

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan metode statistik yang digunakan untuk menentukan hubungan antara variabel respon dan variabel prediktor, sehingga memungkinkan dilakukannya prediksi terhadap variabel respon berdasarkan variabel prediktor (Araveeporn, 2024). Tujuan utama dalam melakukan analisis regresi adalah untuk mencari bentuk estimasi kurva regresi (Dani dkk., 2021). Pada umumnya terdapat tiga pendekatan dalam mengestimasi kurva regresi, yaitu pendekatan parametrik, nonparametrik, dan semiparametrik (Tanaty dkk., 2023). Pendekatan regresi parametrik merupakan pendekatan yang digunakan ketika bentuk hubungannya diketahui dan diasumsikan mengikuti pola data tertentu, sedangkan pendekatan regresi nonparametrik merupakan pendekatan yang digunakan ketika bentuk hubungannya tidak diketahui polanya yang selanjutnya kurva regresinya diasumsikan mulus. Sementara itu, pendekatan regresi semiparametrik merupakan kombinasi dari nonparametrik dan parametrik (Dani & Adrianingsih, 2021). Namun, sering kali dalam prakteknya bentuk pola hubungan antara variabel respon dan prediktor tidak diketahui, sehingga pendekatan yang disarankan adalah regresi nonparametrik.

Pendekatan nonparametrik merupakan salah satu metode dalam analisis regresi yang tidak memerlukan asumsi khusus mengenai bentuk hubungan antara variabel respon dan variabel prediktor (Sanusi dkk., 2019). Model ini bersifat fleksibel karena bentuk fungsi regresinya ditentukan langsung oleh pola data, bukan oleh asumsi peneliti (Dani dkk., 2021). Dalam regresi nonparametrik, terdapat beberapa pendekatan estimator untuk kurva regresi, seperti *spline*, *kernel*, *deret fourier*, dan lainnya (Ramli dkk., 2024). Metode estimasi yang banyak mendapat perhatian dari beberapa peneliti regresi nonparametrik salah satunya adalah estimator *spline* (Mariati dkk., 2021). *Spline* merupakan segmen polinomial yang bersifat fleksibel yang mampu menyesuaikan pola data pada interval tertentu melalui penggunaan titik knot. Titik tersebut menandakan adanya perubahan pola perilaku fungsi pada interval yang berbeda (Davala dkk., 2024). Pendekatan penaksir *spline* dikenal dengan beberapa metode seperti *Truncated Spline*, *Smoothing Spline*, *B-Spline*, *Exponential Spline* dan lainnya (Mariati dkk., 2021).

Metode *spline* pada penerapannya sering kali menghadapi kendala dalam menentukan jumlah dan posisi titik knot yang optimal (Luo dkk., 2019). Jumlah knot yang terlalu banyak dapat menyebabkan *overfitting*, sedangkan jumlah knot yang terlalu sedikit dapat mengabaikan variasi penting pada data (Fajar & Fajariyanto, 2022). Oleh karena itu, diperlukan pendekatan yang mampu mengatasi permasalahan tersebut, salah satunya dengan menambahkan komponen penalti pada fungsi *spline*. Pendekatan ini dikenal sebagai *Penalized Spline*, yang merupakan pengembangan dari *spline* yang lebih stabil dan fleksibel dalam menangkap pola data.

Penalized Spline merupakan metode nonparametrik yang menggabungkan penggunaan titik knot dengan penambahan komponen penalti untuk mengontrol tingkat kelengkungan kurva (Wood, 2017). Metode ini memungkinkan estimasi yang lebih akurat karena parameter penghalus dan titik knot diestimasi secara simultan dalam mengatur kemulusan fungsi regresi (Usrah dkk., 2022). Dengan adanya penalti, model menjadi lebih stabil serta tidak terlalu sensitif terhadap jumlah maupun letak knot yang digunakan. Penelitian sebelumnya yang dilakukan oleh Octavia dkk. (2024) menunjukkan bahwa metode regresi *Penalized Spline* efektif dalam memprediksi pergerakan Indeks Harga Saham Gabungan (IHSG) selama periode Januari 2020 hingga Desember 2022. Hasil analisis menunjukkan bahwa model memiliki kemampuan prediksi yang sangat baik dengan pengaruh variabel prediktor terhadap IHSG sebesar 92,1%. Dengan demikian, metode *Penalized Spline* efektif dalam memodelkan hubungan yang tidak diketahui antara variabel serta mampu menangkap pola data yang kompleks.

Metode *Penalized Spline* yang fleksibel dalam menangkap pola data kompleks dapat dikembangkan lebih lanjut untuk menganalisis fenomena yang memiliki variasi antarwilayah. Hal ini penting karena dalam banyak kasus, ketika penelitian berfokus pada suatu wilayah, hubungan antara variabel respon dan variabel prediktor tidak bersifat seragam di seluruh area pengamatan, melainkan berbeda di setiap lokasi (Safitri & Amaliana, 2021). Kondisi ini dikenal sebagai heterogenitas spasial, yang menggambarkan perbedaan pengaruh suatu variabel terhadap variabel lain sesuai dengan karakteristik masing-masing wilayah (Pratiwi dkk., 2020). Untuk mengakomodasi perbedaan tersebut, digunakan metode *Geographically Weighted Regression* (GWR) yang memungkinkan setiap lokasi memiliki parameter regresi tersendiri, sehingga mampu menggambarkan hubungan lokal secara lebih akurat (Janah & Utami, 2021). Namun, pendekatan GWR yang dikembangkan sebelumnya masih terbatas pada regresi parametrik dan mengasumsikan bentuk hubungan tertentu antara variabel respon dan prediktor (Laome dkk., 2024). Untuk mengatasi keterbatasan tersebut, beberapa penelitian menggabungkan GWR dengan pendekatan nonparametrik, di antaranya yaitu Laome dkk. (2024) mengusulkan pengujian *goodness of fit* pada model *Geographically Weighted Nonparametric Regression* (GWNR) yang dibangun dengan estimator *spline truncated* dan *deret fourier*, sehingga mampu mengatasi keterbatasan model GWR parametrik yang hanya mengasumsikan bentuk hubungan tertentu antara variabel respon dan prediktor. Sementara itu, Arman (2025) memodelkan GWNR dengan pendekatan *Spline Penalized*, yang memberikan fleksibilitas lebih dalam mengontrol tingkat kehalusan kurva estimasi sekaligus mempertahankan kestabilan model.

Selain memperhitungkan efek spasial, variasi antarwaktu atau efek temporal juga perlu diperhatikan karena hubungan antara variabel respon dan variabel prediktor dapat berubah seiring waktu. Dalam konteks ini, metode *Geographically and Temporally Weighted Regression* (GTWR) menjadi pengembangan dari GWR karena mampu mengakomodasi heterogenitas baik secara spasial maupun temporal, sehingga model dapat menangkap dinamika

hubungan lokal yang berubah dari waktu ke waktu (Praditya dkk., 2025). Beberapa penelitian telah mengembangkan pendekatan ini dengan regresi nonparametrik, salah satunya dilakukan oleh Alkaf (2024) yang menggunakan model *Nonparametric Geographically and Temporally Weighted Regression* dengan pendekatan *Spline Truncated* orde satu pada kasus tingkat pengangguran terbuka di Provinsi Sulawesi Selatan. Sementara itu, Sifriyani dkk. (2025) mengembangkan model *Geographically and Temporally Weighted Spline Nonparametric Regression* (GTWSNR) yang menggabungkan pembobot spasial dan temporal dengan estimator *spline truncated* pada data produktivitas padi dari 34 provinsi di Indonesia. Hasil penelitian menunjukkan bahwa model GTWSNR memiliki kinerja terbaik dengan nilai R^2 sebesar 90,89%, sehingga membuktikan efektivitasnya dalam menangkap pola nonlinier serta heterogenitas spasial-temporal. Penelitian sebelumnya masih terbatas pada penggunaan estimator *spline truncated* dalam kerangka regresi nonparametrik spasial-temporal. Oleh karena itu, penelitian ini mengintegrasikan keunggulan fleksibilitas regresi nonparametrik *penalized spline* dalam menangkap pola hubungan yang tidak diketahui dengan pendekatan GTWR yang mampu mengakomodasi heterogenitas spasial dan temporal. Pendekatan ini disebut *Nonparametric Geographically and Temporally Weighted Regression* dengan pendekatan *Penalized Spline* (NGTWR-PS), yang diharapkan dapat menghasilkan estimasi parameter yang lebih akurat. Metode ini diterapkan untuk menganalisis Tingkat Pengangguran Terbuka (TPT) di Provinsi Sulawesi Selatan.

Tingkat Pengangguran Terbuka (TPT) merupakan salah satu indikator penting dalam mengukur kondisi ketenagakerjaan dan kinerja ekonomi suatu wilayah. Nilai TPT mencerminkan tingkat keseimbangan antara jumlah angkatan kerja yang tersedia dan kesempatan kerja yang ada. Secara nasional, kondisi ketenagakerjaan di Indonesia pada tahun 2025 terus mengalami perbaikan yang tercermin dari penurunan TPT menjadi 4,76% pada Februari 2025, atau turun sebesar 0,06% dibandingkan Februari 2024 yang sebesar 4,82% (BPS, 2025). Meskipun demikian, Menteri Ketenagakerjaan menegaskan bahwa tantangan ketidakpastian ekonomi global dan perang tarif masih perlu menjadi perhatian serius (Kemnaker, 2025). Kondisi makroekonomi tersebut tentunya tidak berdampak secara seragam pada seluruh wilayah di Indonesia, mengingat adanya perbedaan struktur ekonomi dan karakteristik pasar kerja yang menyebabkan tingkat serta dinamika pengangguran antar wilayah dapat bervariasi.

Provinsi Sulawesi Selatan sebagai salah satu pusat pertumbuhan ekonomi di Kawasan Timur Indonesia, memiliki struktur ekonomi yang beragam, meliputi sektor pertanian, industri, perdagangan, dan jasa, dengan tingkat perkembangan yang berbeda antar kabupaten/kota. Kondisi ini berimplikasi pada perbedaan kemampuan penyerapan tenaga kerja di setiap wilayah. Berdasarkan data Badan Pusat Statistik (2024), nilai TPT di Provinsi Sulawesi Selatan pada tahun 2024 menunjukkan variasi yang cukup besar antar kabupaten/kota, dengan kisaran antara 1,51% hingga 9,71%. Selain itu, perubahan TPT dari tahun ke tahun juga menunjukkan adanya dinamika ekonomi yang tidak seragam antar wilayah. Untuk memahami kondisi tersebut, diperlukan identifikasi faktor-faktor yang memengaruhi

TPT di Provinsi Sulawesi Selatan. Faktor-faktor tersebut melibatkan berbagai aspek sosial, ekonomi, dan pendidikan yang saling berinteraksi, sehingga menciptakan pola hubungan yang kompleks antara faktor-faktor penjelas dan TPT (Adriyanto dkk., 2020). Berdasarkan hal tersebut, maka dilakukan penelitian mengenai Pemodelan *Nonparametric Geographically and Temporally Weighted Regression* dengan pendekatan *Penalized Spline* pada data Tingkat Pengangguran Terbuka di Sulawesi Selatan tahun 2022-2024.

1.2 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter model *Nonparametric Geographically and Temporally Weighted Regression* dengan pendekatan *Penalized Spline*.
2. Memperoleh model TPT di Provinsi Sulawesi Selatan menggunakan metode *Nonparametric Geographically and Temporally Weighted Regression* dengan pendekatan *Penalized Spline*.

Manfaat penelitian ini adalah sebagai berikut:

1. Meningkatkan pengetahuan dan menambah wawasan bagi penulis maupun pembaca tentang pemodelan menggunakan *Nonparametric Geographically and Temporally Weighted Regression* dengan pendekatan *Penalized Spline* serta menjadi rujukan penelitian selanjutnya.
2. Memberikan pengetahuan dan informasi mengenai faktor-faktor yang memengaruhi TPT pada setiap kabupaten/kota di Provinsi Sulawesi Selatan Tahun 2022-2024.

1.3 Batasan Penelitian

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data TPT di Sulawesi Selatan pada tahun 2022 sampai 2024 dengan empat variabel prediktor yang diduga memengaruhi TPT.
2. Penentuan titik knot serta parameter penghalus optimal menggunakan metode *Generalized Cross Validation* (GCV).
3. Metode penentuan jarak antar lokasi pengamatan menggunakan jarak Euclidean dengan fungsi pembobot yang digunakan adalah fungsi *fixed kernel* berdasarkan nilai *Cross Validation* (CV) minimum.
4. Metode estimasi parameter yang digunakan adalah *Penalized Weighted Least Square* (PWLS).

1.4 Teori

1.4.1 Regresi Linier

Analisis regresi merupakan metode statistika yang digunakan untuk menganalisis hubungan antara variabel respon (y) dengan variabel prediktor (x) pada suatu data. Model regresi yang paling sering digunakan adalah model regresi linier. Regresi linier merupakan suatu metode analisis yang digunakan untuk memodelkan

hubungan linier antara variabel respon dan satu atau lebih variabel prediktor, dengan tujuan untuk memprediksi nilai variabel respon berdasarkan variabel prediktor (Kartiningrum dkk., 2022). Regresi linier terbagi menjadi dua jenis, yaitu regresi linier sederhana yang melibatkan satu variabel prediktor dan regresi linier berganda yang melibatkan lebih dari satu variabel prediktor. Keunggulan analisis regresi linier berganda yaitu mampu menduga keadaan di masa depan melalui pengukuran beberapa variabel prediktor $(x_1, x_2, \dots, x_d)$ dengan variabel respon (y) (Prasetyo & Helma, 2022). Bentuk umum regresi linier berganda dengan d variabel prediktor dituliskan pada Persamaan (1) (Prahutama & Rahmawati, 2023):

$$y_i = \beta_0 + \sum_{j=1}^d \beta_j x_{ji} + \varepsilon_i \quad (1)$$

$$i = 1, 2, \dots, n; j = 1, 2, \dots, d$$

Keterangan:

y_i : variabel respon pada pengamatan ke- i

β_0 : nilai intersep model regresi

β_j : nilai koefisien parameter regresi untuk variabel prediktor ke- j

x_{ji} : variabel prediktor ke- j pada pengamatan ke- i

d : banyaknya variabel prediktor

ε_i : residual pada pengamatan ke- i dengan asumsi $\varepsilon_i \sim NIID(0, \sigma^2)$

Model regresi linier berganda pada persamaan (1) dapat ditulis dalam notasi matriks seperti pada Persamaan (2).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

dengan:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{d1} \\ 1 & x_{12} & x_{22} & \cdots & x_{d2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{dn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Keterangan:

$\mathbf{y}$: vektor variabel respon berukuran $n \times 1$

$\mathbf{X}$: matriks variabel prediktor berukuran $n \times (d + 1)$

$\boldsymbol{\beta}$: vektor parameter regresi berukuran $(d + 1) \times 1$

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

n : banyak pengamatan

Parameter pada model regresi linier berganda dapat diestimasi menggunakan metode *Ordinary Least Squares* (OLS). Pada metode OLS, estimasi parameter dilakukan dengan meminimalkan jumlah kuadrat error seperti pada Persamaan (3) (Pangesti dkk., 2021).

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (3)$$

untuk mendapatkan estimator kuadrat terkecil ($\hat{\boldsymbol{\beta}}$) yang meminimumkan $S(\boldsymbol{\beta})$ disyaratkan bahwa $\left. \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0$. Turunan pertama dari $S(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ adalah:

$$\left. \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} \\ -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = 0$$

sehingga diperoleh hasil estimasi parameter pada Persamaan (4) berikut.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4)$$

Hubungan antara variabel respon dan variabel prediktor dapat diamati melalui *scatter plot*. Namun, tidak semua hubungan tersebut mengikuti pola parametrik. Dalam beberapa kondisi, bentuk kurva regresi antara variabel belum diketahui, sehingga metode regresi nonparametrik digunakan untuk memodelkan hubungan tersebut (Nurhuda dkk., 2022).

1.4.2 Regresi Nonparametrik *Penalized Spline*

Regresi nonparametrik merupakan pendekatan yang digunakan untuk mengestimasi kurva regresi, khususnya ketika bentuk hubungan antara variabel respon dengan variabel prediktor tidak diketahui bentuk kurva regresinya atau tidak terdapat informasi distribusi mengenai bentuk pola data. Pendekatan ini, memiliki fleksibilitas dalam mencari sendiri bentuk kurva regresinya tanpa dipengaruhi subjektivitas peneliti (Dani dkk., 2021). Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu (Famalika & Sihombing, 2022). Model regresi nonparametrik secara umum dituliskan pada Persamaan (5).

$$y_i = f(x_i) + \varepsilon_i \quad ; i = 1, 2, \dots, n \quad (5)$$

Keterangan:

y_i : variabel respon ke- i

$f(x_i)$: fungsi regresi dengan bentuk kurva regresi yang tidak diketahui

x_i : variabel prediktor

ε_i : residual pada pengamatan ke- i dengan asumsi $\varepsilon_i \sim NIID(0, \sigma^2)$

Salah satu metode yang banyak digunakan dalam regresi nonparametrik adalah *spline*. Metode ini mampu memberikan pendekatan yang fleksibel terhadap data dengan pola yang tidak diketahui secara pasti. *Spline* merupakan potongan polinomial tersegmen yang memiliki keunggulan dalam menangkap pola data yang mengalami perubahan tajam naik maupun turun melalui bantuan titik-titik knot, serta menghasilkan kurva yang relatif mulus. Titik-titik knot pada analisis regresi *spline* berfungsi sebagai batas antarsegmen yang menandai perubahan perilaku kurva pada selang tertentu. Secara umum, estimasi fungsi regresi f pada Persamaan (5) dengan pendekatan *spline* berorde m dan sejumlah titik knot $K_1, K_2, \dots, K_r$ untuk satu variabel prediktor dapat dinyatakan pada Persamaan (6) (Nurhuda dkk., 2022).

$$f(x_i) = \sum_{k=0}^m \beta_k x_i^k + \sum_{h=1}^r \beta_{m+h} (x_i - K_h)_+^m \quad (6)$$

dengan $i = 1, 2, \dots, n$ adalah banyaknya pengamatan, dan fungsi *truncated* (potongan)

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

Keterangan:

β_k : parameter fungsi polinomial ke- k

k : orde dengan $k = 0, 1, 2, \dots, m$

β_{m+h} : parameter komponen *truncated*, $h = 1, 2, \dots, r$

r : banyak knot

K_h : letak titik knot ke- h

Apabila fungsi *spline* pada Persamaan (6) disubstitusikan ke dalam Persamaan (5) maka akan diperoleh persamaan regresi nonparametrik *spline* berorde m dan titik-titik knot dengan satu variabel prediktor seperti pada Persamaan (7).

$$y_i = \sum_{k=0}^m \beta_k x_i^k + \sum_{h=1}^r \beta_{m+h} (x_i - K_h)_+^m + \varepsilon_i \quad (7)$$

Sedangkan untuk fungsi *spline* multiprediktor dengan satu variabel respon dan lebih dari satu variabel prediktor $(x_{1i}, x_{2i}, \dots, x_{di}, y_i)$, bentuk fungsi *spline* diperoleh berdasarkan penjumlahan antara fungsi polinomial dengan fungsi *truncated*. Kurva regresi $f_j(x_{ji})$ diasumsikan termuat pada ruang *spline* berorde m dengan titik-titik knot $K_{j1}, K_{j2}, \dots, K_{jr}; j = 1, 2, \dots, d$ dituliskan pada Persamaan (8)

$$f_j(x_{ji}) = \sum_{k=0}^m \beta_{jk} x_{ji}^k + \sum_{h=1}^r \beta_{j,m+h} (x_{ji} - K_{jh})_+^m \quad (8)$$

Berdasarkan Persamaan (8), model regresi nonparametrik *spline* multiprediktor dapat dituliskan pada Persamaan (9).

$$y_i = \beta_0 + \sum_{j=1}^d \left(\sum_{k=1}^m \beta_{jk} x_{ji}^k + \sum_{h=1}^r \beta_{j,m+h} (x_{ji} - K_{jh})_+^m \right) + \varepsilon_i \quad (9)$$

dengan fungsi *truncated* (potongan) seperti pada Persamaan (10)

$$(x_{ji} - K_{jh})_+^m = \begin{cases} (x_{ji} - K_{jh})^m, & x_{ji} \geq K_{jh} \\ 0, & x_{ji} < K_{jh} \end{cases} \quad (10)$$

Model regresi nonparametrik *spline* multiprediktor pada persamaan (9) dapat ditulis dalam notasi matriks seperti pada Persamaan (11) (Sholicha dkk., 2018).

$$y = X\beta + \varepsilon \quad (11)$$

dengan:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \vdots \\ \beta_{1m} \\ \beta_{1,m+1} \\ \vdots \\ \beta_{1,m+r} \\ \vdots \\ \beta_{d1} \\ \vdots \\ \beta_{dm} \\ \beta_{d,m+1} \\ \vdots \\ \beta_{d,m+r} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{11}^m & (x_{11} - K_{11})_+^m & \dots & (x_{11} - K_{1r})_+^m & \dots & x_{d1} & \dots & x_{d1}^m & (x_{d1} - K_{d1})_+^m & \dots & (x_{d1} - K_{dr})_+^m \\ 1 & x_{12} & \dots & x_{12}^m & (x_{12} - K_{11})_+^m & \dots & (x_{12} - K_{1r})_+^m & \dots & x_{d2} & \dots & x_{d2}^m & (x_{d2} - K_{d1})_+^m & \dots & (x_{d2} - K_{dr})_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{1n}^m & (x_{1n} - K_{11})_+^m & \dots & (x_{1n} - K_{1r})_+^m & \dots & x_{dn} & \dots & x_{dn}^m & (x_{dn} - K_{d1})_+^m & \dots & (x_{dn} - K_{dr})_+^m \end{bmatrix}$$

Pemilihan jumlah dan letak titik knot sangat berpengaruh terhadap fleksibilitas model *spline*. Jumlah knot yang terlalu sedikit dapat menyebabkan model mengalami *underfitting*, sedangkan jumlah knot yang terlalu banyak dapat menimbulkan *overfitting* terhadap data. Penentuan jumlah knot yang tepat dapat dilakukan melalui serangkaian percobaan (*trial and error*), tetapi proses tersebut membutuhkan waktu yang cukup lama, terutama jika data yang digunakan berukuran besar dan kompleks. Untuk mengatasi permasalahan tersebut, digunakan pendekatan *penalized spline* (Griggs, 2013).

Penalized spline merupakan metode yang memberikan penalti pada parameter koefisien *spline* ($\beta_{j,m+h}$) untuk mengoptimalkan hasil estimasi dan menghindari *overfitting*. Dengan kata lain, metode ini memberikan penalti pada komponen *spline* agar tercapai keseimbangan antara kemulusan (*smoothness*) kurva dan ketepatan model dalam menyesuaikan data (*goodness of fit*). Pendekatan ini dikenal sebagai *Penalized Least Square* (PLS) karena menggabungkan dua komponen utama, yaitu fungsi kesesuaian model terhadap data dan fungsi penalti terhadap kemulusan kurva. Estimasi *penalized spline* diperoleh dengan meminimumkan fungsi PLS. Kriteria minimisasi ini efektif karena mampu mengurangi pengaruh keseluruhan dari fungsi-fungsi *spline* tersegmen individu, sehingga dapat menghindari *overfitting* pada data. Fungsi PLS dirumuskan pada Persamaan (12).

$$PLS = \sum_{i=1}^n (y_i - f_j(x_{ji}))^2 + \lambda \sum_{h=1}^r \beta_{j,m+h}^2 \quad (12)$$

dengan $\lambda \geq 0$ merupakan parameter penghalus yang berfungsi mengontrol keseimbangan antara kemulusan kurva dan ketepatan model terhadap data (Griggs, 2013). Fungsi PLS pada Persamaan (12) dapat dituliskan dalam bentuk matriks sebagai berikut (Andriani dkk., 2015).

$$\begin{aligned} PLS &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{y} - 2 \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \end{aligned} \quad (13)$$

dengan $\mathbf{D}$ merupakan matriks diagonal dan $\mathbf{I}$ matriks identitas yang didefinisikan sebagai:

$$\mathbf{D} = \begin{bmatrix} \mathbf{0}_{(m+1) \times (m+1)} & \mathbf{0}_{(m+1) \times r} \\ \mathbf{0}_{r \times (m+1)} & \mathbf{I}_{r \times r} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I} \end{bmatrix} \quad (14)$$

Untuk meminimumkan Persamaan (13) maka turunan pertama terhadap $\boldsymbol{\beta}^T$ harus sama dengan nol. Sehingga didapatkan nilai estimasi $\hat{\boldsymbol{\beta}}$ pada Persamaan (15):

$$\begin{aligned}
\frac{\partial PLS}{\partial \beta^T} &= 0 \\
\frac{\partial(y^T y - 2\beta^T X^T y + \beta^T X^T X \beta + \lambda \beta^T D \beta)}{\partial \beta^T} &= 0 \\
0 - 2X^T y + 2X^T X \beta + 2\lambda D \beta &= 0 \\
2(X^T X + \lambda D)\beta &= 2X^T y \\
\hat{\beta} &= (X^T X + \lambda D)^{-1} X^T y \quad (15)
\end{aligned}$$

Selanjutnya substitusi nilai estimasi $\hat{\beta}$ pada Persamaan (15) pada model sebagai berikut:

$$\hat{y} = X\hat{\beta} = X(X^T X + \lambda D)^{-1} X^T y \quad (16)$$

Estimasi fungsi *penalized spline* pada persamaan (16) dapat dinyatakan sebagai berikut:

$$\hat{y} = A_\lambda y \quad (17)$$

dengan

$$A_\lambda = X(X^T X + \lambda D)^{-1} X^T$$

A_λ : matriks hat yang memuat parameter penghalus.

1.4.3 Pemilihan Jumlah Knot dan Parameter Penghalus Optimum

Kemulusan kurva regresi nonparametrik *penalized spline* bergantung pada pemilihan jumlah titik knot dan parameter penghalus (λ) yang optimal (Anggraeni dkk., 2022). Pada regresi *penalized spline*, letak titik knot umumnya ditentukan berdasarkan kuantil dari nilai unik (tunggal) variabel prediktor. Letak titik knot ke- h berada pada sampel kuantil ke- $\frac{h}{r+1}$ dari nilai unik variabel prediktor, sehingga letak titik knot membagi sekumpulan pengamatan menjadi $(r + 1)$ bagian yang sama besar (Zia dkk., 2017). Jumlah knot optimal ditentukan dengan menguji seluruh nilai $r = 1, 2, 3, \dots$ yang memenuhi syarat $r < (n_{unique} - m - 1)$, dengan n_{unique} adalah banyaknya nilai *unique* dari variabel prediktor sehingga jumlah knot (r) kurang dari jumlah pengamatan. Jumlah dan letak titik knot memiliki pengaruh yang signifikan terhadap fleksibilitas model. Semakin banyak titik knot yang digunakan, kurva regresi akan menjadi lebih lentur dan sensitif terhadap perubahan data, namun hal ini dapat meningkatkan risiko *overfitting*. Sebaliknya, penggunaan jumlah titik knot yang terlalu sedikit dapat menyebabkan kurva menjadi terlalu kaku (*underfitting*), sehingga tidak mampu merepresentasikan pola data dengan baik (Fajar & Fajariyanto, 2022).

Kemulusan dan fleksibilitas model regresi *penalized spline* tidak hanya ditentukan oleh jumlah serta letak titik knot, tetapi juga oleh pemilihan parameter penghalus (λ). Parameter penghalus merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan kurva. Nilai λ yang besar memberikan penalti tinggi terhadap perubahan bentuk kurva, sehingga model yang dihasilkan menjadi lebih halus namun cenderung memiliki bias tinggi. Sebaliknya, nilai λ yang terlalu kecil akan menyebabkan penalti terhadap kelengkungan kurva menjadi lemah, sehingga kurva dapat mengikuti data secara berlebihan dan

menyebabkan *overfitting* (Saputri dkk., 2015). Oleh karena itu, pada regresi *penalized spline*, pemilihan nilai λ yang tepat sangat penting agar model dapat mencapai keseimbangan optimal antara bias dan variansi, sehingga menghasilkan estimasi yang stabil dan representatif terhadap pola data yang sebenarnya.

Penentuan orde *spline*, jumlah titik knot, serta parameter penghalus optimum dilakukan menggunakan Algoritma *Full-Search*, yaitu prosedur eksplorasi menyeluruh terhadap berbagai kombinasi parameter hingga diperoleh nilai *Generalized Cross Validation* (GCV) minimum (Ruppert dkk., 2003). GCV mengevaluasi kesalahan prediksi dengan mempertimbangkan penalti atas kompleksitas model melalui *trace of smoother matrix* $tr(A_\lambda)$. Nilai GCV minimum menunjukkan kombinasi parameter yang menghasilkan keseimbangan terbaik antara ketepatan dan kelulusan model (Anggraeni dkk., 2022). Adapun fungsi GCV secara umum dapat dituliskan pada Persamaan (18).

$$GCV = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left[1 - \frac{tr(A_\lambda)}{n}\right]^2} \quad (18)$$

Keterangan:

n : banyak pengamatan

$tr(A_\lambda)$: hasil penjumlahan elemen-elemen pada diagonal matriks A_λ .

1.4.4 Heterogenitas Spasial dan Temporal

Data yang terkait dengan suatu lokasi tertentu cenderung tidak homogen dan bervariasi antar lokasi dari unit spasial. Perbedaan antara satu wilayah dengan wilayah lainnya menyebabkan terjadinya heterogenitas spasial. Uji heterogenitas spasial bertujuan untuk mengetahui adanya variasi karakteristik atau keunikan yang bersifat spesifik pada masing-masing lokasi pengamatan. Keberadaan heterogenitas spasial dapat menyebabkan perbedaan parameter regresi di setiap lokasi pengamatan. Salah satu metode yang dapat digunakan untuk menguji heterogenitas spasial adalah uji *Breusch-Pagan* (Mujiarti dkk., 2024). Adapun hipotesis dan statistik uji dari uji *Breusch-Pagan* dituliskan pada persamaan (19):

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

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

Statistik Uji:

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \sim \chi_{\alpha, d}^2 \quad (19)$$

dengan elemen vektor $\mathbf{f}$ adalah

$$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1 \right)$$

Keterangan:

BP : statistik uji *Breusch-Pagan*

$\mathbf{Z}$: matriks berukuran $n \times (d + 1)$ yang berisi variabel prediktor terstandarisasi

ε_i : residual pada pengamatan ke- i , $i = 1, 2, 3, \dots, n$

σ^2 : variansi residual ε_i

Pengambilan keputusan dalam uji *Breusch-Pagan* dilakukan berdasarkan nilai statistik uji *BP* atau *p - value*. Hipotesis H_0 ditolak apabila statistik uji $BP > \chi^2_{\alpha,d}$ atau nilai *p - value* $< \alpha$ yang menunjukkan terjadinya heterogenitas spasial (Mujiarti dkk., 2024).

Sedangkan, uji heterogenitas temporal bertujuan untuk menilai apakah data yang digunakan menunjukkan variasi atau perubahan nilai dari waktu ke waktu. Jika terdapat homogenitas temporal, hal ini mengindikasikan bahwa setiap tahun memiliki karakteristik yang sama, sedangkan jika terdapat heterogenitas, maka setiap tahun dianggap memiliki karakteristik yang berbeda. Pengujian ini penting dilakukan dalam penerapan metode GTWR. Untuk mengevaluasi heterogenitas temporal dapat dilakukan dengan visualisasi *boxplot* (Oktarina dkk., 2024). *Boxplot* digunakan untuk membandingkan karakter distribusi nilai data secara individual atau kelompok kategori dari suatu variabel serta untuk menangkap perubahan *outlier* pada tiap waktu. Perubahan nilai *outlier* dan besar *box* mengindikasikan bahwa terdapat heterogenitas temporal pada data (Aisyiah dkk., 2014).

1.4.5 Geographically Weighted Regression

Geographically Weighted Regression (GWR) merupakan pengembangan dari regresi linier yang diterapkan secara lokal. Metode ini memungkinkan setiap lokasi memiliki parameter regresi yang berbeda karena penaksiran dilakukan pada titik pengamatan tertentu (Sunusi & Subarkah, 2023). GWR memanfaatkan karakteristik geografis setiap wilayah sebagai faktor yang memengaruhi variabel respon dengan memasukkan koordinat geografis ke dalam model. Model yang dihasilkan kemudian digunakan untuk memprediksi besarnya variabel respon berdasarkan parameter pada masing-masing lokasi. Inti dari metode GWR terletak pada kedekatan antarwilayah yang direpresentasikan melalui matriks pembobot sehingga wilayah yang berjarak lebih dekat memperoleh bobot yang lebih besar (Zulhijrah dkk., 2024). Model GWR dituliskan pada Persamaan (20).

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^d \beta_j(u_i, v_i)x_{ji} + \varepsilon_i \quad (20)$$

dengan

$$i = 1, 2, \dots, n; j = 1, 2, \dots, d$$

Keterangan:

- y_i : variabel respon pada lokasi pengamatan ke- i
- (u_i, v_i) : titik koordinat lintang (*latitude*) dan bujur (*longitude*) pada lokasi ke- i
- $\beta_0(u_i, v_i)$: intersep pada lokasi pengamatan ke- i
- $\beta_j(u_i, v_i)$: koefisien regresi variabel prediktor ke- j pada lokasi pengamatan ke- i
- x_{ji} : variabel prediktor ke- j pada lokasi pengamatan ke- i
- ε_i : residual pada lokasi pengamatan ke- i

Estimasi parameter model GWR menggunakan metode *Weighted Least Square* (WLS), yaitu metode kuadrat terkecil dengan memberikan pembobot yang berbeda pada setiap titik lokasi pengamatan (Oktarina dkk., 2024). Estimasi

parameter pada lokasi ke- i dituliskan pada Persamaan (21):

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \quad (21)$$

dengan $\mathbf{W}(u_i, v_i)$ merupakan matriks berukuran $n \times n$ dengan diagonal pembobot spasial pada lokasi pengamatan ke- i . Elemen diagonal tersebut ditentukan berdasarkan kedekatan lokasi ke- i terhadap lokasi lainnya.

1.4.6 Geographically and Temporally Weighted Regression

Geographically and Temporally Weighted Regression (GTWR) merupakan pengembangan dari model GWR yang mampu mengakomodasi heterogenitas spasial (lokasi) sekaligus temporal (waktu). Model ini dibangun dengan menambahkan unsur waktu pada GWR sehingga analisis tidak hanya mempertimbangkan variasi antarwilayah, tetapi juga perubahan dari waktu ke waktu. GTWR memanfaatkan matriks pembobot yang mengintegrasikan dimensi spasial dan temporal untuk menangkap keragaman keduanya secara lebih komprehensif. Bentuk umum model GTWR dengan d variabel prediktor dan satu variabel respon pada koordinat (u_i, v_i, t_i) untuk setiap pengamatan dituliskan pada Persamaan (22) (Oktarina dkk., 2024).

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{j=1}^d \beta_j(u_i, v_i, t_i) x_{ji} + \varepsilon_i \quad (22)$$

dengan

$$i = 1, 2, \dots, n; j = 1, 2, \dots, d$$

Keterangan:

- y_i : variabel respon pada lokasi pengamatan ke- i
- (u_i, v_i, t_i) : koordinat geografis dan waktu lokasi pengamatan ke- i
- $\beta_0(u_i, v_i, t_i)$: intersep pada lokasi pengamatan (u_i, v_i) dan waktu amatan t_i
- $\beta_j(u_i, v_i, t_i)$: koefisien regresi variabel prediktor ke- j pada lokasi dan waktu ke- i
- x_{ji} : variabel prediktor ke- j pada lokasi pengamatan ke- i
- ε_i : residual pada lokasi pengamatan ke- i

Berdasarkan pendekatan WLS, estimasi parameter regresi pada metode GTWR diperoleh dengan meminimalkan Persamaan (23).

$$WLS = \sum_{i=1}^n W(u_i, v_i, t_i) \varepsilon_i^2 \quad (23)$$

Bentuk matriks dari Persamaan (23) dapat dituliskan sebagai berikut:

$$\begin{aligned} WLS &= \boldsymbol{\varepsilon}^T \mathbf{W}(u_i, v_i, t_i) \boldsymbol{\varepsilon} \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W}(u_i, v_i, t_i) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} - \mathbf{y}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} + \\ &\quad \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X}\boldsymbol{\beta} \end{aligned} \quad (24)$$

Koefisien regresi β diperoleh dengan melakukan diferensiasi fungsi WLS pada Persamaan (24) terhadap β^T seperti berikut:

$$\begin{aligned} \frac{\partial WLS}{\partial \beta^T} &= 0 \\ \frac{\partial (\mathbf{y}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} - 2\beta^T \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} + \beta^T \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X} \beta)}{\partial \beta^T} &= 0 \\ 0 - 2\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} + 2\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X} \beta &= 0 \\ \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X} \beta &= \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} \\ \hat{\beta}(u_i, v_i, t_i) &= (\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} \end{aligned} \quad (25)$$

dengan $\mathbf{W}(u_i, v_i, t_i) = \text{diag}(w_{i1}, w_{i2}, \dots, w_{in})$ merupakan matriks pembobot berukuran $n \times n$ pada lokasi pengamatan (u_i, v_i) dan waktu t_i , dengan n adalah jumlah pengamatan (Putra dkk., 2019). Elemen diagonal $w_{ij} (1 \leq j \leq n)$ merepresentasikan fungsi jarak spasial-temporal pada (u_i, v_i, t_i) . Persamaan (26) mendefinisikan fungsi jarak spasial-temporal yang menggabungkan karakteristik fungsi jarak spasial (d_{ij}^S) dan fungsi jarak temporal (d_{ij}^T) (Ma'rifah dkk., 2025).

$$(d_{ij}^{ST})^2 = \gamma (d_{ij}^S)^2 + \mu (d_{ij}^T)^2 \quad (26)$$

Fungsi jarak spasial-temporal dilambangkan dengan (d_{ij}^S) dan (d_{ij}^T) dengan faktor skala γ dan μ yang menyeimbangkan dampak yang berbeda dari pengukuran jarak spasial dan temporal. Sehingga, diperoleh jarak Euclidean spasial-temporal yang dituliskan pada Persamaan (27):

$$\begin{aligned} (d_{ij}^S)^2 &= (u_i - u_j)^2 + (v_i - v_j)^2 \\ (d_{ij}^T)^2 &= (t_i - t_j)^2 \\ (d_{ij}^{ST})^2 &= \gamma [(u_i - u_j)^2 + (v_i - v_j)^2] + \mu [(t_i - t_j)^2] \end{aligned} \quad (27)$$

u_i menunjukkan koordinat garis lintang (*latitude*) pada lokasi ke- i , u_j menunjukkan koordinat garis lintang (*latitude*) pada lokasi ke- j , v_i menunjukkan koordinat garis bujur (*longitude*) pada lokasi ke- i , v_j menunjukkan koordinat garis bujur (*longitude*) pada lokasi ke- j , t_i menunjukkan waktu atau tahun amatan lokasi ke- i , dan t_j menunjukkan waktu atau tahun amatan lokasi ke- j .

Misalkan τ merupakan parameter rasio $\frac{\mu}{\gamma}$ dengan $\gamma \neq 0$, maka diperoleh Persamaan (28):

$$\frac{(d_{ij}^{ST})^2}{\gamma} = [(u_i - u_j)^2 + (v_i - v_j)^2] + \tau (t_i - t_j)^2 \quad (28)$$

Parameter τ berperan dalam memperbesar atau memperkecil efek temporal agar sesuai dengan efek spasial (Ma'rifah dkk., 2025). Penentuan parameter ini dapat dilakukan dengan melihat nilai koefisien determinasi (R^2) melalui proses iterasi berdasarkan inisialisasi nilai awal τ . Selanjutnya, estimasi parameter γ dan μ diperoleh dengan metode iterasi berdasarkan hasil estimasi τ yang menghasilkan R^2 terbesar.

1.4.7 Fungsi Pembobot

Analisis model yang mempertimbangkan aspek spasial memerlukan penerapan pembobot dalam proses perhitungannya (Oktafiani dkk., 2024). Pembobot memiliki peran yang sangat penting dalam model GWR karena nilainya mencerminkan kedekatan lokasi antar data observasi. Proses pembobotan dapat dilakukan dengan berbagai metode, salah satunya menggunakan fungsi kernel. Dalam pembentukan matriks pembobot, elemen diagonal matriks diisi dengan hasil perhitungan kernel baik menggunakan *fixed bandwidth* maupun *adaptive bandwidth* pada setiap lokasi pengamatan, sedangkan elemen lainnya bernilai nol. Berikut merupakan contoh bentuk matriks pembobot spasial untuk kernel *fixed bandwidth* (Pamungkas dkk., 2016):

$$W(u_i, v_i) = \begin{bmatrix} w_{i1}^S & 0 & \cdots & 0 \\ 0 & w_{i2}^S & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in}^S \end{bmatrix} \quad (29)$$

dengan:

$W(u_i, v_i)$: matriks pembobot spasial untuk lokasi pengamatan ke- i

w_{in}^S : nilai dari kernel *fixed bandwidth* untuk data pada titik ke- n dalam pengujian model di lokasi pengamatan ke- i .

Pembobotan pada model GTWR diberikan berdasarkan bobot fungsi kernel seperti yang digunakan pada model GWR, kemudian bobot tersebut diterapkan pada setiap periode waktu. Ada dua jenis fungsi kernel berdasarkan lokasi yaitu fungsi kernel bersifat tetap (*fixed kernel*) dan fungsi kernel bersifat adaptif (*adaptive kernel*). Fungsi *fixed kernel* memiliki *bandwidth* yang sama di setiap titik lokasi observasi, sedangkan fungsi *adaptive kernel* memiliki *bandwidth* yang berbeda di setiap titik lokasi observasi (Martha dkk., 2021). Fungsi pembobot spasial-temporal dapat diperoleh dengan mengkombinasikan jarak spasial-temporal dan fungsi kernel dengan peluruhan jarak. Persamaan fungsi pembobot *fixed kernel* terbagi menjadi beberapa fungsi sebagai berikut (Vianti dkk., 2024):

Fungsi *Gaussian Kernel*

$$w_{ij}^{ST} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}^{ST}}{h^{ST}} \right)^2 \right] \quad (30)$$

Fungsi *Bi-Square Kernel*

$$w_{ij}^{ST} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{ST}}{h^{ST}} \right)^2 \right]^2, & d_{ij}^{ST} \leq h^{ST} \\ 0, & d_{ij}^{ST} > h^{ST} \end{cases} \quad (31)$$

Fungsi *Tricube Kernel*

$$w_{ij}^{ST} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{ST}}{h^{ST}} \right)^3 \right]^3, & d_{ij}^{ST} \leq h^{ST} \\ 0, & d_{ij}^{ST} > h^{ST} \end{cases} \quad (32)$$

Fungsi *Exponential Kernel*

$$w_{ij}^{ST} = \exp \left[-\frac{d_{ij}^{ST}}{h^{ST}} \right] \quad (33)$$

Keterangan:

w_{ij}^{ST} : matriks pembobot spasial-temporal

d_{ij}^{ST} : jarak spasial-temporal antara observasi ke- i dengan observasi ke- j

h^{ST} : parameter *bandwidth* spasial-temporal

Misalnya, digunakan fungsi *Exponential Kernel* dengan peluruhan jarak sebagaimana ditunjukkan pada Persamaan (33). Selanjutnya, perhitungan fungsi pembobot spasial-temporal berdasarkan Persamaan (33) dapat didefinisikan ke dalam Persamaan (34) (Zhang dkk., 2020).

$$w_{ij}^{ST} = \exp \left[-\left(\frac{\sqrt{\gamma [(u_i - u_j)^2 + (v_i - v_j)^2] + \mu (t_i - t_j)^2}}{h^{ST}} \right) \right] \quad (34)$$

1.4.8 Pemilihan *Bandwidth* Optimal

Bandwidth didefinisikan sebagai suatu lingkaran dengan radius h yang berpusat pada titik lokasi pengamatan dan berfungsi sebagai dasar dalam menentukan pembobot untuk setiap titik observasi. Titik-titik pengamatan yang berada di dalam lingkaran dengan radius h dari lokasi ke- i akan memberikan pengaruh dalam proses estimasi parameter pada titik pengamatan tersebut (Purnamasari & Widyaningsih, 2023). Pemilihan *bandwidth* optimum merupakan salah satu hal penting karena akan memengaruhi akurasi hasil estimasi, yaitu mengatur varians dan bias dari model (Putra dkk., 2022). Salah satu metode yang umum digunakan untuk menentukan *bandwidth* optimal adalah metode *Cross Validation* (CV). Metode CV merupakan metode yang memanfaatkan data yang ada untuk menggambarkan apa yang seharusnya dilakukan apabila pengulangan observasi dimungkinkan (Adiningrum dkk., 2018).

Dalam konteks penentuan *bandwidth*, metode CV dilakukan dengan cara mengeluarkan satu titik data ke- i dari proses estimasi y_i untuk kemudian dibandingkan dengan nilai aktualnya. Prosedur ini diulang secara iteratif hingga diperoleh nilai CV yang paling kecil. *Bandwidth* dianggap optimum apabila menghasilkan nilai CV minimum. Secara matematis, nilai CV dapat diperoleh dengan Persamaan (35) (Purnamasari & Widyaningsih, 2023).

$$CV(h^{ST}) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h^{ST}))^2 \quad (35)$$

dengan $\hat{y}_{\neq i}(h^{ST})$ adalah nilai estimasi y_i saat pengamatan ke- i dihilangkan dari estimasi.

1.4.9 Geographically and Temporally Weighted Spline Nonparametric Regression

Model *Geographically and Temporally Weighted Spline Nonparametric Regression* (GTWSNR) merupakan perluasan dari model GTWR. GTWSNR menggabungkan pendekatan *nonparametric spline regression* dengan pembobotan spasial dan temporal secara simultan. Model ini mengintegrasikan informasi geografis dan temporal untuk membentuk fungsi regresi yang tidak diketahui secara eksplisit. Sehingga, GTWSNR mampu menangkap variasi spasial dan temporal secara bersamaan serta memberikan estimasi yang lebih fleksibel terhadap fenomena yang berubah menurut ruang dan waktu. Beberapa penelitian sebelumnya menggunakan *truncated spline* sebagai estimator regresi nonparametrik untuk mengatasi keterbatasan GTWR yang masih mengasumsikan bentuk hubungan tertentu (Sifriyani dkk., 2025). Pada model regresi asumsi yang digunakan adalah error berdistribusi normal dengan mean nol dan variansi $\sigma^2(u_i, v_i, t_i)$ pada setiap lokasi (u_i, v_i) dan waktu (t_i) (Serena dkk., 2021). Secara matematis, hubungan antara variabel respon dan variabel prediktor pada observasi ke- i dalam model GTWSNR dapat dinyatakan melalui Persamaan (36) (Sifriyani dkk., 2025).

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{j=1}^d \sum_{k=1}^m \beta_{jk}(u_i, v_i, t_i) x_{ji}^k + \sum_{j=1}^d \sum_{h=1}^r \beta_{j,m+h}(u_i, v_i, t_i) (x_{ji} - K_{jh})_+^m + \varepsilon_i \quad (36)$$

Keterangan:

- y_i : variabel respon pada lokasi ke- i , dengan $i = 1, 2, \dots, n$
- x_{ji} : variabel prediktor ke- j pada lokasi ke- i , dengan $j = 1, 2, \dots, d$
- K_{jh} : titik knot ke- h pada komponen variabel prediktor ke- j , dengan $h = 1, 2, \dots, r$
- $\beta_{jk}(u_i, v_i, t_i)$: parameter fungsi polinomial ke- k dari variabel prediktor ke- j pada lokasi ke- i
- $\beta_{j,m+h}(u_i, v_i, t_i)$: parameter regresi fungsi *spline* ke- $m + h$ pada titik knot ke- h dan variabel prediktor ke- j pada lokasi pengamatan ke- i

1.4.10 Penalized Weighted Least Square

Metode *Penalized Weighted Least Square* (PWLS) merupakan pendekatan estimasi parameter yang digunakan ketika asumsi ragam konstan pada model regresi tidak terpenuhi, atau terjadi heteroskedastisitas pada residual (Oktarina dkk., 2025). Metode ini merupakan pengembangan dari PLS, yang merupakan kriteria standar dalam proses penghalusan. PWLS mengombinasikan konsep WLS dengan penalti terhadap kompleksitas model, sehingga menghasilkan estimasi parameter yang lebih stabil dan akurat. Kriteria estimasi PWLS secara simultan mempertimbangkan unsur pembobotan serta pemilihan parameter penghalus untuk memperoleh keseimbangan antara fleksibilitas model dan tingkat penghalusan yang optimal. Kriteria ini telah banyak dikembangkan oleh berbagai peneliti untuk diterapkan dalam model regresi nonparametrik, antara lain melalui penggunaan estimator *smoothing spline* dan *penalized spline* (Islamiyati dkk., 2022). Secara umum, bentuk kriteria estimasi pada metode PWLS dapat dinyatakan pada

Persamaan (37) (Anggraeni dkk., 2022).

$$PWLS = \sum_{i=1}^n W(u_i, v_i, t_i) \varepsilon_i^2 + \lambda \sum_{h=1}^r \beta_{j,m+h}^2(u_i, v_i, t_i) \quad (37)$$

Keterangan:

- $W(u_i, v_i, t_i)$: fungsi pembobot spasial-temporal pada lokasi (u_i, v_i) dan waktu t_i , yang menunjukkan kedekatan antara titik pengamatan ke- i
- λ : parameter penghalus variabel prediktor ke- j
- $\beta_{j,m+h}^2(u_i, v_i, t_i)$: parameter regresi fungsi *spline* ke- $(m + h)$ pada titik knot ke- h dan variabel prediktor ke- j pada lokasi pengamatan ke- i

Bentuk matriks dari kriteria PWLS pada Persamaan (37) diberikan pada Persamaan (38).

$$\begin{aligned} PWLS &= \boldsymbol{\varepsilon}^T \mathbf{W}(u_i, v_i, t_i) \boldsymbol{\varepsilon} + \lambda \boldsymbol{\beta}^T(u_i, v_i, t_i) \mathbf{D} \boldsymbol{\beta}(u_i, v_i, t_i) \\ &= (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}(u_i, v_i, t_i))^T \mathbf{W}(u_i, v_i, t_i) (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}(u_i, v_i, t_i)) + \lambda \boldsymbol{\beta}^T(u_i, v_i, t_i) \mathbf{D} \boldsymbol{\beta}(u_i, v_i, t_i) \end{aligned} \quad (38)$$

1.4.11 Pengujian Signifikansi Parameter

Setelah dilakukan estimasi terhadap parameter, selanjutnya dilakukan pengujian signifikansi parameter. Uji signifikansi parameter digunakan untuk mengidentifikasi parameter apa saja yang berpengaruh secara signifikan terhadap variabel respon. Dalam hal ini, pengujian signifikansi parameter dapat dilakukan secara simultan dan parsial.

1. Uji Signifikansi Parameter Simultan

Prosedur pengujian signifikansi parameter secara simultan dalam model dilakukan dengan menggunakan uji F. Uji ini bertujuan untuk mengetahui signifikansi parameter model secara bersama-sama. Prosedur pengujian signifikansi parameter secara simultan dapat dilakukan dengan langkah-langkah berikut (Mukrom dkk., 2021):

Hipotesis:

$$H_0 : \beta_{11}(u_i, v_i, t_i) = \beta_{12}(u_i, v_i, t_i) = \dots = \beta_{dm}(u_i, v_i, t_i) = \beta_{1,m+1}(u_i, v_i, t_i) = \dots = \beta_{d,m+r}(u_i, v_i, t_i) = 0$$

$$H_1 : \text{Setidaknya ada satu } \beta_{jk}(u_i, v_i, t_i) \neq 0 \text{ atau } \beta_{j,m+h}(u_i, v_i, t_i) \neq 0$$

Statistik Uji:

$$F_{hitung} = \frac{MSR}{MSE} = \frac{SSR/d}{SSE/(n - d - 1)} \quad (39)$$

Keterangan:

SSR : jumlah kuadrat regresi

SSE : jumlah kuadrat error

d : banyaknya variabel prediktor

Kriteria Uji:

Tolak H_0 jika nilai $F_{hitung} > F_{(\alpha; db_1; db_2)}$ atau $p - \text{value} < \alpha$.

2. Uji Signifikansi Parameter Parsial

Uji signifikansi parameter secara parsial dilakukan untuk mengetahui pengaruh masing-masing parameter dalam model terhadap variabel respon secara individu. Uji parsial dilakukan setelah model dinyatakan signifikan secara simultan agar dapat diketahui parameter yang memiliki pengaruh nyata dalam model. Prosedur pengujian signifikansi parameter secara parsial dilakukan melalui tahapan berikut (Sifriyani dkk., 2025):

Hipotesis:

$$H_0 : \beta_{jk}(u_i, v_i, t_i) = 0 \text{ dan } \beta_{j,m+h}(u_i, v_i, t_i) = 0$$

$$H_1 : \beta_{jk}(u_i, v_i, t_i) \neq 0 \text{ atau } \beta_{j,m+h}(u_i, v_i, t_i) \neq 0$$

Statistik Uji:

$$t = \frac{\hat{\beta}(u_i, v_i, t_i)}{SE[\hat{\beta}(u_i, v_i, t_i)]} \quad (40)$$

dengan $\hat{\beta}(u_i, v_i, t_i)$ adalah estimator dari $\beta(u_i, v_i, t_i)$ dan komponen $SE[\hat{\beta}(u_i, v_i, t_i)] = \sqrt{g_{kk}}$, dengan g_{kk} merupakan elemen diagonal ke- $(k + 1)$ dari matriks $(\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X} + \lambda \mathbf{D})^{-1} \hat{\sigma}^2(u_i, v_i, t_i)$.

Kriteria Uji:

Tolak H_0 jika nilai $|t_{hitung}| > t_{(\frac{\alpha}{2}, (n-1))}$ atau $p\text{-value} < \alpha$.

1.4.12 Ukuran Keakuratan Model

Keakuratan model NGTWR dengan pendekatan *penalized spline* dalam penelitian ini diukur menggunakan *Mean Square Error* (MSE), seperti yang digunakan dalam penelitian Sifriyani dkk. (2025). MSE merupakan nilai rata-rata kesalahan kuadrat antara nilai aktual dan nilai hasil estimasi. Metode MSE biasanya digunakan untuk mengevaluasi metode pengukuran dalam model regresi. Model terbaik ditentukan berdasarkan nilai MSE terkecil, yang menunjukkan tingkat kesalahan prediksi paling rendah. Perhitungan nilai MSE disajikan pada Persamaan (41).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (41)$$

Keterangan:

y_i : nilai variabel respon

$\hat{y}_i$: nilai estimasi variabel respon

1.4.13 Tingkat Pengangguran Terbuka

Pengangguran adalah penduduk usia 15 tahun ke atas yang tidak bekerja namun sedang mencari pekerjaan, mempersiapkan usaha baru, sudah diterima bekerja/sudah siap berusaha tetapi belum mulai bekerja/berusaha, atau merasa tidak mungkin mendapatkan pekerjaan (putus asa). Tingkat Pengangguran Terbuka (TPT) merupakan indikator yang digunakan untuk mengukur tenaga kerja yang tidak terserap oleh pasar kerja dan menggambarkan kurang termanfaatkannya pasokan tenaga kerja. TPT menggambarkan sejauh mana penduduk usia kerja yang aktif mencari pekerjaan tetapi belum memperoleh pekerjaan. Dengan kata

lain, TPT menjadi salah satu ukuran penting dalam menilai kondisi ketenagakerjaan dan efisiensi pasar tenaga kerja di suatu daerah. Semakin tinggi nilai TPT, semakin besar ketidakseimbangan antara penawaran dan permintaan tenaga kerja di suatu wilayah (Badan Pusat Statistik, 2025). Rumus perhitungan TPT seperti pada Persamaan (42):

$$TPT = \frac{\text{Jumlah Pengangguran}}{\text{Jumlah Angkatan Kerja}} \times 100\% \quad (42)$$

dengan, Jumlah Pengangguran adalah penduduk usia kerja yang tidak bekerja tetapi sedang mencari pekerjaan, mempersiapkan usaha, atau merasa tidak mungkin mendapat pekerjaan. Sementara itu, Jumlah Angkatan Kerja mencakup penduduk usia kerja yang bekerja maupun yang sedang menganggur.

Berdasarkan penelitian-penelitian terdahulu, beberapa faktor yang diduga memengaruhi TPT adalah sebagai berikut:

1. Jumlah Penduduk Miskin merupakan salah satu faktor penting yang memengaruhi tingkat pengangguran. Penduduk miskin sering kali memiliki tingkat pendidikan yang rendah dan keterbatasan keterampilan, sehingga sulit untuk bersaing di pasar tenaga kerja formal dan cenderung bekerja di sektor informal yang tidak stabil serta berpendapatan rendah. Semakin besar jumlah penduduk miskin, semakin tinggi potensi terjadinya pengangguran akibat keterbatasan akses terhadap pendidikan dan pelatihan kerja yang dibutuhkan untuk meningkatkan daya saing tenaga kerja. Kondisi ini memperpanjang ketidakmampuan individu untuk memperoleh pekerjaan yang layak, sehingga jumlah penduduk miskin perlu dipertimbangkan dalam analisis TPT karena berhubungan dengan kualitas sumber daya manusia dan keterbatasan kesempatan kerja (Abdi dkk., 2024).
2. Kepadatan Penduduk merupakan indikator yang menunjukkan jumlah penduduk per satuan luas wilayah dan mencerminkan tingkat tekanan terhadap sumber daya ekonomi serta lapangan kerja di suatu daerah. Tingginya kepadatan penduduk sering kali berkaitan dengan meningkatnya persaingan dalam memperoleh pekerjaan dan terbatasnya akses terhadap fasilitas publik yang memadai. Kondisi ini dapat memperburuk ketimpangan ekonomi dan memperbesar peluang terjadinya pengangguran (Tervia dkk., 2022).
3. Harapan Lama Sekolah merupakan indikator yang mencerminkan rata-rata jumlah tahun pendidikan yang diharapkan dapat ditempuh oleh anak usia sekolah di suatu wilayah. Angka harapan lama sekolah dihitung untuk penduduk berusia 7 tahun ke atas. Nilai harapan lama sekolah yang tinggi menunjukkan adanya peluang bagi masyarakat untuk memperoleh pendidikan yang lebih baik dan keterampilan yang lebih luas. Peningkatan indikator ini dapat menurunkan risiko pengangguran karena tenaga kerja menjadi lebih siap dan kompetitif dalam pasar kerja. Oleh karena itu, harapan lama sekolah menjadi faktor penting dalam analisis tingkat pengangguran terbuka (Siskawati dkk., 2021).
4. Laju pertumbuhan Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan digunakan untuk mengukur pertumbuhan ekonomi suatu wilayah dari periode ke periode, baik secara tahunan maupun triwulanan. PDRB harga

konstan dihitung menggunakan tahun dasar 2010, sehingga mencerminkan pertumbuhan ekonomi riil tanpa dipengaruhi oleh perubahan harga. Indikator ini dapat digunakan untuk menunjukkan laju pertumbuhan ekonomi secara keseluruhan maupun menurut sektor usaha dari tahun ke tahun. PDRB yang meningkat menunjukkan adanya pertumbuhan aktivitas ekonomi dan potensi penyerapan tenaga kerja yang lebih besar. Meskipun demikian, peningkatan PDRB tidak selalu diikuti oleh penurunan tingkat pengangguran apabila pertumbuhan tersebut tidak disertai pemerataan kesempatan kerja. Oleh karena itu, laju pertumbuhan PDRB penting diperhitungkan dalam analisis TPT sebagai indikator kapasitas ekonomi daerah dalam menyediakan lapangan kerja (Darmawan & Mifrahi, 2022).

Berdasarkan data Badan Pusat Statistik (2025), TPT nasional pada Februari 2025 tercatat sebesar 4,76%, menurun dibandingkan dengan Februari 2024 yaitu sebesar 4,91%. Penurunan ini menunjukkan adanya perbaikan kondisi ketenagakerjaan terutama pada sektor informal dan industri pengolahan. Namun, di Provinsi Sulawesi Selatan, TPT pada Februari 2025 justru mencapai 4,96%, meningkat sebesar 0,06 persen poin dibandingkan Februari 2024, yang mengindikasikan masih adanya tantangan dalam menciptakan lapangan kerja di daerah tersebut (BPS Sulsel, 2025). TPT yang tinggi dapat menimbulkan berbagai dampak negatif terhadap perekonomian dan sosial masyarakat. Dampak tersebut meliputi penurunan pendapatan dan daya beli masyarakat, peningkatan angka kemiskinan dan ketimpangan ekonomi, munculnya masalah sosial seperti kriminalitas dan ketidakstabilan sosial, serta penurunan produktivitas nasional.

BAB II METODOLOGI PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan yang dapat diakses pada laman <https://sulsel.bps.go.id/id>. Unit pengamatan yang digunakan adalah tingkat Kabupaten/Kota di Sulawesi Selatan yang terdiri dari 24 kabupaten/kota dalam rentang periode 2022-2024 sehingga dalam penelitian ini terdapat 72 pengamatan. Data yang digunakan dalam penelitian ini dapat dilihat pada Lampiran 1.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon dan variabel prediktor yang diuraikan pada Tabel 1 sebagai berikut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Definisi Operasional
Y	Tingkat Pengangguran Terbuka	Persentase jumlah pengangguran terhadap jumlah angkatan kerja. Angkatan Kerja adalah penduduk usia kerja (15 tahun ke atas) yang bekerja atau punya pekerjaan namun sementara tidak bekerja, dan pengangguran.
X_1	Jumlah Penduduk Miskin	Jumlah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan.
X_2	Kepadatan Penduduk per km ²	Jumlah rata-rata penduduk per kilometer persegi wilayah.
X_3	Harapan Lama Sekolah	Lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang.
X_4	Laju Pertumbuhan Produk Domestik Regional Bruto Atas Dasar Harga Konstan	Persentase pertumbuhan produksi barang dan jasa di suatu wilayah perekonomian dalam selang waktu tertentu yang dihitung menggunakan harga konstan tahun 2010, sehingga menggambarkan pertumbuhan ekonomi riil tanpa dipengaruhi oleh perubahan harga (inflasi).

2.3 Metode Analisis

Pendekatan analisis yang digunakan dalam penelitian ini adalah regresi nonparametrik spasial-temporal dengan metode NGTWR-PS. Selanjutnya, dilakukan pengolahan dan analisis data menggunakan *software* R dan QGIS. Tahapan analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan estimasi parameter model NGTWR-PS pada Persamaan (36) menggunakan metode *Penalized Weighted Least Square* (PWLS) pada Persamaan (38).
2. Memperoleh model NGTWR-PS pada faktor-faktor yang memengaruhi TPT di Provinsi Sulawesi Selatan.
 - a. Melakukan eksplorasi data menggunakan peta tematik untuk mendapatkan gambaran awal mengenai pola sebaran variabel respon.
 - b. Melakukan analisis hubungan linier melalui *scatter plot* antara variabel respon dan masing-masing variabel prediktor untuk mengidentifikasi pola hubungan antarvariabel.
 - c. Melakukan uji multikolinearitas dengan melihat nilai *Variances Inflation Factor* (VIF) berdasarkan persamaan sebagai berikut (Sriningsih dkk., 2018):

$$VIF = \frac{1}{1 - R_j^2}$$

dengan R_j^2 adalah koefisien determinansi variabel ke- j . Jika nilai VIF > 10 menunjukkan adanya multikolinearitas.

- d. Melakukan uji heterogenitas spasial menggunakan uji *Breusch-Pagan* pada Persamaan (19), serta menguji heterogenitas temporal melalui visualisasi *boxplot* untuk melihat variasi antarwaktu.
- e. Menentukan jumlah orde, titik knot, dan parameter penghalus yang optimum dengan menggunakan metode *Generalized Cross Validation* (GCV) sesuai persamaan (18).
- f. Menentukan nilai *bandwidth* optimum dan fungsi pembobot menggunakan nilai CV minimum berdasarkan Persamaan (35).
- g. Menentukan parameter rasio spasial temporal (τ), parameter spasial (γ) dan parameter temporal (μ).
- h. Menghitung jarak spasial temporal (d_{ij}^{ST}) menggunakan Persamaan (27).
- i. Membentuk matriks pembobot spasial-temporal menggunakan fungsi pembobot spasial temporal.
- j. Mengestimasi parameter model NGTWR-PS.
- k. Melakukan pengujian signifikansi parameter secara simultan menggunakan Persamaan (39), dan uji parsial menggunakan Persamaan (40).
- l. Melakukan pengujian keakuratan model berdasarkan nilai *Mean Square Error* (MSE) menggunakan Persamaan (41).
- m. Melakukan interpretasi model NGTWR-PS.