# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Regresi linear merupakan suatu metode statistik yang digunakan untuk menganalisis hubungan sebab-akibat yang dinyatakan dalam bentuk fungsi matematik antara variabel prediktor $(X)$, dengan variabel respon $(Y)$ (Sa'adah, 2024). Terdapat tiga jenis pendekatan untuk mengestimasi fungsi regresi, yaitu parametrik, nonparametrik, dan semiparametrik. Pada regresi parametrik, diasumsikan pola hubungan yang terbentuk antara variabel prediktor dan variabel respon diketahui, sedangkan pada regresi nonparametrik, pola hubungan yang terbentuk diasumsikan tidak diketahui (Femadiyanti dkk., 2020). Dalam kasus regresi multivariabel, ketika pola hubungan salah satu variabel sudah diketahui sementara pola variabel lainnya tidak diketahui maka digunakan pendekatan regresi semiparametrik. Pendekatan ini mengintegrasikan elemen dari regresi parametrik dan nonparametrik untuk memberikan analisis yang lebih fleksibel (Fadlirhohim dkk., 2024). Bentuk kurva regresi yang tidak diketahui pada data menyebabkan pendekatan parametrik tidak dapat digunakan, sehingga untuk mengatasi permasalahan tersebut dapat menggunakan pendekatan nonparametrik.

Pendekatan nonparametrik adalah metode yang tidak memerlukan syarat-syarat tertentu yang harus dipenuhi oleh data. Pendekatan ini lebih sesuai digunakan ketika asumsi-asumsi dalam regresi parametrik tidak terpenuhi, terutama ketika bentuk hubungan antara variabel tidak diketahui atau bersifat kompleks. Regresi nonparametrik telah banyak dikembangkan antara lain menggunakan *spline*, *kernel*, polinomial lokal, dan deret *fourier* (Syam dkk., 2019). Salah satu metode regresi nonparametrik yang menawarkan fleksibilitas tinggi dalam memodelkan hubungan tidak diketahui adalah spline. Metode ini membagi data menjadi segmen-segmen polinomial dan memiliki fleksibilitas untuk menyesuaikan berbagai bentuk kurva yang kompleks. (Susnawati dkk., 2019).

Metode regresi *spline* bergantung pada titik knot, yaitu titik peralihan yang menunjukkan perubahan pola perilaku dari suatu fungsi pada interval yang berbeda. Dengan kata lain, metode ini sangat fleksibel dalam mengikuti bentuk data yang sebenarnya, sehingga kurva regresi yang dihasilkan dapat lebih akurat. Jenis-jenis *spline* yang umum digunakan yaitu*, B-spline, truncated, smoothing,* dan *penalized*. Kemampuan fleksibilitas regresi *spline* sangat berguna untuk menangkap pola data yang kompleks. Namun, hal ini menjadi kendala karena regresi *spline* cenderung berisiko mengalami *overfitting*, yaitu model cenderung mengikuti data, sehingga tidak dapat memberikan prediksi yang akurat pada data baru, terutama ketika menggunakan banyak titik *knot*. Salah satu cara untuk mengatasi permasalahan ini adalah menggunakan pendekatan regresi nonparametrik *spline penalized* (Apriani, 2017).

*Spline penalized* adalah metode estimasi dalam regresi *spline* yang didapatkan dengan cara meminimalkan fungsi *Penalized Least Square* (PLS). *Spline* *penalized* memiliki parameter *smoothing* yang merupakan parameter penghalus pengontrol keseimbangan antara kecocokan data dan kemulusan kurva (Kurniasari dkk., 2019). Parameter penghalus dapat mencegah terjadinya *overfitting* namun *spline* tetap dapat menyesuaikan pola yang ada dalam data (Maulida, 2023). Penelitian sebelumnya yang dilakukan oleh Octavia dkk. (2024) telah menunjukkan bahwa metode regresi *spline penalized* efektif dalam memprediksi pergerakan Indeks Harga Saham Gabungan (IHSG) selama periode Januari 2020 hingga Desember 2022. Model ini memperoleh nilai *R-squared* sebesar 92,1% yang mengindikasikan bahwa model tersebut mampu menjelaskan sebagian besar variabilitas data IHSG. Berdasarkan hal tersebut, model *spline penalized* efektif dalam memodelkan hubungan tidak diketahui antara variabel dan mampu menangkap pola yang kompleks dalam data.

Ketika penelitian berfokus pada suatu wilayah maka peluang bentuk kurva regresi untuk tidak diketahui semakin besar karena adanya perbedaan karakteristik setiap wilayah. Perbedaan lokasi wilayah dapat memberikan data yang bersifat spasial yang dikenal dengan istilah heterogenitas spasial (Ardhani dkk., 2023). Salah satu metode yang dapat digunakan ketika terdapat efek heterogenitas spasial yakni *Geographically Weighted Regression* (GWR). Metode ini memberikan bobot yang berbeda untuk setiap observasi berdasarkan lokasinya, serta memungkinkan identifikasi pola hubungan yang bervariasi secara spasial antara variabel. Model ini menghasilkan estimasi parameter yang bersifat lokal atau berbeda-beda untuk setiap titik lokasi (Daulay & Simamora, 2023). Meskipun model ini efektif dalam mengakomodasi heterogenitas spasial, model ini memiliki keterbatasan yang terletak pada asumsi linearitas lokal yang membatasi kemampuannya dalam menangkap pola data yang kompleks. Permasalahan tersebut dapat diatasi melalui pengembangan lebih lanjut dengan mengintegrasikan GWR dengan regresi nonparametrik *spline*, yang akan menghasilkan model yang lebih fleksibel dan akurat dalam menangkap pola data spasial yang kompleks (Fitri dkk., 2019).

Beberapa penelitian sebelumnya yang telah menggunakan regresi *spline* dalam model GWR di antaranya, yaitu Rifaldi (2023) menggunakan GWR dengan pendekatan *spline truncated* orde 1 pada data Kasus Balita Gizi Kurang di Provinsi Sulawesi Selatan dan Serena dkk. (2021) menggunakan pendekatan *spline truncated* orde 2 dalam model GWR pada data pencemaran daerah aliran Sungai Mahakam. Penelitian-penelitian sebelumnya menunjukkan bahwa pengembangan regresi *spline* dalam model GWR menjadi pendekatan yang efektif untuk memodelkan data spasial yang memiliki pola hubungan yang kompleks. Penelitian ini mengintegrasikan keunggulan GWR dalam mengakomodasi heterogenitas spasial dengan fleksibilitas regresi nonparametrik *spline penalized* dalam menangkap pola tidak diketahui. Pendekatan ini dipilih karena dapat menangani data yang memiliki pola hubungan yang tidak diketahui dan dapat memperhitungkan perbedaan yang ada di berbagai lokasi. Selain itu, pendekatan ini memiliki kemampuan untuk menghindari risiko *overfitting* ketika menggunakan banyak titik *knot*. Penelitian sebelumnya telah menerapkan regresi *spline* dalam model GWR di bidang kesehatan dan lingkungan. Oleh karena itu, penelitian ini akan menganalisis bidang lainnya, yaitu bidang ekonomi terkait persentase penduduk miskin.

Persentase penduduk miskin berfungsi sebagai indikator penting untuk mengukur tingkat kemiskinan suatu wilayah, yang dihitung berdasarkan proporsi penduduk dengan pengeluaran per kapita di bawah garis kemiskinan yang ditetapkan (Badan Pusat Statistik, 2024). Penelitian sebelumnya oleh Maulida (2023) tentang persentase penduduk miskin di Provinsi Jawa Barat menggunakan metode regresi *spline penalized* membuktikan bahwa dapat dilakukan penelitian terhadap persentase penduduk miskin. Adapun persentase penduduk miskin di Provinsi Sulawesi Selatan pada maret 2023 tercatat sebesar 8,7%, dengan kabupaten/kota yang memiliki angka tertinggi adalah Pangkep sebesar 13,4%, sedangkan kabupaten/kota dengan angka terendah yaitu Makassar sebesar 5,07%. Variasi persentase penduduk miskin di Provinsi Sulawesi Selatan menunjukkan ketidakmerataan akibat perbedaan karakteristik daerah. Karakteristik ini dapat mencakup faktor-faktor seperti akses terhadap layanan kesehatan, infrastruktur, serta kebijakan pemerintah yang berbeda-beda di setiap wilayah. Hal ini menciptakan pola hubungan yang kompleks antara faktor-faktor yang memengaruhi kemiskinan, serta menunjukkan pola tidak diketahui yang memerlukan pendekatan analisis yang lebih fleksibel. Berdasarkan hal tersebut, maka peneliti akan melakukan penelitian dengan judul "**Pemodelan *Geographically Weighted Nonparametric* Regression Dengan *Spline Penalized* (Studi Kasus: Persentase Penduduk Miskin di Provinsi Sulawesi Selatan Tahun 2023)"**.

## 1.2 Batasan Penelitian

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data persentase penduduk miskin di Sulawesi Selatan tahun 2023.
2. Pendekatan *spline* yang digunakan yaitu *Spline penalized* dengan pemilihan titik knot dan parameter penghalus optimal menggunakan metode *Generalized Cross Validation* (GCV).
3. Penentuan *bandwidth* optimal dilakukan dengan meminimumkan *Cross Validation* (CV).

## 1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter model *Geographically Weighted Nonparametric Regression* dengan *Spline Penalized*.
2. Memperoleh model *Geographically Weighted Nonparametric Regression* dengan *Spline Penalized* pada data persentase penduduk miskin di Provinsi Sulawesi Selatan.

Berdasarkan tujuan penelitian di atas maka setelah melakukan penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Sebagai sumber pengetahuan mengenai tahapan estimasi parameter pada model *Geographically Weighted Nonparametric Regression* dengan *Spline Penalized*.
2. Sebagai sumber pengetahuan dan informasi mengenai pemodelan persentase penduduk miskin berdasarkan 24 kabupaten/kota di Provinsi Sulawesi Selatan sehingga bermanfaat bagi pemerintah untuk meningkatkan kesejahteraan masyarakat.

## 1.4 Teori

### 1.4.1 Analisis Regresi

Analisis regresi adalah metode analisis yang digunakan untuk mengeksplorasi pola hubungan antara variabel prediktor $(x_1, x_2, \dots, x_p)$ dan variabel respon $(y)$ pada suatu data. Pola hubungan antar variabel dapat dinyatakan melalui bentuk model regresi. Bentuk umum persamaan regresi linear untuk data berpasangan $(x_i, y_i)$ dengan $i = 1,2, \dots, n$ adalah sebagai berikut (Khotijah, 2020):

$$y_i = \beta_0 + \sum_{p=1}^{l} \beta_p x_{pi} + \varepsilon_i \qquad (1)$$

Atau dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2)$$

Keterangan:

$y_i$ : variabel respon pada pengamatan ke-$i$
$x_{pi}$ : variabel prediktor ke-$p$ pada pengamatan ke-$i$; $p = 1,2, \dots, l$
$n$ : banyaknya pengamatan
$l$ : banyaknya variabel prediktor
$\beta_0$ : nilai intersep model regresi
$\beta_p$ : nilai koefisien parameter variabel prediktor ke-$p$
$\varepsilon_i$ : nilai *error* untuk pengamatan ke-$i$

Berdasarkan persamaan nilai estimasi yang diharapkan untuk prediksi $\hat{y}_i$ pada titik $x_i$ dinyatakan sebagai berikut (Nurani dkk., 2023) :

$$\hat{y}_i = b_0 + \sum_{p=1}^{l} b_p x_{pi} \qquad (3)$$

Keterangan:

$\hat{y}_i$ : nilai estimasi variabel respon pada pengamatan ke-$i$
$b_0$ : nilai intersep model regresi
$b_p$ : nilai koefisien parameter variabel prediktor ke-$p$

Pola hubungan antara variabel respon dan variabel prediktor dapat dianalisis menggunakan *scatter plot*. Namun, hubungan antara variabel tidak selalu mengikuti pola parametrik. Dalam beberapa kasus, satu atau lebih variabel prediktor memiliki bentuk kurva regresi tidak diketahui sehingga diperlukan penggunaan regresi nonparametrik (Khotijah, 2020).

### 1.4.2 Heterogenitas Spasial

Heterogenitas spasial menggambarkan kondisi ketika hubungan antara variabel prediktor dan variabel respon berbeda-beda di lokasi yang berbeda dalam suatu wilayah penelitian. Hal ini disebabkan oleh adanya perbedaan karakteristik antar lokasi, seperti faktor lingkungan, sosial, ekonomi, budaya, atau faktor geografis lainnya. Model regresi konvensional mengasumsikan bahwa hubungan antara variabel bersifat konstan di seluruh wilayah penelitian (homogenitas) yang disebut sebagai model regresi global. Heterogenitas spasial pada model regresi dapat menghasilkan parameter regresi yang berbeda-beda di setiap lokasi pengamatan. Heterogenitas spasial dapat diidentifikasi menggunakan pengujian *Breusch-Pagan* (Daulay & Simamora, 2023). Uji *Breusch-Pagan* memiliki hipotesis sebagai berikut:
$H_0$ : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$ (tidak terdapat heterogenitas spasial)
$H_1$ : minimal ada satu $\sigma_i^2 \neq \sigma^2$ (terdapat heterogenitas spasial)

Statistik uji *Breusch-Pagan* dinyatakan dalam persamaan berikut:

$$BP = \frac{1}{2}\boldsymbol{f}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{f} \tag{4}$$

Keterangan:
$BP$ : nilai *Breusch-Pagan*
$f$ : vektor *error* dari model regresi dengan $f_i = \frac{\varepsilon_i^2}{\sigma^2} - 1$, dan $i = 1,2,\dots,n$.
$\sigma^2$ : ragam *error*
$Z$ : Matriks $X$ yang distandarisasi berukuran $n \times (l+1)$.

Kriteria Uji:
Tolak $H_0$ apabila nilai $BP > \chi^2_{\alpha,l}$ atau $p-value < \alpha$.

### 1.4.3 Regresi Nonparametrik Regresi *Spline penalized*

Regresi nonparametrik adalah salah satu metode regresi yang tidak memerlukan syarat-syarat tertentu yang harus dipenuhi oleh data, salah satunya pola hubungan yang tidak diketahui atau bersifat kompleks. Dengan menggunakan pendekatan ini, kurva yang dihasilkan hanya diasumsikan mulus (*smooth*) dan terletak dalam ruang fungsi tertentu, sehingga memberikan fleksibilitas yang tinggi dalam memodelkan data. Model regresi nonparametrik secara umum dinyatakan sebagai berikut (Salam dkk., 2022):

$$y_i = f(x_i) + \varepsilon_i \tag{5}$$

Keterangan:

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

Salah satu metode regresi nonparametrik yang fleksibel dalam memodelkan hubungan yang tidak diketahui adalah *spline*. Metode ini menggunakan potongan polinomial untuk memodelkan data. Dengan menghubungkan potongan (*truncated*) polinomial ini melalui titik-titik knot, *spline* dapat menyesuaikan diri dengan berbagai bentuk kurva, sehingga efektif untuk memodelkan hubungan yang kompleks dan tidak diketahui sebelumnya. Titik knot adalah titik-titik yang digunakan untuk membentuk dan menyesuaikan bentuk kurva *spline* (Ferryan, 2023). Secara umum, bentuk fungsi regresi *spline* yaitu (Kurniasari dkk., 2019):

$$f(x_i) = \sum_{m=0}^{q} \beta_m x_i^m + \sum_{k=1}^{r} \beta_{q+k}(x_i - K_k)_+^q \tag{6}$$

Keterangan:

| | |
|---|---|
| $m$ | : orde dengan $m = 0, 1, 2, \ldots, q$ |
| $k$ | : jumlah knot dengan $k = 1, 2, \ldots, r$ |
| $K$ | : letak titik knot ke-$k$ |
| $(x_i - K_k)_+^q$ | : fungsi *truncated* pada variabel ke-$i$ dengan letak titik knot $K_k$ dan orde $q$. |

dengan fungsi *truncated* sebagai berikut:

$$(x_i - K_k)_+^q = \begin{cases} (x_i - K_k)_+^q & , x_{ji} > K_k \\ 0 & , x_{ji} \leq K_k \end{cases}$$

Persamaan regresi nonparametrik *spline* dapat diperoleh melalui subtitusi Persamaan (6) ke dalam Persamaan (5) sehingga diperoleh persamaan sebagai berikut:

$$y_i = \sum_{m=0}^{q} \beta_m x_i^m + \sum_{k=1}^{r} \beta_{q+k}(x_i - K_k)_+^q + \varepsilon_i \tag{7}$$

Permasalahan yang sering terjadi pada model regresi *spline* adalah *overfitting*, yaitu model cenderung mengikuti data. Hal ini dapat terjadi ketika jumlah titik knot yang digunakan terlalu banyak, sehingga model menjadi sangat kompleks. Meskipun model tersebut dapat memberikan hasil yang baik pada data, kemampuan prediksinya pada data baru menjadi terbatas. Akibatnya, model tidak dapat memberikan prediksi yang akurat, terutama ketika dihadapkan pada variasi yang tidak terlihat dalam data baru. Salah satu metode *spline* yang dapat mengatasi permasalahan ini adalah *spline penalized.*

*Spline penalized* juga dikenal sebagai *P-spline* adalah metode statistik yang mengintegrasikan fleksibilitas dari model nonparametrik dengan kemampuan generalisasi yang efektif. Metode ini beroperasi dengan mencari kurva yang paling sesuai dengan data sambil tetap mempertahankan kehalusan kurva tersebut. (Ferryan, 2023). *Spline penalized* diperoleh dengan meminimumkan fungsi estimasi dari *Penalized Least Square* (PLS), yang merupakan kriteria estimasi model yang menggabungkan antara fungsi penalti dengan *goodness of fit*. Fungsi penalti memiliki parameter penghalus ($\lambda$) yang diterapkan pada model bertujuan untuk mencegah terjadinya *overfitting* (Islamiyati & Herdiani, 2019). Fungsi PLS adalah sebagai berikut:

$$PLS = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \sum_{k}^{r} \beta_{q+k}^2 \tag{8}$$

dengan $\lambda$ sebagai parameter penghalus yang berfungsi untuk mengontrol tingkat kemulusan kurva yang dihasilkan dengan nilai $\lambda \geq 0$. Sebelum meminimumkan fungsi PLS, Persamaan (8) akan diubah ke dalam bentuk matriks sebagai berikut:

$$\begin{aligned} PLS &= (\boldsymbol{Y} - \boldsymbol{X\beta})^T(\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{D\beta} \\ &= \boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta} + \lambda\boldsymbol{\beta}^T\boldsymbol{D\beta} \end{aligned} \tag{9}$$

dengan $\boldsymbol{D}$ didefinisikan sebagai matriks diagonal sebagai berikut:

$$\boldsymbol{D} = \begin{bmatrix} 0_{(q+1)\times(q+1)} & 0_{(q+1)\times r} \\ 0_{r\times(q+1)} & I_{r\times r} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{I} \end{bmatrix}$$

dengan $\boldsymbol{I}$ merupakan matriks identitas untuk $\beta_{q+1}, \beta_{q+2}, \dots, \beta_{q+k}$.

Koefisien regresi $\beta$ diperoleh melalui differensiasi dari fungsi PLS pada Persamaan (9) terhadap $\boldsymbol{\beta}^T$ sebagai berikut:

$$\frac{\partial PLS}{\partial \boldsymbol{\beta}^T} = 0$$

$$\frac{\partial(\boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta} + \lambda\boldsymbol{\beta}^T\boldsymbol{D\beta})}{\partial \boldsymbol{\beta}^T} = 0$$

$$0 - \boldsymbol{2X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X\beta} + 2\lambda\boldsymbol{D\beta} = \mathbf{0}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{D})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{10}$$

Selanjutnya subtitusi nilai $\widehat{\boldsymbol{\beta}}$ pada model sebagai berikut:

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{D})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{11}$$

atau dapat dituliskan sebagai berikut:

$$\widehat{\boldsymbol{y}} = \boldsymbol{H}(\lambda)\boldsymbol{Y} \tag{12}$$

dengan

$$\boldsymbol{H}(\lambda) = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{D})^{-1}\boldsymbol{X}^T \tag{13}$$

Keterangan:
$\boldsymbol{H}(\lambda)$ : matriks hat yang memuat parameter penghalus.

### 1.4.4 Titik Knot dan Parameter Penghalus Optimal

Persamaan regresi *spline* memiliki fungsi *truncated* yang merupakan potongan polinomial yang tersegmentasi dan kontinu. Fungsi *truncated* bertujuan untuk menunjukkan variasi pada interval tertentu, sehingga memberikan fleksibilitas dalam pemodelan hubungan antara variabel prediktor dan respon. Pemilihan jumlah dan titik knot merupakan fokus paling penting di dalam regresi *spline* (Kusunartutik & Dwidayati, 2022). Pemilihan jumlah titik knot optimal pada regresi *spline truncated* melibatkan perhitungan kombinasi yang sangat banyak, sebanding dengan jumlah data. Hal ini menyebabkan kebutuhan waktu dan memori yang lebih besar saat menggunakan perangkat lunak. Regresi *spline penalized* menawarkan alternatif dalam penentuan jumlah knot.

Pada regresi *spline penalized*, posisi titik knot ditentukan pada kuantil dari nilai unik (tunggal) variabel prediktor dan terdapat parameter penghalus pada modelnya. Jika digunakan sebanyak $k$ titik knot, maka nilai kuantil yang berfungsi sebagai titik knot akan membagi keseluruhan data pengamatan menjadi $(k + 1)$ bagian yang sama besar, yaitu kuantil ke-$\frac{m}{k+1}$ dari *unique* dari variabel prediktor $\{x_i\}_{i=1}^{n}$ dengan $m = 1,2, \ldots, k$. Jumlah maksimal jumlah knot yang dihitung dibatasi dengan ketentuan $k < (n_{unique} - q - 1)$ dengan $n_{unique}$ adalah banyaknya nilai *unique* dari variabel prediktor (Nurfadhillah, 2023).

Parameter penghalus $(\lambda)$ pada *spline penalized* memiliki pengaruh besar pada model regresi karena berperan sebagai pengontrol keseimbangan antara kemulusan fungsi regresi terhadap data. Penggunaan nilai $\lambda$ yang semakin besar menyebabkan hasil estimasi fungsi yang semakin mulus, sedangkan nilai $\lambda$ yang semakin kecil akan menyebabkan estimasi fungsi semakin fluktuatif (Maulida, 2023). Titik knot dan parameter penghalus merupakan parameter yang sangat penting dalam estimasi model *spline penalized*. Oleh karena itu, penentuan jumlah dan letak titik knot, serta parameter penghalus yang digunakan dalam mengestimasi model *spline penalized* diharapkan memilih nilai yang optimal agar memperoleh hasil estimasi fungsi yang optimal.

Penentuan orde, jumlah titik knot, dan parameter penghalus dapat diperoleh melalui Algoritma *Full-Search.* Dalam algoritma ini, jumlah knot, orde, serta parameter penghalus akan dihitung hingga memiliki nilai *Generalized Cross Validation* (GCV) minimum (Ruppert dkk., 2003). GCV merupakan salah satu metode untuk memperoleh nilai optimal melalui pengukuran kesalahan prediksi dan kompleksitas model. Metode ini merupakan modifikasi dari *Cross Validation (CV).* Nilai GCV dihitung dengan menjumlahkan kuadrat *error* yang telah dikoreksi dengan kuadrat faktor-faktor tertentu. GCV menggunakan metode dalam memilih model berdasarkan pada kemampuan prediksi dari model tersebut. Fungsi GCV dapat dinyakan sebagai berikut (Pratiwi, 2017):

$$GCV = \frac{MSE}{(1 - n^{-1}\left(trace(\boldsymbol{H}(\lambda))\right)^2} \qquad (14)$$

dengan

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 \qquad (15)$$

Keterangan:

$MSE$ : *Mean Square Error* atau rata-rata kuadrat *error*.

### 1.4.5 Model Geographically Weighted Regression

*Geographically Weighted Regression* (GWR) merupakan pengembangan metode dari regresi linear yang berguna untuk mengolah data spasial dengan memberikan pembobot yang berbeda untuk setiap lokasi amatan. Model ini mengakomodasi heterogenitas spasial dan menghasilkan penaksir parameter model yang bersifat lokal untuk setiap amatan (Kusnandar dkk., 2021). Koefisien regresi pada model GWR berbeda-beda bergantung pada titik lokasi amatan. Model GWR pada koordinat lokasi bujur ($u$) dan lintang ($v$) dapat dituliskan sebagai berikut (Sihombing dkk., 2023):

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^{l} \beta_p(u_i, v_i)x_{pi} + \varepsilon_i \tag{16}$$

Keterangan:
$(u_i, v_i)$ : koordinat letak geografis (bujur, lintang) lokasi pengamatan ke$-i$.

Metode penaksiran parameter pada model GWR yaitu metode *Weighted Least Square* (WLS) yang memberikan pembobot untuk setiap amatan. Parameter lokasi $(u_i\,, v_i)$ diestimasi dengan menambahkan unsur pembobot $W(u_i\,, v_i)$ kemudian meminimumkan jumlah kuadrat *error* dengan mengkombinasikan pada Persamaan (1) sebagai berikut (Daulay & Simamora, 2023):

$$WLS = \sum_{i=1}^{n} W(u_i, v_i)\varepsilon_i^2 \tag{17}$$

Persamaan (17) dalam bentuk matriks adalah sebagai berikut:

$$\begin{aligned} WLS &= \boldsymbol{\varepsilon}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{\varepsilon} \\ &= (\boldsymbol{Y} - \boldsymbol{X\beta})^T \boldsymbol{W}(u_i, v_i)(\boldsymbol{Y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{Y}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{Y} - 2\boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{Y}\boldsymbol{W}(u_i, v_i) + \boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{X\beta} \end{aligned} \tag{18}$$

Koefisien regresi $\beta$ diperoleh melalui differensiasi dari fungsi WLS pada Persamaan (18) terhadap $\boldsymbol{\beta}^T$ sebagai berikut:

$$\frac{\partial WLS}{\partial \boldsymbol{\beta}^T} = 0$$

$$\frac{\partial(\boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X\beta}\,)}{\partial \boldsymbol{\beta}^T} = 0$$

$$0 - 2\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{W}(u_i, v_i)\boldsymbol{\beta} = 0$$

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{Y} \tag{19}$$

Karena terdapat $n$ lokasi amatan maka penaksir yang diperoleh merupakan penaksir setiap baris dari matriks lokal parameter seluruh lokasi amatan sebagai berikut:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0(u_1, v_1) & \beta_1(u_1, v_1) & \beta_2(u_1, v_1) & \cdots & \beta_l(u_1, v_1) \\ \beta_0(u_2, v_2) & \beta_1(u_2, v_2) & \beta_2(u_2, v_2) & \cdots & \beta_l(u_2, v_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \beta_1(u_n, v_n) & \beta_2(u_n, v_n) & \cdots & \beta_l(u_n, v_n) \end{bmatrix} \tag{20}$$

### 1.4.6 Pemilihan Bandwidth Optimal

*Bandwidth* adalah lingkaran dengan radius ($h$) dari titik pusat lokasi yang digunakan sebagai dasar untuk menentukan bobot setiap pengamatan terhadap model regresi pada lokasi tersebut. Titik-titik pengamatan yang berada dalam radius lingkaran dianggap memiliki pengaruh terhadap pembentukan model di lokasi pengamatan ke$-i$. Pemilihan *bandwidth* yang optimal bertujuan untuk menghindari ragam yang tidak homogen akibat peningkatan nilai pendugaan koefisien parameter (Farichah, 2020). Salah satu metode yang sering digunakan untuk memilih *bandwidth* optimal adalah metode *Cross Validation* (CV). Penerapan metode *Cross Validation* dalam pemilihan *bandwidth* optimal dilakukan dengan memilih nilai CV terendah. Pemilihan *bandwidth* optimal menggunakan metode CV dapat dituliskan menggunakan persamaan berikut (Razak dkk., 2019):

$$CV(h) = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(h)\right)^2 \tag{21}$$

Keterangan:

$\hat{y}_{\neq i}(h)$ : Nilai estimasi untuk lokasi pengamatan ke$-i$ yang dihitung tanpa melibatkan data dari lokasi pengamatan ke$-i$.

### 1.4.7 Fungsi Pembobot Spasial

Matriks pembobot spasial merupakan matriks diagonal yang berfungsi untuk mengidentifikasi hubungan spasial antar lokasi amatan. Pembobot yang digunakan bergantung pada jarak antar titik amatan. Setiap elemen pada matriks diagonal merupakan fungsi pembobot dari masing-masing titik amatan. Dengan demikian, matriks pembobot ini berperan penting dalam menaksir parameter yang dapat bervariasi di setiap lokasi pengamatan (Lestari, 2023). Matriks pembobot spasial diperoleh dari perhitungan jarak *Euclidean* antara lokasi berdasarkan derajat bujur (*longitude*) dan lintang (*latitude*). Perhitungan jarak *Euclidean* menggunakan persamaan berikut (Yulianti dkk., 2022):

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{22}$$

Keterangan:

$d_{ij}$ : jarak *Euclidean* pada baris ke$-i$ dan kolom ke$-j$
$u_i$ : bujur pada baris ke$-i$
$u_j$ : bujur pada kolom ke$-j$
$v_i$ : lintang pada baris ke$-i$
$v_j$ : lintang pada baris ke$-j$

Setelah perhitungan jarak *Euclidean*, selanjutnya matriks pembobot untuk setiap lokasi amatan dapat diperoleh melalui fungsi pembobot *kernel*. Fungsi pembobot *kernel*. Fungsi *kernel* berperan sebagai fungsi pembobot yang menentukan seberapa besar pengaruh observasi di sekitar lokasi tertentu terhadap estimasi parameter regresi di lokasi tersebut. Data yang memiliki jarak lebih dekat ke titik regresi akan mendapatkan bobot yang lebih tinggi dibandingkan dengan data yang berjarak lebih jauh (Farichah, 2020). Terdapat dua jenis fungsi *kernel* yaitu fungsi *fixed kernel* dan *adaptive kernel*. Setiap fungsi *kernel* memiliki tiga jenis pembobot, yaitu pembobot *Gaussian*, *Bisquare*, dan *Tricube*. Perbedaan utama antara fungsi *adaptive kernel* dan *fixed kernel* terletak pada penggunaan *bandwidth*. *Adaptive kernel* memiliki *bandwidth* yang berbeda-beda sesuai dengan setiap titik lokasi amatan, sedangkan *fixed kernel* menerapkan *bandwidth* yang sama di semua lokasi pengamatan. Hal ini disebabkan oleh kemampuan fungsi *adaptive kernel* untuk menyesuaikan diri dengan kondisi distribusi titik-titik pengamatan. Persamaan fungsi pembobot *kernel* sebagai berikut (Lestari, 2023):

a. *Adaptive gaussian kernel*

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right] ; i, j = 1,2, \dots, n ; i \neq j \tag{23}$$

b. *Adaptive bisquare kernel*

$$w_{ij} = \begin{cases} \left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right], & d_{ij} < h \\ 0 \quad , & d_{ij} > h \end{cases} \tag{24}$$

c. *Adaptive tricube kernel*

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h_i}\right)^3\right]^3, & d_{ij} < h \\ 0 \quad , & d_{ij} > h \end{cases} \tag{25}$$

d. *Fixed gaussian kernel*

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right] ; i, j = 1,2, \dots, n ; i \neq j \tag{26}$$

e. *Fixed bisquare kernel*

$$w_{ij} = \begin{cases} \left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right], & d_{ij} < h \\ 0 \quad , & d_{ij} > h \end{cases} \tag{27}$$

f. *Fixed tricube kernel*

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h}\right)^3\right]^3, & d_{ij} < h \\ 0 \quad , & d_{ij} > h \end{cases} \tag{28}$$

Keterangan:

$w_{ij}$ : matriks pembobot spasial
$h$ : jarak *bandwidth* pada lokasi pengamatan ke$-i$

### 1.4.8 Regresi Nonparametrik *Spline* pada Model *Geographically Weighted Regression*

Regresi nonparametrik *spline truncated* dalam model *Geographically Weighted Regression* (GWR) adalah pengembangan dari regresi nonparametrik yang mempertimbangkan faktor geografis atau spasial. Metode ini memadukan analisis nonparametrik dengan elemen geografis, memungkinkan variasi data dipengaruhi oleh lokasi geografisnya. Secara matematis, hubungan antara variabel respon dan variabel prediktor pada lokasi ke-$i$ adalah sebagai berikut (Serena dkk., 2021):

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^{l}\sum_{m=1}^{q} \beta_m(u_i, v_i)x_{pi}^m + \sum_{p=1}^{l}\sum_{k=1}^{r} \beta_{p,q+k}(u_i, v_i)\left(x_{pi} - K_k\right)_+^q + \varepsilon_i \quad (29)$$

Keterangan:

| | |
|---|---|
| $\beta_{p,q+k}(u_i, v_i)$ | : parameter regresi fungsi *truncated* ke$-(q+k)$ pada titik knot ke-$k$ dan variabel prediktor ke$-p$ pada lokasi pengamatan ke$-i$ |
| $l$ | : banyaknya variabel prediktor |
| $m$ | : orde dengan $m = 1,2, \dots, q$ |
| $k$ | : jumlah knot dengan $k = 1,2, \dots, r$ |
| $K$ | : letak titik knot ke-$K$ |
| $(x_i - K_k)_+^q$ | : fungsi *truncated* pada variabel ke-$i$ dengan letak titik knot $K_k$ dan orde $q$. |

### 1.4.9 *Penalized Weighted Least Square*

*Penalized Weighted Least Square* (PWLS) merupakan pengembangan dari *Penalized Least Square* yang melibatkan pembobotan, titik knot, dan parameter penghalus secara bersamaan. PWLS efektif pada data yang memiliki karakteristik tertentu seperti heteroskedastisitas atau *outlier.* Kriteria dalam PWLS telah dikembangkan oleh peneliti dalam model regresi nonparametrik. PWLS telah digunakan dalam model regresi nonparametrik, dengan berbagai jenis pembobotan seperti matriks ragam peragam, jumlah parameter, atau bootstrap. Pada PWLS, setiap amatan diberikan bobot yang berbeda dan juga menambahkan fungsi penalti ke dalam fungsi regresi yang bertujuan untuk mengontrol kompleksitas model dan mencegah *overfitting.* Dalam proses estimasi parameter, dilakukan modifikasi terhadap fungsi regresi nonparametrik dengan menambahkan pembobot dan fungsi penalti. (Islamiyati dkk., 2022). Kriteria estimasi untuk model regresi nonparametrik dengan PWLS adalah sebagai berikut:

$$PWLS = \sum_{i=1}^{n} W(u_i, v_i)\varepsilon_i^2 + \lambda \sum_{k}^{r} \beta_{q+k}^2(u_i, v_i) \quad (30)$$

Keterangan:

| | |
|---|---|
| $\beta_{q+k}(u_i, v_i)$ | : parameter regresi fungsi *truncated* ke-$(q+k)$ pada titik knot ke-$k$ dan variabel prediktor ke$-p$ pada lokasi pengamatan ke$-i$ |
| $m$ | : orde dengan $m = 1,2, \dots, q$ |

$k$ : jumlah knot dengan $k = 1,2, \dots, r$
$K$ : letak titik knot ke-$k$
$(x_i - K_k)_+^q$ : fungsi *truncated* pada variabel ke-$i$ dengan letak titik knot $K_k$ dan orde $q$.
$\lambda$ : parameter penghalus

Persamaan (30) untuk PWLS dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\begin{aligned} PWLS &= \boldsymbol{\varepsilon}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{\varepsilon} + \lambda \boldsymbol{\beta}^T(u_i, v_i)\boldsymbol{D}\boldsymbol{\beta}(u_i, v_i) \\ &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(u_i, v_i))^T \boldsymbol{W}(u_i, v_i)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(u_i, v_i)) + \lambda \boldsymbol{\beta}^T(u_i, v_i)\boldsymbol{D}\boldsymbol{\beta}(u_i, v_i) \end{aligned} \quad (31)$$

### 1.4.10 Pengujian Signifikansi Parameter

1. Uji Simultan

Pengujian parameter dilakukan secara simultan bertujuan untuk mengidentifikasi pengaruh signifikan variabel prediktor secara keseluruhan terhadap variabel respon. Prosedur pengujian signifikansi parameter dalam model secara simultan menggunakan uji F adalah sebagai berikut (Mukrom dkk., 2021):

Hipotesis:

$H_0$ : $\beta_{1,0}(u_i, v_i) = \beta_{1,1}(u_i, v_i) = \cdots = \beta_{l,q}(u_i, v_i) = \cdots = \beta_{l,q+r}(u_i, v_i) = 0$
(Tidak terdapat pengaruh variabel prediktor secara bersamaan terhadap variabel respon)

$H_1$ : Minimal terdapat satu $\beta_{l,q}(u_i, v_i) \neq 0$ atau $\beta_{l,q+k}(u_i, v_i) \neq 0$
(Terdapat pengaruh variabel prediktor secara bersamaan terhadap variabel respon)

Statistik Uji:

$$F_{hitung} = \frac{SSR/l}{SSE/(n-l-1)} \quad (32)$$

Keterangan:

SSR : jumlah kuadrat regresi
SSE : jumlah kuadrat *error*
$l$ : banyaknya variabel prediktor

Kriteria Uji:

Tolak $H_0$ jika nilai $F_{hitung} > F_{(\alpha;db1;db2)}$ atau $p - value < \alpha$.

2. Uji Parsial

Pengujian parameter secara parsial untuk mengetahui variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon secara individu. Adapun prosedur pengujian signifikansi parameter model GWNR dengan *spline penalized* secara parsial menggunakan uji *Wald* adalah sebagai berikut (Mursidin, 2024):

Hipotesis

$H_0$ : $\beta_j(u_i, v_i) = 0$
(Tidak terdapat pengaruh variabel prediktor secara parsial terhadap variabel respon)

$H_1$ : $\beta_j(u_i, v_i) \neq 0$
(Terdapat pengaruh variabel prediktor secara parsial terhadap variabel respon)

Statistik Uji

$$Wald = \frac{\hat{\beta}_j^2(u_i, v_i)}{var(\hat{\beta}_j(u_i, v_i))} \quad (33)$$

Keterangan:
$\hat{\beta}_j(u_i, v_i)$ : nilai estimator parameter ke-$j$
$var\left(\hat{\beta}(u_i, v_i)\right)$ : matriks ragam yang berkorespondensi terhadap $\beta_j$
Kriteria Uji:
Tolak $H_0$ jika nilai $W > \chi^2_{\alpha,1}$ atau $p - value < \alpha$.

### 1.4.11 Persentase Penduduk Miskin

Kemiskinan adalah salah satu masalah utama yang selalu menjadi fokus perhatian pemerintah di setiap negara. Menurut data dari Badan Pusat Statistik (BPS), tingkat kemiskinan di Indonesia menunjukkan fluktuasi yang dipengaruhi oleh berbagai faktor ekonomi, sosial, dan politik. Pada tahun 2023, jumlah penduduk miskin di Indonesia tercatat mencapai sekitar 26,36 juta jiwa, dengan persentase kemiskinan nasional berada di angka 9,57%. Persentase kemiskinan menunjukkan adanya penurunan dalam beberapa tahun terakhir. Tingkat kemiskinan di Provinsi Sulawesi Selatan juga mencerminkan tren yang serupa. Berdasarkan data terbaru dari BPS Provinsi Sulawesi Selatan, jumlah penduduk miskin pada Desember 2023 meningkat sebesar 0,83% menjadi 788,85 ribu jiwa dibandingkan dengan data sebelumnya. Persentase penduduk miskin di Sulawesi Selatan bervariasi antar kabupaten dan kota, dengan beberapa daerah mengalami tingkat kemiskinan yang lebih tinggi dibandingkan yang lain. Misalnya, Kabupaten Toraja Utara dan Makassar menunjukkan persentase kemiskinan yang signifikan, sementara daerah seperti Bulukumba mencatat penurunan dalam tingkat kemiskinan (Risantika dkk., 2023).

Secara umum, masalah kemiskinan membutuhkan perhatian serius dari berbagai pihak. Kerjasama antara pemerintah, masyarakat sipil, dan sektor swasta sangat penting untuk menciptakan solusi yang efektif dan berkelanjutan. Dengan pendekatan yang terintegrasi dan berfokus pada pemberdayaan masyarakat yang mencakup pendidikan, pelatihan keterampilan, dan akses terhadap peluang ekonomi, diharapkan tingkat kemiskinan terus menurun dan kesejahteraan masyarakat meningkat secara signifikan.