

BAB I PENDAHULUAN

1.1 Latar Belakang

Regresi spasial merupakan pendekatan statistik yang mempertimbangkan letak geografis setiap pengamatan, sehingga hubungan antarvariabel dapat dianalisis dengan memperhatikan kedekatan lokasi (Sanusi dkk., 2019). Perbedaan kondisi sosial, ekonomi, dan lingkungan antarwilayah seringkali menyebabkan pengaruh variabel independen terhadap variabel dependen tidak bersifat seragam yang disebut dengan heterogenitas spasial (Majid dan Pasaribu, 2019). *Geographically Weighted Regression* (GWR) merupakan metode analisis spasial yang memberikan bobot berbeda bagi setiap lokasi amatan sehingga setiap lokasi memiliki koefisien regresi yang berbeda sesuai karakteristik wilayahnya (Eliezer, 2024). Meskipun model GWR mampu menangkap variasi antarwilayah, karakteristik suatu daerah tidak bersifat tetap setiap waktu. Perubahan kondisi lingkungan, penerapan kebijakan tertentu, serta dinamika sosial masyarakat dapat menyebabkan hubungan antarvariabel mengalami perubahan pada periode waktu yang berbeda. Perbedaan pengaruh yang muncul akibat faktor waktu disebut sebagai heterogenitas temporal (Trezenki dkk., 2025).

Geographically and Temporally Weighted Regression (GTWR) merupakan pengembangan dari GWR yang mampu menangkap perbedaan baik dalam spasial maupun temporal. Keunggulan penerapan GTWR pada penelitian sebelumnya yaitu pemodelan Indeks Pembangunan Manusia (IPM) di Provinsi Bengkulu menunjukkan bahwa GTWR memberikan performa estimasi yang lebih unggul dalam menangkap struktur heterogenitas spasial-temporal dengan lebih efektif dibandingkan GWR. Hal ini ditunjukkan oleh nilai koefisien determinasi R^2 pada model GTWR sebesar 99,98%, lebih tinggi dari GWR dengan nilai R^2 99,74% (Oktarina dkk., 2024). Meskipun GTWR sesuai untuk memetakan perbedaan pengaruh antarwilayah dan antarwaktu, karakteristik modelnya berorientasi pada variabel dependen yang bersifat kontinu sehingga penerapannya perlu disesuaikan ketika jenis data memiliki struktur distribusi yang berbeda (Widayaka dkk., 2016).

Data kesehatan masyarakat umumnya dicatat sebagai jumlah kejadian (*count data*), sehingga regresi Poisson merupakan salah satu pendekatan statistik yang digunakan untuk menganalisis hubungan antara variabel dependen berupa jumlah kasus dan sejumlah variabel independen yang diduga berpengaruh (Schober dan Vetter, 2021). Sering kali, pada data kesehatan masyarakat, varians observasi melampaui nilai rata-ratanya, sehingga muncul kondisi overdispersi (Menssen dan Rathjens, 2025). Ketidaksesuaian ini menyebabkan asumsi dasar Poisson tidak terpenuhi dan menjadikan model tersebut kurang akurat dalam menghasilkan estimasi parameter. *Negative Binomial Regression* (NBR) merupakan perluasan dari model regresi Poisson dengan menambahkan parameter dispersi untuk menangani varians data yang tinggi serta menghasilkan estimasi parameter yang lebih stabil (Fatmala dkk., 2024). Meski demikian, pendekatan global pada model NBR tidak memperhitungkan perbedaan karakter antarwilayah. Oleh karena itu, NBR

dikembangkan lebih lanjut menjadi *Geographically Weighted Negative Binomial Regression* (GWNBR), sehingga estimasi parameter dapat disesuaikan dengan kondisi masing-masing wilayah. Berdasarkan hasil evaluasi model dalam pemodelan faktor risiko penyakit pneumonia pada balita di Provinsi Sulawesi Tengah, GWNBR mampu menghasilkan parameter regresi yang berbeda di setiap lokasi sesuai kondisi wilayah yang di teliti, sehingga heterogenitas spasial pada kasus pneumonia dapat tertangani (Mar'ah dkk., 2023).

Penerapan GWNBR sebagai metode yang mampu menangkap perbedaan karakteristik antarwilayah melalui pembobot spasial seringkali menghadapi masalah multikolinearitas apabila dalam analisis melibatkan banyak variabel independen. Multikolinearitas merupakan kondisi ketika terdapat korelasi tinggi antarvariabel independen yang menyebabkan ketidakstabilan estimasi parameter dan menurunkan akurasi hasil pemodelan (Andini dkk., 2025). *Principal Component Analysis* (PCA) merupakan metode yang digunakan untuk mereduksi variabel yang saling berkorelasi menjadi beberapa komponen utama yang representatif (Chen dkk., 2025). Penelitian yang dilakukan oleh Bavar dkk. (2023) menerapkan kombinasi metode PCA dan GWNBR untuk menganalisis pengaruh faktor-faktor lingkungan terhadap frekuensi kecelakaan lalu lintas di wilayah perkotaan Shiraz, Iran. Hasil penelitian tersebut menunjukkan bahwa penerapan PCA sebelum pemodelan spasial dapat menurunkan efek kolinearitas dan meningkatkan kestabilan estimasi parameter pada model GWNBR.

Pendekatan GWNBR-PCA sesuai digunakan untuk menganalisis kasus pneumonia pada balita karena mampu menangkap perbedaan pengaruh faktor-faktor risiko antarwilayah sekaligus mengatasi masalah multikolinearitas yang sering muncul pada data kesehatan masyarakat yang melibatkan banyak variabel independen. Faktor penentu risiko pneumonia seperti kepadatan penduduk, status gizi balita, cakupan imunisasi, kualitas udara, dan sanitasi lingkungan umumnya memiliki hubungan yang saling berkorelasi, sehingga perubahan pada satu variabel sering beriringan dengan perubahan pada variabel lainnya. Kondisi ini dapat menimbulkan multikolinearitas dalam pemodelan regresi yang berdampak pada ketidakstabilan estimasi parameter dan menurunkan akurasi hasil prediksi.

Penyebaran kasus pneumonia balita di Indonesia menunjukkan keragaman spasial. Wilayah dengan kepadatan penduduk tinggi seperti Pulau Jawa cenderung memiliki risiko penularan yang lebih besar. Selain itu, faktor kerentanan seperti status gizi buruk dan berat badan lahir rendah berperan dalam menentukan kekuatan sistem imunitas balita, sementara intervensi kesehatan seperti pemberian ASI eksklusif, vitamin A, dan cakupan imunisasi memberikan perlindungan terhadap risiko infeksi (Adawiyah, R. dan Duarsa, 2016). Faktor lingkungan seperti kualitas udara dan sanitasi layak juga turut memengaruhi tingkat paparan balita terhadap sumber penyakit (Suryati dkk., 2018). Namun, pola kejadian pneumonia balita dapat berubah antarwaktu akibat perubahan kondisi lingkungan, perbedaan musim, mobilitas penduduk, serta kebijakan intervensi kesehatan yang diterapkan pada waktu tertentu, sehingga penyebarannya tidak hanya dipengaruhi oleh perbedaan karakteristik antarwilayah tetapi juga oleh dinamika temporal. Berdasarkan uraian

sebelumnya, penelitian ini akan mengembangkan model GWNBR-PCA dengan menambahkan dimensi waktu (temporal), sehingga memperoleh model *Geographically and Temporally Weighted Negative Binomial Regression Principal Component Analysis* (GTWNBR-PCA) pada jumlah kasus pneumonia balita di Provinsi Jawa Timur beserta faktor-faktor yang memengaruhinya dalam rentang waktu tahun 2021 hingga 2024.

1.2 Batasan Masalah

Berdasarkan masalah yang telah diuraikan sebelumnya, berikut batasan masalah pada penelitian ini.

1. Data yang digunakan merupakan data temporal mengenai kasus pneumonia pada balita di Provinsi Jawa Timur beserta faktor-faktor yang memengaruhinya dalam rentang waktu tahun 2021 hingga 2024.
2. Penentuan jarak antar lokasi pengamatan menggunakan jarak *euclidean* dengan fungsi pembobot *Fixed Kernel*.
3. Kriteria pemilihan komponen utama dalam *Principal Component Analysis* dibatasi pada proporsi keragaman kumulatif yang mampu menjelaskan data sebesar 80%.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini ialah sebagai berikut :

1. Memperoleh model GTWNBR-PCA pada data jumlah kasus pneumonia balita di Provinsi Jawa Timur.
2. Memperoleh faktor-faktor yang memengaruhi jumlah kasus pneumonia pada balita di Provinsi Jawa Timur dalam rentang waktu tahun 2021 hingga 2024.

Manfaat yang dapat diperoleh dari penelitian ini ialah sebagai berikut

1. Memberikan informasi mengenai kondisi penyebaran kasus pneumonia balita di Provinsi Jawa Timur sejak tahun 2021 hingga 2024.
2. Menjadi referensi bagi peneliti selanjutnya yang ingin mengembangkan pemodelan spasial-temporal pada data diskrit dengan karakteristik data serupa.
3. Menjadi masukan bagi pemerintah dan instansi kesehatan dalam menetapkan prioritas intervensi dan kebijakan penanganan pneumonia berdasarkan perbedaan kebutuhan antarwilayah dan periode waktu.

1.4 Landasan Teori

1.4.1 Regresi Linear Berganda

Regresi linear berganda merupakan pengembangan regresi linear yang digunakan untuk menganalisis pengaruh beberapa variabel independen terhadap satu variabel dependen dengan asumsi bahwa hubungan antarvariabel tersebut bersifat linear. Model ini digunakan untuk memperoleh estimasi parameter yang menggambarkan kontribusi masing-masing variabel independen (Siagian, 2023). Bentuk umum yang digunakan dalam regresi linear berganda ditunjukkan pada Persamaan (1).

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (1)$$

$i = 1, 2, \dots, n; k = 1, 2, 3, \dots, p$

dengan

- y_i : nilai observasi variabel dependen ke- i
- X_{ki} : nilai observasi variabel independen ke- k pada amatan ke- i
- β_0 : nilai konstanta model regresi
- $\beta_1, \beta_2, \dots, \beta_k$: koefisien regresi variabel independen ke- k
- ε_i : residual pada amatan ke- i

1.4.2 Distribusi dan Regresi Poisson

Distribusi Poisson merupakan distribusi peluang untuk data diskrit yang menggambarkan probabilitas banyaknya kejadian dalam suatu periode waktu tertentu ketika nilai mean kejadiannya telah diketahui, dengan asumsi bahwa setiap peristiwa terjadi secara independen dari peristiwa sebelumnya (Pahdian dkk., 2021). Distribusi peluang variabel acak Poisson x yang menyatakan jumlah kejadian dalam suatu interval waktu atau wilayah tertentu t dapat dituliskan sebagai Persamaan (2).

$$p(x; \mu t) = \frac{e^{-\mu t} (\mu t)^x}{x!}; x = 1, 2, \dots, n \quad (2)$$

dengan μt menyatakan mean frekuensi terjadinya suatu peristiwa pada tiap satuan waktu atau daerah. Jika seluruh selang waktu kejadian sama, maka fungsi peluang untuk variabel acak Poisson y dengan parameter μ dapat dinyatakan sebagai Persamaan (3).

$$p(y; \mu) = \frac{[\exp(-\mu)](\mu)^y}{y!}, y = 0, 1, 2, \dots, n; \mu > 0 \quad (3)$$

Regresi Poisson merupakan model regresi yang didasarkan pada distribusi Poisson dan digunakan untuk menganalisis hubungan antara variabel dependen berbentuk data diskrit dengan satu atau lebih variabel independen (Tendriyawati dkk., 2023). Model ini mengasumsikan bahwa variabel dependen memiliki nilai rata-rata yang sama dengan ragamnya, yaitu $E[Y] = \mu = \text{var}(Y)$. Secara umum, bentuk model regresi Poisson dinyatakan dalam Persamaan (4).

$$y_i = \mu_i + \varepsilon_i; i = 1, 2, 3, \dots, n \quad (4)$$

dengan y_i sebagai jumlah kejadian dan μ_i merupakan mean kejadian yang diasumsikan tidak berubah dari data ke data.

1.4.3 Overdispersi

Salah satu asumsi yang digunakan dalam regresi Poisson menyatakan bahwa variabel dependen memiliki varians yang sama dengan nilai mean-nya. Namun pada beberapa kasus, keragaman yang muncul justru melebihi nilai rata-rata sehingga menghasilkan kondisi overdispersi. Kondisi ketika varians data lebih besar dibandingkan dengan nilai mean-nya disebut overdispersi. Kondisi ini dapat menyebabkan standar residual menjadi terlalu kecil, nilai statistik uji meningkat namun hasil menjadi tidak akurat, serta menurunkan kemampuan model dalam memprediksi data (Khan dkk., 2025). Overdispersi dapat dideteksi dengan

menentukan rasio dispersi (ϕ) melalui perbandingan nilai *chi-square* (χ^2) dan derajat bebas (df), sebagaimana dituliskan dalam Persamaan (5) (Salaki dkk., 2020).

$$\phi = \frac{\chi^2}{df} \quad (5)$$

Nilai *chi-square* (χ^2) diperoleh dari Persamaan (6).

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - E(Y))^2}{Var(Y)} \quad (6)$$

dengan df sebagai derajat bebas dan y_i sebagai nilai pengamatan ke- i dari variabel dependen. Apabila nilai rasio dispersi ϕ lebih dari satu, maka model mengalami overdispersi.

1.4.4 Regresi Binomial Negatif

Regresi Binomial Negatif (*Negative Binomial Regression*) merupakan pengembangan regresi Poisson yang dirancang untuk mengatasi permasalahan overdispersi. Model ini umumnya diterapkan pada data diskrit berupa jumlah kejadian dan dirumuskan dalam struktur *Generalized Linear Model* (GLM) menggunakan fungsi penghubung *log*. Bentuk umum regresi binomial negatif ditunjukkan pada Persamaan (7) (Widyaningsih dkk., 2021).

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} \quad (7)$$

dengan μ_i merupakan nilai ekspektasi dari y_i yang berdistribusi binomial negatif, β_0 adalah konstanta, $\beta_k X_{ki}$ adalah koefisien regresi ke- k dengan nilai variabel independen ke- k untuk observasi ke- i . Estimasi parameter $\beta_0, \beta_1, \dots, \beta_k$ dilakukan menggunakan metode *Maximum Likelihood Estimator* (MLE) dengan fungsi *likelihood* seperti pada Persamaan (8).

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^n \left(\prod_{r=0}^{y_i-1} (1 + \theta r) \left(\frac{1}{y_i!} \right) \left(\frac{\theta \mu_i}{1 + \theta \mu_i} \right)^{y_i} \left(\frac{1}{1 + \theta \mu_i} \right)^{\frac{1}{\theta}} \right) \quad (8)$$

dengan, θ merupakan parameter dispersi, sedangkan r berperan sebagai indeks pengulangan pada komponen perkalian di dalam fungsi *likelihood* yang digunakan untuk menyesuaikan efek dispersi berdasarkan banyaknya kejadian y_i .

1.4.5 Pengujian Heterogenitas Spasial dan Temporal

Heterogenitas spasial merupakan kondisi ketika hubungan antara variabel dependen dan independen tidak bersifat seragam antarwilayah. Pengujian heterogenitas spasial dilakukan untuk menilai varians residual bersifat konstan pada seluruh area pengamatan atau berubah mengikuti karakteristik lokasi tertentu. Keberadaan varians yang tidak homogen menunjukkan bahwa model global tidak lagi sesuai dan memerlukan pendekatan lokal yang mampu menyesuaikan parameter untuk setiap

wilayah. Salah satu uji yang umum digunakan adalah uji *Breusch–Pagan*, dengan hipotesis yang menyatakan:

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ (Tidak terjadi heterogenitas spasial)

$H_1 : \text{setidaknya terdapat satu } \sigma_i^2 \neq \sigma_j^2 \text{ dengan } i \neq j$ (Terjadi heterogenitas spasial)

Nilai uji *Breusch–Pagan* dinyatakan dalam Persamaan (9) (Munika dkk., 2014).

$$BP = \left(\frac{1}{2}\right) \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} + \left(\frac{1}{T}\right) \left[\frac{\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon}}{\sigma^2} \right]^2 \sim \chi_{(k)}^2 \quad (9)$$

Keterangan:

BP : Statistik uji *Breusch–Pagan*

$\mathbf{f}^T$: $(f_1, f_2, \dots, f_n)$

$\mathbf{Z}$: Matriks $\mathbf{X}$ yang distandarisasi, berukuran $n \times (p + 1)$

$\mathbf{W}$: Matriks pembobot

$\boldsymbol{\varepsilon}$: Vektor residual berukuran $n \times 1$

Heterogenitas temporal merupakan kondisi ketika hubungan antara variabel dependen dan independen berubah sepanjang waktu, sehingga parameter model tidak dapat diasumsikan konstan antarwaktu. Uji yang dapat digunakan untuk mengidentifikasi adanya pengaruh temporal ialah visualisasi *boxplot* (Sholihin dkk., 2017). Perbedaan lebar *boxplot* serta perpindahan posisi *maximum outlier* pada setiap periode menunjukkan adanya ketidakstabilan varians dari waktu ke waktu, sehingga mengindikasikan adanya heterogenitas temporal (Oktarina dkk., 2024).

1.4.6 Geographically and Temporally Weighted Regression

Geographically and Temporally Weighted Regression (GTWR) merupakan metode regresi yang memodelkan hubungan antarvariabel dengan mempertimbangkan keragaman spasial maupun temporal menggunakan matriks pembobot (Oktarina dkk., 2024). Model GTWR untuk k variabel independen dengan variabel dependen pada koordinat (u_i, v_i, t_i) untuk setiap pengamatan ditunjukkan pada Persamaan (10) (Liu dkk., 2017).

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{Y} \quad (10)$$

Vektor dan matriks yang berada pada Persamaan (10) ialah sebagai berikut:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}}(u_i, v_i, t_i) = \begin{pmatrix} \beta_0(u_i, v_i, t_i) \\ \beta_1(u_i, v_i, t_i) \\ \vdots \\ \beta_q(u_i, v_i, t_i) \end{pmatrix}$$

Matriks W merupakan matriks pembobot spasial temporal pada lokasi ke- i dan waktu ke- t yang disusun *bandwidth* spasial temporal optimum dengan metode *Cross Validation* (CV). Persamaan (11) menunjukkan penduga parameter GTWR.

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^p \beta_k(u_i, v_i, t_i)x_{ki} + \varepsilon_i \quad (11)$$

Keterangan:

- y_i : Variabel dependen pada lokasi amatan ke- i
- u_i, v_i, t_i : Koordinat geografis dan waktu lokasi amatan ke- i
- $\beta_0(u_i, v_i, t_i)$: Nilai konstanta model GTWR
- $\beta_k(u_i, v_i, t_i)$: Parameter regresi variabel independen ke- k pada lokasi dan waktu amatan ke- i
- x_{ki} : nilai observasi variabel independen ke- k pada amatan ke- i
- ε_i : Residual pengamatan ke- i

Jarak spasial-temporal dibangun melalui kombinasi jarak spasial (d^S) dan jarak temporal (d^T) yang dihitung menggunakan jarak *euclidean*. Jarak spasial diperoleh dari selisih koordinat lokasi dan jarak temporal berasal dari selisih waktu antar pengamatan. Jarak spasial (d^S) dan jarak temporal (d^T) berturut-turut didefinisikan pada Persamaan (12) dan Persamaan (13).

$$d_{ij}^S = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (12)$$

$$d_{ij}^T = \sqrt{(t_i - t_j)^2} \quad (13)$$

Karena terdapat perbedaan skala antara spasial dan temporal, maka fungsi jarak spasial-temporal dibentuk melalui kombinasi fungsi jarak spasial (d^S) dan fungsi jarak temporal (d^T). Fungsi jarak spasial-temporal ditunjukkan pada Persamaan (14).

$$(d_{ij}^{ST})^2 = \lambda [(u_i - u_j)^2 + (v_i - v_j)^2] + \mu (t_i - t_j)^2 \quad (14)$$

Parameter λ dan μ berfungsi sebagai penyeimbang efek yang berbeda pada jarak spasial dan temporal. Setelah jarak Euclidean di substitusikan, maka Persamaan (14) dapat dituliskan kembali dalam Persamaan (15).

$$(d_{ij}^{ST}) = \lambda(u_i - u_j) + (v_i - v_j) + \mu(t_i - t_j) \quad (15)$$

Rasio $\tau = \frac{\mu}{\lambda}$ digunakan ketika $\lambda \neq 0$, sehingga untuk Persamaan (15) dapat dituliskan kembali dalam Persamaan (16).

$$\frac{(d_{ij}^{ST})^2}{\lambda} = (u_i - u_j) + (v_i - v_j) + \tau(t_i - t_j) \quad (16)$$

Parameter τ digunakan untuk pengendali efek jarak temporal terhadap efek jarak spasial. Pemilihan parameter tersebut dapat dilakukan melalui metode CV dengan

menggunakan kriteria nilai minimum yang diperoleh melalui proses iterasi berdasarkan inisialisasi nilai awal τ pada Persamaan (17).

$$CV(\tau) = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}]^2 \quad (17)$$

Fungsi jarak spasial-temporal dapat digunakan untuk menghitung bobot spasial-temporal melalui fungsi pembobot kernel, diantaranya (Vianti dkk., 2024):

Fungsi Kernel *Exponential*

$$w_{ij}^{ST} = \exp \left[-\frac{d_{ij}^{ST}}{h} \right] \quad (18)$$

Fungsi Kernel *Gaussian*

$$w_{ij}^{ST} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}^{ST}}{h} \right)^2 \right] \quad (19)$$

Fungsi Kernel *Bisquare*

$$w_{ij}^{ST} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{ST}}{h} \right)^2 \right]^2, & d_{ij}^{ST} \leq h \\ 0, & d_{ij}^{ST} > h \end{cases} \quad (20)$$

Fungsi Kernel *Tricube*

$$w_{ij}^{ST} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{ST}}{h} \right)^3 \right]^3, & d_{ij}^{ST} \leq h \\ 0, & d_{ij}^{ST} > h \end{cases} \quad (21)$$

dengan h sebagai *bandwidth* spasial-temporal. Fungsi pembobot spasial-temporal dapat diturunkan melalui penggabungan fungsi pembobot spasial dan temporal secara terpisah dengan memanfaatkan formulasi jarak spasial-temporal. *Bandwidth* spasial (γ_S) dan *bandwidth* temporal (γ_T) berturut-turut didefinisikan berdasarkan parameter dan pada Persamaan (22) dan Persamaan (23).

$$(\gamma_S)^2 = \frac{(h)^2}{\lambda} \quad (22)$$

$$(\gamma_T)^2 = \frac{(h)^2}{\mu} \quad (23)$$

Estimasi *bandwidth* kernel pada model GTWR menggunakan CV dapat dilakukan menggunakan Persamaan (24)

$$CV(\tau) = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(h)]^2 \quad (24)$$

dengan $\hat{y}_{\neq i}$ sebagai nilai y_i saat pengamatan ke- i dihilangkan dari estimasi (Ma dkk., 2018).

1.4.7 Geographically Weighted Negative Binomial Regression

Geographically Weighted Negative Binomial Regression (GWNBR) merupakan model yang digunakan untuk menganalisis hubungan antara variabel dependen dan variabel independen dengan mempertimbangkan adanya heterokedastisitas spasial

pada data. Berbeda dengan model regresi global yang mengasumsikan bahwa hubungan antarvariabel bersifat homogen pada seluruh wilayah, GWNBR mengasumsikan bahwa hubungan antarvariabel dapat berubah pada lokasi yang berbeda. Sebagai bentuk pengembangan dari *Geographically Weighted Regression* (GWR), model ini berguna untuk menangani data diskrit (*count data*) yang mengalami overdispersi, yaitu kondisi ketika varians data lebih besar dibandingkan nilai meannya sehingga distribusi Poisson tidak lagi memadai.

Variabel dependen pada model GWNBR diasumsikan mengikuti distribusi binomial negatif dengan mean μ_{it} dan parameter dispersi θ dengan model yang dirumuskan pada Persamaan (25) (Yasin dkk., 2021).

$$Y_{it} \sim \text{Negatif Binomial} \left[\exp \sum_k \beta_k(u_i, v_i) x_{ik}, \theta_k(u_i, v_i) \right] \quad (25)$$

Fungsi peluang dinyatakan pada Persamaan (25) (Alwansyah dkk., 2025).

$$P(Y_i = y_i | \mu_i, \theta) = \frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta}) y_i!} \left(\frac{1}{1 + \theta \mu_i} \right)^{\frac{1}{\theta}} \left(\frac{\theta \mu_i}{1 + \theta \mu_i} \right)^{y_i} \quad (26)$$

dengan $\mu_i = \exp(X_i^T \beta_{(u_i, v_i)})$ dan $\theta = \theta_{(u_i, v_i)}$. Estimasi parameter lokal diperoleh menggunakan matriks pembobot spasial yang memberikan bobot lebih besar pada pengamatan yang berdekatan secara geografis, dan semakin kecil pada lokasi yang lebih jauh. Bobot spasial dihitung menggunakan fungsi pembobot kernel, dengan parameter β dan θ diestimasi menggunakan metode MLE melalui algoritma *Newton-Raphson Iterative Optimization*. Estimasi *bandwidth* optimum diperoleh menggunakan metode CV dengan kriteria nilai minimum yang dinyatakan pada Persamaan (24).

1.4.8 *Geographically and Temporally Weighted Negative Binomial Regression*

Geographically and Temporally Weighted Negative Binomial Regression (GTWNBR) merupakan pengembangan dari GWNBR dengan menambahkan efek waktu (temporal) ke dalam proses pembobotan. Model ini digunakan untuk menganalisis data diskrit yang mengalami overdispersi serta menunjukkan adanya heterogenitas spasial dan temporal, sehingga parameter regresi yang diperoleh berbeda pada setiap lokasi dan periode waktu. Model GTWNBR mengasumsikan bahwa variabel dependen mengikuti distribusi binomial negatif dengan mean μ_{it} dan parameter dispersi θ , dengan fungsi peluang yang dinyatakan pada Persamaan (27).

$$P(Y_{it} = y_{it} | \mu_{it}, \theta) = \frac{\Gamma(y_{it} + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta}) y_{it}!} \left(\frac{1}{1 + \theta \mu_{it}} \right)^{\frac{1}{\theta}} \left(\frac{\theta \mu_{it}}{1 + \theta \mu_{it}} \right)^{y_{it}} \quad (27)$$

Parameter β dan θ pada model GTWNBR diestimasi menggunakan metode MLE dengan algoritma *Newton-Raphson Iterative Optimization*, sehingga parameter diperoleh melalui turunan pertama dan turunan kedua dari fungsi *log-likelihood*.

Estimasi parameter dilakukan secara lokal untuk setiap lokasi (u_i, v_i) dan waktu amatan ke (t_i) dengan menerapkan pembobot lokal yang dibentuk menggunakan fungsi pembobot kernel dan ditentukan berdasarkan jarak spasial–temporal antar titik observasi. Jarak spasial dan jarak temporal dihitung menggunakan jarak *euclidean* sebagaimana ditunjukkan pada Persamaan (12) dan (13). Selanjutnya, jarak spasial–temporal dibentuk melalui kombinasi linear dari kedua jenis jarak tersebut seperti pada Persamaan (14). Fungsi jarak spasial–temporal ini digunakan untuk menghasilkan bobot spasial–temporal melalui fungsi pembobot kernel. Estimasi *bandwidth* optimum diperoleh menggunakan metode CV dengan kriteria nilai minimum yang dinyatakan pada Persamaan (24).

1.4.9 Multikolinearitas

Multikolinearitas merupakan masalah yang muncul ketika variabel independen dalam suatu model regresi memiliki hubungan linear yang sangat tinggi satu sama lain. Meskipun tidak memengaruhi kemampuan prediksi model secara keseluruhan, keberadaan multikolinearitas menyebabkan ketidakstabilan estimasi parameter. *Variance Inflation Factors* (VIF) umumnya digunakan untuk mendeteksi multikolinearitas. Rumus yang digunakan untuk memperoleh nilai VIF ditunjukkan pada Persamaan (28)(Salaki dkk., 2020).

$$VIF_k = \frac{1}{1 - R_k^2} \quad (28)$$

dimana, R_k^2 merupakan koefisien determinasi antara variabel independen X_k dengan variabel independen lainnya. Nilai VIF yang berada di bawah 10 mengindikasikan bahwa model tidak mengalami multikolinearitas yang mengganggu stabilitas estimasi (Manoppo dkk., 2025).

1.4.10 Principal Component Analysis

Principal Component Analysis (PCA) merupakan metode statistik yang digunakan untuk mereduksi dimensi data dengan mentransformasikan sekumpulan variabel yang saling berkorelasi menjadi variabel baru yang saling tidak berkorelasi. Metode ini bertujuan untuk menyederhanakan struktur data tanpa menghilangkan informasi penting (Rizky dkk., 2025). Sebelum PCA dilakukan, data perlu melalui pengujian kelayakan untuk memastikan bahwa struktur korelasinya memenuhi syarat. Uji *Kaiser Meyer Olkin* (KMO) merupakan uji yang digunakan untuk mengetahui kelayakan data observasi dalam analisis PCA. Persamaan (29) merupakan rumus yang digunakan untuk melakukan uji tersebut (Haqq dkk., 2025).

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2}, k = 1, 2, \dots, p; j = 1, 2, \dots, p \quad (29)$$

dimana,

p : Jumlah variabel yang di analisis

r_{ij}^2 : Koefisien korelasi sederhana antara variabel ke- k dan ke- j

a_{ij}^2 : Koefisien korelasi parsial antara variabel ke- k dan ke- j

Nilai $KMO < 0,5$ mengindikasikan bahwa penggunaan variabel dalam penelitian perlu dipertimbangkan kembali, karena korelasi antara variabel tidak dapat dijelaskan oleh variabel lain. Sehingga, data tidak layak di analisis dengan metode PCA. Lalu, untuk mengetahui apakah variabel layak dianalisis lebih lanjut dalam pembentukan komponen utama, dilakukan pengukuran *Measure of Sampling Adequacy* (MSA) sebagai indikator kecukupan sampel dan kelayakan korelasi antarvariabel untuk diolah menggunakan analisis PCA. Jika nilai $MSA < 0,5$ maka variabel penelitian harus dikeluarkan dari analisis karena tidak memenuhi kelayakan untuk dianalisis lebih lanjut dalam pembentukan komponen utama (Setianingsih dan Asmianto, 2024).

Kelayakan data untuk dianalisis menggunakan PCA yang telah ditunjukkan melalui uji KMO dan MSA menjadi dasar untuk melanjutkan proses ekstraksi komponen utama. Selanjutnya ialah menentukan komponen utama yang akan dipertahankan dengan memeriksa nilai eigen. Nilai eigen merupakan ukuran yang merepresentasikan besarnya proporsi varian yang dijelaskan oleh setiap komponen. Suatu komponen dikategorikan layak dipertahankan apabila memiliki nilai eigen lebih besar atau sama dengan satu (Damanik dan Sitepu, 2024). Selain itu, pemilihan komponen utama yang layak digunakan dalam PCA ialah dengan mempertimbangkan persentase keragaman kumulatif. Komponen yang secara kumulatif menjelaskan minimal 80% varians umumnya dianggap memadai untuk menggambarkan struktur data secara efisien (Delsen dkk., 2017). Pengamatan melalui *scree plot* turut digunakan dengan melihat titik siku sebagai batas yang menunjukkan jumlah komponen optimal (Gour dkk., 2017).

Kontribusi setiap variabel asal terhadap komponen utama yang terbentuk diinterpretasikan menggunakan nilai *loading*. Kekuatan dan arah korelasi antara variabel dan komponen utama ditunjukkan oleh nilai *loading*, sehingga semakin besar nilai absolut *loading* maka semakin besar pula pengaruh variabel tersebut dalam membentuk komponen terkait. Variabel dengan nilai *loading* absolut lebih besar dari 0,50 umumnya dianggap memiliki kontribusi yang signifikan dan layak dipertahankan dalam komponen utama (Enzellina dan Suhaedi, 2022).

1.4.11 Pengujian Signifikansi dan Evaluasi Kelayakan Model

1.4.11.1 Uji Parsial

Uji parsial merupakan pengujian yang digunakan untuk menilai pengaruh masing-masing variabel independen terhadap variabel dependen. Statistik uji yang digunakan untuk uji parsial adalah uji Wald ($Z - test$), dengan Persamaan (30) sebagai rumus yang digunakan untuk menghitung nilai statistik uji tersebut (Mar'ah dkk., 2023).

$$Z_{ki} = \frac{\hat{\beta}_{ki}}{SE(\hat{\beta}_{ki})} \quad (30)$$

dengan $\hat{\beta}_{ki}$ sebagai estimasi koefisien variabel independen ke- k waktu amatan ke- i dan SE adalah standar residual variabel independen ke- k . Berikut hipotesis uji yang digunakan:

$H_0 : \beta_k = 0$ (tidak ada pengaruh variabel ke- k terhadap observasi amatan)

$H_1 : \beta_k \neq 0$ (terdapat pengaruh variabel ke- k terhadap observasi amatan)
Jika nilai $|Z| > Z_{\frac{\alpha}{2}}$, maka hipotesis nol ditolak.

1.4.11.2 Pemilihan Model Terbaik

Devians merupakan salah satu ukuran statistik yang digunakan untuk mengevaluasi kesesuaian model terhadap data pengamatan. Umumnya, nilai devians dituliskan sebagai Persamaan (31) (Mar'ah dkk., 2023).

$$G^2 = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\lambda_i} \right) \quad (31)$$

dimana, y_i merupakan nilai aktual variabel dependen pada amatan ke- i dan λ_i merupakan nilai prediksi dari variabel dependen model pada amatan ke- i . Semakin kecil nilai devians pada suatu amatan, maka semakin baik model dalam memprediksi nilai variabel dependen pada amatan tersebut. Selanjutnya, pemilihan model terbaik dilakukan dengan mempertimbangkan tingkat kecocokan model serta efisiensi model dalam penggunaan parameter.

Akaike Information Criterion (AIC) merupakan metode statistik yang digunakan untuk menilai kualitas suatu model sekaligus membandingkannya dengan model alternatif. AIC mengevaluasi keseimbangan antara tingkat kecocokan model terhadap data dan kompleksitas parameter yang digunakan dengan menghitung *log-likelihood* maksimum. Nilai AIC yang lebih kecil menunjukkan bahwa model tersebut mampu menggambarkan pola data dengan lebih efisien. Persamaan (32) merupakan rumus statistik uji yang digunakan untuk menghitung nilai AIC (Mar'ah dkk., 2023).

$$AIC = 2p - 2\ln \hat{L}(\hat{\theta}) \quad (32)$$

dengan p sebagai jumlah parameter dalam model dan $\hat{L}(\hat{\theta})$ sebagai nilai *likelihood* yang berasal dari fungsi *log-likelihood* model regresi.

Bayesian Information Criterion (BIC) merupakan metode statistik yang digunakan untuk mengevaluasi kinerja suatu model dengan mempertimbangkan tingkat kesesuaian model terhadap data pengamatan serta jumlah parameter yang digunakan dalam model tersebut. Berbeda dengan AIC, kriteria BIC memasukkan ukuran sampel ke dalam komponen penalti, sehingga model dengan jumlah parameter yang lebih banyak akan dikenakan penalti yang lebih besar. Oleh karena itu, model dengan nilai BIC yang lebih kecil dianggap lebih optimal karena mampu menjelaskan data secara memadai dengan struktur model yang lebih sederhana. Persamaan (33) merupakan rumus statistik uji yang digunakan untuk menghitung nilai BIC (Tarumingkeng, 2025).

$$BIC = p \ln(n) - 2\ln \hat{L}(\hat{\theta}) \quad (33)$$

dengan, p sebagai jumlah parameter dalam model, n jumlah data pengamatan, dan $\hat{L}(\hat{\theta})$ sebagai nilai *likelihood* yang berasal dari fungsi *log-likelihood* model regresi.

1.4.12 Kasus Pneumonia pada Balita

Pneumonia merupakan penyakit infeksi pernapasan akut menular yang menyerang jaringan paru-paru dan menjadi penyebab utama kematian balita dengan jumlah kasus kematian mencapai 740.000 tiap tahunnya (World Health Organization, 2022).

Penyakit ini sering disebabkan oleh infeksi bakteri (seperti *Streptococcus pneumoniae* dan *Haemophilus influenzae*) atau virus (seperti RSV), dengan faktor risiko utama meliputi malnutrisi, polusi udara, kekurangan akses kesehatan, dan kondisi lingkungan buruk (Aprilia dkk., 2024). Di negara berkembang, kematian balita akibat pneumonia dengan prevalensi tertinggi berada di Afrika Sub-Sahara dan Asia Selatan (Putri dan Sufriani, 2025). Pencegahan melalui vaksinasi, pemberian nutrisi, dan sanitasi telah mengurangi angka kematian dalam beberapa waktu terakhir (Zairina, 2025). Namun tantangan seperti perubahan iklim dan pandemi masih memperburuk situasi (Kemenkes RI, 2021). Di Indonesia, pneumonia merupakan penyebab kematian terbesar kedua pada balita setelah diare (Kulsum dkk., 2023). Wilayah dengan tingkat kepadatan penduduk yang tinggi, memiliki kerentanan lebih besar terhadap penyebaran pneumonia akibat intensitas interaksi sosial yang tinggi.

Status gizi, praktik pemberian makanan, cakupan imunisasi, kondisi sanitasi, kualitas udara, dan berat badan lahir sebagai faktor risiko utama pneumonia pada balita (Indah dan Refialdinata, 2025). Malnutrisi melemahkan sistem kekebalan tubuh, meningkatkan risiko infeksi bakteri dan virus (Gamal dkk., 2023). Pemberian ASI eksklusif selama enam bulan pertama kehidupan terbukti meningkatkan antibodi dan proteksi imunologis, sehingga kegagalan pemberian ASI eksklusif meningkatkan risiko infeksi pernapasan (Saputra dkk., 2025). Cakupan imunisasi DPT-Hb-Hib dan imunisasi dasar lengkap sangat penting untuk mencegah infeksi bakteri penyebab pneumonia seperti *Haemophilus Influenzae type B* dan *Streptococcus Pneumoniae*, sehingga rendahnya cakupan imunisasi meningkatkan kerentanan pneumonia (Iswari dkk., 2017). Selain itu, pemberian vitamin A merupakan salah satu intervensi yang berperan dalam menjaga fungsi epitel saluran pernapasan dan meningkatkan kekebalan tubuh, sehingga defisiensi vitamin A meningkatkan risiko pneumonia berat. Berat Badan Lahir Rendah (BBLR) juga menjadi faktor penting, karena bayi dengan berat lahir kurang dari 2500 gram biasanya memiliki sistem imun yang belum berkembang optimal (Jasmine dkk., 2023). Di sisi lain, kondisi lingkungan seperti cuaca ekstrem turut memengaruhi tingginya kasus pneumonia, terutama saat perubahan suhu signifikan yang melemahkan daya tahan tubuh balita (Makrufardi dkk., 2024).

BAB II

METODE PENELITIAN

2.1 Sumber Data dan Variabel

Data dalam penelitian ini merupakan data sekunder sejak tahun 2021 hingga 2024 yang diperoleh dari publikasi “Profil Kesehatan Provinsi Jawa Timur” yang diterbitkan oleh Dinas Kesehatan Jawa Timur, “Laporan Kerja Instansi Pemerintah (LKJiP)” yang dipublikasikan oleh Dinas Perumahan Rakyat, Kawasan Pemukiman dan Cipta Karya Provinsi Jawa Timur, serta Badan Pusat Statistik Provinsi Jawa Timur yang terpublikasi pada laman jatim.bps.go.id. Tabel 1 menunjukkan variabel yang digunakan dalam penelitian dan Lampiran 1 menampilkan dataset yang digunakan dalam penelitian.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Definisi Operasional	Sumber
Y	Kasus Pneumonia pada balita setiap kabupaten/kota di Provinsi Jawa Timur	Jumlah balita (usia 0-59 bulan) yang terdiagnosis pneumonia	Profil Kesehatan Provinsi Jawa Timur
X_1	Persentase Balita Gizi Kurang	Persentase balita dengan status gizi kurang (kondisi berat badan balita lebih rendah dari berat badan ideal menurut umur berdasarkan kurva pertumbuhan WHO).	Profil Kesehatan Provinsi Jawa Timur
X_2	Pemberian Eksklusif	ASI Persentasi bayi usia 0-6 bulan yang hanya diberi ASI (Air Susu Ibu) tanpa tambahan makanan atau minuman lain.	Profil Kesehatan Provinsi Jawa Timur
X_3	Cakupan imunisasi DPT-Hb-Hib	Jumlah bayi yang telah menerima imunisasi DPT-Hb-Hib, yaitu vaksin kombinasi untuk melindungi dari difteri, pertusis, tetanus, hepatitis B, dan infeksi <i>Haemophilus influenzae</i> tipe B	Profil Kesehatan Provinsi Jawa Timur
X_4	Imunisasi Lengkap	Dasar Jumlah bayi yang telah mendapatkan semua imunisasi dasar lengkap (BCG, DPT-HB-Hib dan Campak).	Profil Kesehatan Provinsi Jawa Timur

Variabel	Keterangan	Definisi Operasional	Sumber
X_5	Persentase Berat Badan Lahir Rendah	Persentase bayi yang lahir dengan berat badan kurang dari 2500 gram.	Profil Kesehatan Provinsi Jawa Timur
X_6	Pemberian Vitamin A	Jumlah balita yang mendapatkan kapsul Vitamin A dua kali dalam setahun.	Profil Kesehatan Provinsi Jawa Timur
X_7	Persentase Sanitasi Layak	Persentase rumah tangga yang memiliki fasilitas sanitasi yang memenuhi standar kesehatan.	Laporan Kerja Instansi Pemerintah (LKJiP) Provinsi Jawa Timur
X_8	Jumlah Kejadian Bencana Alam (Cuaca Ekstrem)	Jumlah peristiwa cuaca berbahaya yang terjadi dalam satu daerah selama satu tahun.	Badan Pusat Statistik Provinsi Jawa Timur

2.2 Tahapan Analisis Data

Berikut tahapan analisis yang dilakukan dalam penelitian ini.

1. Melakukan estimasi parameter pada model *Geographically and Temporally Weighted Negative Binomial Regression Principal Component Analysis* menggunakan metode *Maximum Likelihood Estimation* dengan pendekatan iteratif *Newton–Raphson*.
2. Melakukan eksplorasi terhadap data jumlah kasus pneumonia pada balita di Provinsi Jawa Timur tahun 2021-2024 serta variabel yang memengaruhinya.
3. Mengidentifikasi overdispersi pada model Poisson dengan membandingkan rasio *deviance* dan *Pearson chi-square* terhadap derajat bebas. Overdispersi terjadi jika nilai rasio > 1 .
4. Mengidentifikasi multikolinearitas menggunakan VIF menggunakan Persamaan (28). Jika nilai VIF lebih besar dari 10, maka terdapat multikolinearitas antarvariabel penelitian.
5. Mengidentifikasi heterogenitas spasial dengan uji *Breusch-Pagan* menggunakan Persamaan (9) dengan hipotesis uji sebagai berikut.
 $H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ (Tidak terjadi heterogenitas spasial)
 $H_1 : \text{setidaknya terdapat satu } \sigma_i^2 \neq \sigma_j^2 \text{ dengan } i \neq j$ (Terjadi heterogenitas spasial)
 Pengambilan keputusan dalam uji *Breusch-Pagan* dilakukan berdasarkan nilai uji *BP* atau *p – value*. Jika nilai $BP > \chi_{\alpha,p}^2$ atau nilai $p – value < \alpha$ (Tingkat signifikansi), maka hipotesis H_0 ditolak. Lalu, mengidentifikasi heterogenitas temporal melalui visualisasi *boxplot*.
6. Melakukan standardisasi data.

7. Mengatasi masalah multikolinearitas menggunakan PCA pada variabel independen dengan mengekstraksi beberapa komponen utama yang memiliki nilai eigen ≥ 1 atau mampu menjelaskan keragaman data 80%.
8. Menganalisis model GTWNBR-PCA dengan langkah-langkah sebagai berikut:
 - a. Menentukan jarak spasial-temporal menggunakan Persamaan (14).
 - b. Menentukan fungsi pembobot optimum menggunakan *Fixed Kernel*.
 - c. Menentukan *bandwidth* optimum dengan meminimalkan nilai CV menggunakan Persamaan (24).
 - d. Memperoleh estimasi parameter GTWNBR-PCA.
 - e. Melakukan uji parsial untuk setiap parameter GTWNBR-PCA menggunakan Persamaan (30) dengan hipotesis uji sebagai berikut.

$$H_o : \beta_k = 0 \text{ (tidak ada pengaruh variabel ke-} k \text{ terhadap observasi amatan)}$$

$$H_1 : \beta_k \neq 0 \text{ (terdapat pengaruh variabel ke-} k \text{ terhadap observasi amatan)}$$
 Pengambilan keputusan uji parsial berdasarkan nilai statistik uji. Apabila nilai $|Z| > Z_{\frac{\alpha}{2}}$, maka hipotesis H_o ditolak.
9. Pemilihan model terbaik berdasarkan nilai AIC dan BIC terkecil.
10. Melakukan interpretasi model GTWNBR-PCA.