.	Persentase usia remaja (X ₂)	Kuantitatif
4.	Persentase kecelakaan terjadi di kawasan jalan kabupaten/kota (X ₃)	Kuantitatif
5.	Persentase pendidikan terakhir pelaku kecelakaan adalah di atas SMP (X ₄)	Kuantitatif
6.	Rasio jenis kelamin (X ₅)	Kuantitatif
7.	Persentase Kecelakaan Terjadi Pada Waktu Gelap (X ₆)	Kuantitatif

C. Langkah Analisis

Berikut adalah langkah-langkah analisis yang digunakan dalam melakukan penelitian ini sesuai dengan tujuan penelitian

1. Mendeskripsikan angka kecelakaan lalu lintas dan faktor yang mempengaruhinya di setiap kabupaten/kota di Jawa Timur
 - a. Mendeskripsikan data dengan menggunakan statistika deskriptif dan peta tematik menjadi 5 kategori (sangat rendah, rendah, sedang, tinggi, sangat tinggi) dengan menggunakan metode *Natural Break* kecuali untuk variabel rasio jenis kelamin hanya dibagi menjadi 2 kategori sebagai gambaran tentang kecelakaan lalu lintas yang terjadi di Jawa Timur dan faktor-faktor yang mempengaruhinya
2. Memodelkan faktor penyebab kecelakaan lalu lintas di Provinsi Jawa Timur dengan menggunakan metode *Geographically Weighted Regression*
 - a. Menguji multikolinearitas pada setiap variabel prediktor dengan menggunakan nilai VIF
 - b. Memodelkan menggunakan regresi *Ordinary Least Square*
 - c. Memeriksa heterogenitas spasial dengan menggunakan statistik uji Breusch-Pagan
 - d. Memeriksa dependensi aspek spasial dengan menggunakan statistik uji Morans'I
 - e. Melakukan pemilihan pembobot terbaik
 - f. Melakukan pemodelan GWR
 - g. Mendapatkan estimasi parameter untuk model GWR pada masing-masing lokasi.
 - h. Melakukan perbandingan model GWR dan OLS.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Angka Kecelakaan Lalu Lintas dan Faktor-faktor yang Diduga Mempengaruhinya.

Sebelum melakukan pemodelan dengan menggunakan metode *Ordinary Least Square* dan *Geographically Weighted Regression*, terlebih dahulu dilakukan analisis secara deskriptif untuk mengetahui karakteristik variabel yang digunakan.

Tabel 2. Statistika Deskriptif Variabel dalam Analisis

Variabel	Rata-rata	Koefisien Varians	Minimum	Maksimum
Y	81,10	103,23	6,90	442,00
X ₁	1817,00	2591,58	387,00	8562,00
X ₂	16,29	0,15	14,48	21,40
X ₃	38,46	12,99	0,00	100,00
X ₄	66,81	5,44	15,45	90,80
X ₅	97,09	0,07	90,65	101,20
X ₆	36,18	1,46	16,48	48,25

Tabel 2 menunjukkan bahwa rata-rata kecelakaan lalu lintas di Jawa Timur cukup tinggi yaitu setiap 100 kilometer terjadi rata-rata 81,10 atau mendekati 82 kecelakaan lalu lintas pada tahun 2014.



Gambar 1. Persebaran Angka Kecelakaan Lalu Lintas di Jawa Timur Tahun 2014 dan Faktor yang Mempengaruhinya

Gambar 1 menunjukkan pembagian angka kecelakaan lalu lintas di Jawa Timur menjadi 5 kategori yaitu sangat rendah, rendah, sedang, tinggi dan sangat tinggi. Berdasarkan Gambar 1 secara visual dapat terlihat bahwa angka kecelakaan lalu lintas menunjukkan pola mengelompok. Daerah-daerah dengan kategori angka kecelakaan lalu lintas sedang hingga tinggi cenderung berada di barat laut Provinsi Jawa Timur. Sementara dengan kategori angka kecelakaan sangat rendah hingga rendah saling berdekatan.

B. Pengujian Asumsi Residual Berdistribusi Normal

Pengujian asumsi residual Normal digunakan untuk menguji apakah dalam model regresi, data memiliki distribusi Normal. Dalam penelitian dengan metode GWR, diperlukan data yang memenuhi asumsi residual Normal. Berdasarkan uji Kolmogorov Smirnov, diperoleh *P-value* sebesar $< 0,010$ atau tolak H_0 yang berarti residual tidak berdistribusi normal. Oleh karena itu dilakukan transformasi terhadap variabel respon menjadi $\ln(y)$. Setelah dilakukan transformasi terhadap variabel respon, diperoleh *P-value Kolmogorov Smirnov* sebesar $> 0,150$. Nilai ini lebih besar dari nilai α sebesar 0,05 sehingga diputuskan gagal tolak H_0 atau data yang digunakan memenuhi asumsi Residual berdistribusi Normal.

C. Pemodelan Regresi Ordinary Least Square (OLS) Angka Kecelakaan Lalu Lintas

Sebelum melakukan pemodelan, terlebih dahulu dilakukan deteksi multikolinearitas. Kriteria terjadi kasus multikolinearitas pada variabel prediktor dapat dilihat dari nilai VIF.

Tabel 3. Nilai VIF Variabel Prediktor

Variabel	VIF
X ₁	1,27
X ₂	1,49
X ₃	1,22
X ₄	1,80
X ₅	1,52
X ₆	1,33

Tabel 3 menunjukkan nilai VIF yang kurang dari 10 artinya tidak terdapat autokorelasi. Diperoleh model regresi OLS sebagai berikut.

$$\ln(y) = 3,62 + 0,000179 x_1 + 0,101 x_2 - 0,00607 x_3 + 0,0247x_4 - 0,0418 x_5 + 0,0288 x_6$$

Untuk memudahkan dalam interpretasi, maka pemodelan dapat dituliskan sebagai berikut.

$$\hat{Y} = e^{(g(x))}$$

dimana:

$$g(x) = 3,62 + 0,000179 x_1 + 0,101 x_2 - 0,00607 x_3 + 0,0247x_4 - 0,0418 x_5 + 0,0288 x_6$$

Selanjutnya dilakukan pengujian signifikansi parameter secara serentak. Berdasarkan analisis diperoleh P-value sebesar 0,001 atau kurang dari nilai α (0,05) sehingga merujuk pada persamaan (1) disimpulkan tolak H₀ yang artinya paling sedikit ada satu variabel prediktor yang memiliki pengaruh signifikan terhadap variabel respon.

Untuk mengetahui variabel mana saja yang berpengaruh secara signifikan terhadap variabel respon, dilakukan pengujian signifikansi parameter secara parsial. Hasil uji signifikansi dapat dilihat pada Tabel 4.

Tabel 4. Uji Signifikansi Parameter Secara Parsial

Variabel	P-value
X ₁	0,006
X ₂	0,273
X ₃	0,296
X ₄	0,005
X ₅	0,459
X ₆	0,125

Berdasarkan Tabel 4, dengan taraf signifikansi 0,05 diketahui bahwa variabel yang berpengaruh secara signifikan terhadap angka kecelakaan lalu lintas atau tolak H₀ berdasarkan persamaan (2) adalah X₁ (kepadatan penduduk) dan X₄ (persentase pendidikan terakhir pelaku kecelakaan di atas SMP).

D. Pengujian Heterogenitas Spasial

Diperoleh P-value pengujian *Breusch Pagan* sebesar 0,008775. Dengan menggunakan taraf signifikansi sebesar 0,05 maka diputuskan tolak H₀ atau berdasarkan persamaan (4) terdapat heterogenitas spasial pada data yang diamati.

E. Pengujian Dependensi Spasial

Pengujian *Moran's I* merupakan pengujian yang dilakukan untuk melihat apakah pengamatan di suatu lokasi berpengaruh terhadap pengamatan di lokasi lain yang letaknya saling berdekatan. Diketahui bahwa pada penelitian ini diperoleh p-value sebesar 0,03 atau kurang dari α sebesar 0,05 sehingga dapat disimpulkan bahwa tolak H₀ atau terdapat dependensi spasial pada pengamatan.

F. Pemodelan GWR

Setelah aspek dependensi spasial dan heterogenitas spasial terpenuhi, selanjutnya dilakukan pemilihan pembobot optimum yang akan dimasukkan dalam pemodelan GWR.

Tabel 5 Pemilihan Pembobot Optimum

Fungsi Pembobot	CV minimum	Bandwidth
<i>Fixed Gaussian</i>	24,03484	3,551274
<i>Adaptive Gaussian</i>	23,88938	0,7368411
<i>Fixed Bisquare</i>	23,71636	1,673124
<i>Adaptive Bisquare</i>	24,89006	0,7631708
<i>Fixed Tricube</i>	23,81214	1,714153
<i>Adaptive Tricube</i>	25,28998	0,76316

Salah satu kriteria pembobot terbaik adalah dengan meminimumkan nilai CV. Berdasarkan Tabel 5, dapat diketahui bahwa nilai CV paling minimum diperoleh dengan menggunakan fungsi pembobot *Fixed Bisquare*, yaitu dengan CV minimum sebesar 23,71636 dan *bandwidth* sebesar 1,673124.

Setelah ditentukan pembobot yang digunakan, selanjutnya dilakukan pemodelan dengan metode GWR. Tabel 6 menunjukkan hasil estimator pada model GWR berupa *Bandwidth*, R² dan SSE.

Tabel 6. Hasil Estimator Model GWR

Bandwidth	1,673124
R²	52,7959 persen
SSE	14,22316

Informasi yang didapatkan dari Tabel 6 yaitu nilai koefisien determinasi (R²) model GWR sebesar 52,7959 persen yang berarti model yang terbentuk dapat menjelaskan variabel angka kecelakaan lalu lintas sebesar 52,7959 persen dan sisanya dijelaskan oleh variabel lain yang tidak diamati dalam model.

G. Pengujian Kesesuaian Model GWR

Berdasarkan penghitungan pada persamaan (8), diperoleh nilai F_{hitung} sebesar 0,88984 yang lebih kecil dibanding F_{tabel(0,05;2,4103;28,5897)}=3,34 sehingga dapat disimpulkan gagal tolak H₀ artinya tidak ada perbedaan yang signifikan antara model regresi OLS dan GWR

H. Pengujian Signifikansi Parameter Model GWR

Pengujian signifikansi parameter model GWR secara parsial dilakukan untuk mengetahui parameter mana saja yang berpengaruh secara signifikan terhadap angka kecelakaan lalu lintas di setiap kabupaten/kota di Jawa Timur. Hipotesis yang digunakan dalam pengujian model GWR secara parsial adalah sebagai berikut.

$$H_0: \beta_k(u_i, v_i) = 0$$

$$H_1: \beta_k(u_i, v_i) \neq 0; k = 1, 2, \dots, 7$$

Jika nilai $|t_{hitung}| > t_{tabel} = t_{(0,05;28,5897)} = 2,048407$ maka parameter signifikan. Berdasarkan uji parsial, diperoleh bahwa variabel yang signifikan untuk seluruh kabupaten/kota di Jawa Timur dengan taraf signifikansi sebesar 5 persen adalah variabel X₁ dan X₄.

I. Interpretasi Model GWR

Pada pemodelan GWR, daerah satu memiliki model berbeda dengan model lainnya. Pemodelan tiap kabupaten/kota di Jawa Timur dapat dilihat pada Lampiran. Pada pembahasan ini dilakukan interpretasi model Kota Surabaya sebagai berikut.

$$\ln \hat{Y} = 3,3789 + 0,000174 X_1 + 0,1011 X_2 - 0,0066 X_3 + 0,0216 X_4 - 0,0061 X_5 + 0,0245 X_6$$

Untuk memudahkan dalam interpretasi, maka pemodelan dapat dituliskan sebagai berikut.

$$\hat{Y} = e^{g(x)}$$

$$g(x) = 3,3789 + 0,000174 X_1 + 0,1011 X_2 - 0,0066 X_3 + 0,0216 X_4 - 0,0061 X_5 + 0,0245 X_6$$

Penjelasan yang diperoleh berdasarkan model di atas untuk Kota Surabaya, yaitu untuk setiap kenaikan kepadatan penduduk sebesar 1 jiwa/km², maka akan menaikkan angka kecelakaan lalu lintas/ 100 kilometer sebesar $e^{0,000174} = 1$ kejadian kecelakaan lalu lintas dengan asumsi variabel lain tetap. Untuk setiap kenaikan persentase usia remaja sebesar 1 persen maka akan menaikkan angka kecelakaan lalu lintas/ 100 kilometer sebesar $e^{0,1011} = 1,10 \approx 1$ kejadian kecelakaan lalu lintas dengan asumsi variabel lain tetap. Angka kecelakaan lalu lintas/ 100 kilometer akan turun sebesar $e^{-0,0066} = 1,0066 \approx 1$ kejadian jika persentase kecelakaan terjadi di kawasan jalan kabupaten/kota naik sebesar 1 persen dengan asumsi variabel lain tetap. Kenaikan persentase pendidikan terakhir pelaku kecelakaan lebih dari SMP sebesar 1 persen akan meningkatkan kecelakaan lalu lintas/100 kilometer sebesar $e^{0,0216} = 1,02 \approx 1$ kejadian kecelakaan dengan asumsi variabel lain tetap. Angka kecelakaan lalu lintas/ 100 kilometer akan turun sebesar $e^{-0,0061} = 1,0061 \approx 1$ jika rasio jenis kelamin naik sebesar 1 persen dengan asumsi variabel lain tetap. Angka kecelakaan lalu lintas/ 100 kilometer akan naik sebesar $e^{0,0245} = 1,02 \approx 1$ jika persentase kecelakaan terjadi saat waktu gelap naik sebesar 1 persen.

J. Pemilihan Model Terbaik

Pemilihan model terbaik merupakan proses evaluasi dari model untuk mengetahui seberapa besar peluang masing-masing model yang terbentuk sudah sesuai dengan data. Pada penelitian ini dibandingkan model OLS dan GWR berdasarkan kriteria R² dan SSE.

Tabel 7. Perbandingan Model OLS dan GWR

Kriteria	Model OLS	Model GWR
R ²	36,66 persen	52,7959 persen
SSE	15,2898	14,22316

Berdasarkan Tabel 7 diperoleh informasi bahwa berdasarkan nilai R² dan nilai SSE, model GWR lebih baik dibandingkan model OLS. Model GWR terbukti mampu meningkatkan nilai R² dan menurunkan nilai *Sum Square Error*.

V. KESIMPULAN DAN SARAN

Hasil analisis menunjukkan karakteristik angka kecelakaan lalu lintas mengelompok sesuai kategori. Faktor-faktor yang secara signifikan mempengaruhi angka kecelakaan lalu lintas di setiap kabupaten/kota di Jawa Timur secara spasial dengan taraf signifikansi sebesar 5% adalah kepadatan penduduk dan persentase pendidikan terakhir pelaku kecelakaan lebih dari SMP. Pemodelan dengan GWR meningkatkan nilai R² dan diperoleh SSE yang lebih kecil dibanding metode OLS. Dengan menggunakan taraf signifikansi sebesar 5%, diperoleh model yang sama untuk setiap kabupaten/kota di Jawa Timur.

Sebaiknya pemerintah daerah berfokus kepada penanggulangan kepadatan penduduk sehingga dapat menekan angka kecelakaan lalu lintas. Selain itu, pendidikan terakhir pelaku kecelakaan terbukti signifikan

terhadap peningkatan angka kecelakaan lalu lintas. Padahal penelitian ini dilakukan kepada pelaku kecelakaan yang memiliki pendidikan terakhir di atas SMP atau kebanyakan memiliki surat izin mengemudi. Hal ini menunjukkan bahwa kualitas kepemilikan SIM masih rendah sehingga perlu dilakukan pengetatan terhadap pemberian SIM oleh pihak berwajib.

Untuk penelitian selanjutnya diharapkan menambah faktor-faktor lain sehingga lebih menggambarkan kondisi yang mempengaruhi angka kecelakaan lalu lintas di Jawa Timur, selain itu bisa dilakukan pendekatan area untuk memodelkan angka kecelakaan lalu lintas secara spasial.

DAFTAR PUSTAKA

- [1] Warpani, S. 2002. *Pengelolaan Lalu Lintas dan Angkutan Jalan*. Bandung: Penerbit ITB.
- [2] Wijayanti, L.A. 2014. *Pemanfaatan Citra Quickbird dan Sistem Informasi Geografis Untuk Identifikasi Daerah Rawan Kecelakaan, Daerah Kajian: Kota Semarang, Jawa Tengah*. Yogyakarta: Universitas Gadjah Mada.
- [3] Erdogan, Saffet. 2009. Exploative Spatial Analysis of Traffic Accident Statistics and Road Mortality among The Provinces of Turkey. *Journal of Safety Research*. Vol 40: hal 341-351.
- [4] Zheng, L., Robinson, R.M., Khattak, A. & Wang, X. 2011. *All Accidents are Not Equal: Using Geographically Weighted Regressions Models to Assess and Forecast Accident Impacts*, Indianapolis, USA: 3rd International Conference on Road Safety and Simulation
- [5] Walpole, R.E. 1993. *Pengantar Statistika Edisi ke-3*. Jakarta: PT Gramedia Pustaka Utama
- [6] Fotheringham, A.S., Brundson, C., & Charlton, M. 2002. *Geographically Weighted Regression*. Chichester, UK: Jhon Wiley & Sons.
- [7] Gumawang, J. 2015. *Kajian Tingkat Kerawanan Kecelakaan Lalu Lintas Sebagian Ruas Jalan di Kota Tangerang Secara Spasial*. Yogyakarta: Universitas Gadjah Mada.
- [8] Goodwin, G.C., Schoby, J., & Council W. 2014. *A Hot Spot Analysis of Teenage Crashes: An Assessment of Crashes in Houston, Texas*. Texas: Texas Southern University.
- [9] Wahyuningtyas, F. 2013. *Perbedaan Sikap Disiplin Berlalu Lintas Ditinjau dari Tingkat Pendidikan*. Surakarta: Universitas Muhammadiyah Surakarta.
- [10] Wedagama, D.M.P. 2010. *Metode Elipsoidal Kesalahan Sebagai Metode Alternatif untuk Analisis Lokasi Rawan Kecelakaan Lalu Lintas*. Jurnal Ilmiah Teknik Sipil Vol. 14 No.1.
- [11] Soediono, W. & Handoko, D. 2004. Pemodelan dan Simulasi Antrian Kendaraan di Gerbang Tol. *Prosiding Semiloka Teknologi Simulasi dan Komputasi serta Aplikasi 2004*.

LAMPIRAN

Kabupaten/Kota	Variabel Signifikan	Model GWR
Pacitan	X ₁ ,X ₄	$\ln \hat{Y} = 0,4862 + 0,000193 X_1 + 0,1011 X_2 - 0,0066 X_3 + 0,0216 X_4 - 0,0061 X_5 + 0,0245 X_6$
Ponorogo	X ₁ ,X ₄	$\ln \hat{Y} = 1,8906 + 0,000185 X_1 + 0,1064 X_2 - 0,0068 X_3 + 0,0243 X_4 - 0,0236 X_5 + 0,0259 X_6$
Trenggalek	X ₁ ,X ₄	$\ln \hat{Y} = 1,1595 + 0,000187 X_1 + 0,1025 X_2 - 0,0066 X_3 + 0,0228 X_4 - 0,0140 X_5 + 0,0247 X_6$
Tulungagung	X ₁ ,X ₄	$\ln \hat{Y} = 1,3301 + 0,000185 X_1 + 0,1025 X_2 - 0,0066 X_3 + 0,0230 X_4 - 0,0159 X_5 + 0,0247 X_6$
Lumajang	X ₁ ,X ₄	$\ln \hat{Y} = 2,1167 + 0,000176 X_1 + 0,1025 X_2 - 0,0065 X_3 + 0,0241 X_4 - 0,0246 X_5 + 0,0248 X_6$
Bondowoso	X ₁ ,X ₄	$\ln \hat{Y} = 2,6926 + 0,000177 X_1 + 0,1069 X_2 - 0,0067 X_3 + 0,0255 X_4 - 0,0326 X_5 + 0,0260 X_6$
Pasuruan	X ₁ ,X ₄	$\ln \hat{Y} = 3,1575 + 0,000170 X_1 + 0,1054 X_2 - 0,0065 X_3 + 0,0259 X_4 - 0,0373 X_5 + 0,0259 X_6$
Jombang	X ₁ ,X ₄	$\ln \hat{Y} = 4,0588 + 0,000154 X_1 + 0,0989 X_2 - 0,0059 X_3 + 0,0260 X_4 - 0,0455 X_5 + 0,0260 X_6$
Nganjuk	X ₁ ,X ₄	$\ln \hat{Y} = 4,3349 + 0,000150 X_1 + 0,0974 X_2 - 0,0058 X_3 + 0,0260 X_4 - 0,0482 X_5 + 0,0263 X_6$
Madiun	X ₁ ,X ₄	$\ln \hat{Y} = 6,0579 + 0,000134 X_1 + 0,0934 X_2 - 0,0051 X_3 + 0,0266 X_4 - 0,0665 X_5 + 0,0290 X_6$
Magetan	X ₁ ,X ₄	$\ln \hat{Y} = 5,1798 + 0,000148 X_1 + 0,1014 X_2 - 0,0058 X_3 + 0,0275 X_4 - 0,0589 X_5 + 0,0273 X_6$
Ngawi	X ₁ ,X ₄	$\ln \hat{Y} = 5,3983 + 0,000147 X_1 + 0,1019 X_2 - 0,0057 X_3 + 0,0278 X_4 - 0,0615 X_5 + 0,0275 X_6$
Bojonegoro	X ₁ ,X ₄	$\ln \hat{Y} = 4,1308 + 0,000159 X_1 + 0,1039 X_2 - 0,0062 X_3 + 0,0268 X_4 - 0,0478 X_5 + 0,0264 X_6$
Tuban	X ₁ ,X ₄	$\ln \hat{Y} = 3,9015 + 0,000163 X_1 + 0,1054 X_2 - 0,0063 X_3 + 0,0268 X_4 - 0,0457 X_5 + 0,0264 X_6$
Lamongan	X ₁ ,X ₄	$\ln \hat{Y} = 3,5252 + 0,000170 X_1 + 0,1080 X_2 - 0,0066 X_3 + 0,0267 X_4 - 0,0423 X_5 + 0,0266 X_6$
Bangkalan	X ₁ ,X ₄	$\ln \hat{Y} = 3,2453 + 0,000172 X_1 + 0,1078 X_2 - 0,0066 X_3 + 0,0263 X_4 - 0,0391 X_5 + 0,0264 X_6$
Pamekasan	X ₁ ,X ₄	$\ln \hat{Y} = 2,9943 + 0,000175 X_1 + 0,1080 X_2 - 0,0067 X_3 + 0,0260 X_4 - 0,0364 X_5 + 0,0264 X_6$
Kota Kediri	X ₁ ,X ₄	$\ln \hat{Y} = 1,9765 + 0,000186 X_1 + 0,1075 X_2 - 0,0069 X_3 + 0,0246 X_4 - 0,0250 X_5 + 0,0263 X_6$
Kota Blitar	X ₁ ,X ₄	$\ln \hat{Y} = 1,5655 + 0,000191 X_1 + 0,1071 X_2 - 0,0070 X_3 + 0,0240 X_4 - 0,0204 X_5 + 0,0265 X_6$
Kota Malang	X ₁ ,X ₄	$\ln \hat{Y} = 1,4076 + 0,000192 X_1 + 0,1065 X_2 - 0,0070 X_3 + 0,0237 X_4 - 0,0185 X_5 + 0,0264 X_6$
Kota Probolinggo	X ₁ ,X ₄	$\ln \hat{Y} = 1,6600 + 0,000191 X_1 + 0,1079 X_2 - 0,0071 X_3 + 0,0242 X_4 - 0,0217 X_5 + 0,0268 X_6$
Kota Pasuruan	X ₁ ,X ₄	$\ln \hat{Y} = 2,2254 + 0,000186 X_1 + 0,1096 X_2 - 0,0071 X_3 + 0,0252 X_4 - 0,0286 X_5 + 0,0271 X_6$
Kota Mojokerto	X ₁ ,X ₄	$\ln \hat{Y} = 3,4855 + 0,000180 X_1 + 0,1148 X_2 - 0,0071 X_3 + 0,0275 X_4 - 0,0444 X_5 + 0,0285 X_6$
Kota Madiun	X ₁ ,X ₄	$\ln \hat{Y} = 3,3961 + 0,000174 X_1 + 0,1100 X_2 - 0,0067 X_3 + 0,0268 X_4 - 0,0416 X_5 + 0,0270 X_6$
Kota Surabaya	X ₁ ,X ₄	$\ln \hat{Y} = 3,3789 + 0,000174 X_1 + 0,1098 X_2 - 0,0067 X_3 + 0,0267 X_4 - 0,0413 X_5 + 0,0270 X_6$
Kota Batu	X ₁ ,X ₄	$\ln \hat{Y} = 3,7727 + 0,000170 X_1 + 0,1101 X_2 - 0,0066 X_3 + 0,0272 X_4 - 0,0458 X_5 + 0,0271 X_6$
Blitar	X ₁ ,X ₄	$\ln \hat{Y} = 4,9136 + 0,000156 X_1 + 0,1067 X_2 - 0,0061 X_3 + 0,0280 X_4 - 0,0575 X_5 + 0,0274 X_6$
Kediri	X ₁ ,X ₄	$\ln \hat{Y} = 5,1604 + 0,000154 X_1 + 0,1062 X_2 - 0,0060 X_3 + 0,0282 X_4 - 0,0602 X_5 + 0,0275 X_6$
Mojokerto	X ₁ ,X ₄	$\ln \hat{Y} = 5,6522 + 0,000150 X_1 + 0,1057 X_2 - 0,0058 X_3 + 0,0286 X_4 - 0,0656 X_5 + 0,0279 X_6$

Banyuwangi	X ₁ ,X ₄	$\ln \hat{Y} = 2,6245 + 0,000177 X_1 + 0,1067 X_2 - 0,0067 X_3 + 0,0254 X_4 - 0,0318 X_5 + 0,0260 X_6$
Gresik	X ₁ ,X ₄	$\ln \hat{Y} = 2,2605 + 0,000174 X_1 + 0,1023 X_2 - 0,0064 X_3 + 0,0243 X_4 - 0,0262 X_5 + 0,0248 X_6$
Jember	X ₁ ,X ₄	$\ln \hat{Y} = 3,1836 + 0,000170 X_1 + 0,1055 X_2 - 0,0065 X_3 + 0,0259 X_4 - 0,0377 X_5 + 0,0259 X_6$
Malang	X ₁ ,X ₄	$\ln \hat{Y} = 4,5878 + 0,000156 X_1 + 0,1040 X_2 - 0,0061 X_3 + 0,0273 X_4 - 0,0530 X_5 + 0,0268 X_6$
Probolinggo	X ₁ ,X ₄	$\ln \hat{Y} = 3,6341 + 0,000167 X_1 + 0,1068 X_2 - 0,0065 X_3 + 0,0266 X_4 - 0,0431 X_5 + 0,0264 X_6$
Sampang	X ₁ ,X ₄	$\ln \hat{Y} = 3,2301 + 0,000173 X_1 + 0,1082 X_2 - 0,0066 X_3 + 0,0263 X_4 - 0,0391 X_5 + 0,0265 X_6$
Sidoarjo	X ₁ ,X ₄	$\ln \hat{Y} = 1,6003 + 0,000190 X_1 + 0,1070 X_2 - 0,0070 X_3 + 0,0240 X_4 - 0,0207 X_5 + 0,0264 X_6$
Situbondo	X ₁ ,X ₄	$\ln \hat{Y} = 3,6709 + 0,000170 X_1 + 0,1090 X_2 - 0,0066 X_3 + 0,0270 X_4 - 0,0443 X_5 + 0,0269 X_6$
Sumenep	X ₁ ,X ₄	$\ln \hat{Y} = 3,1272 + 0,000171 X_1 + 0,1062 X_2 - 0,0065 X_3 + 0,0259 X_4 - 0,0373 X_5 + 0,0260 X_6$