

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Salah satu metode statistika yang paling sering digunakan adalah analisis regresi. Analisis regresi mampu memodelkan hubungan variabel respon dengan prediktornya. Namun, yang sering kali menjadi kesalahan adalah penggunaan metode tersebut yang tidak sesuai dengan kondisi data yang disajikan. Apabila data yang disajikan dipengaruhi oleh faktor geografis atau data spasial, maka tentu penggunaan analisis regresi kurang tepat. Hal tersebut disebabkan karena analisis regresi hanya mampu memodelkan secara global, sedangkan pada data spasial diperlukan model secara lokal karena setiap lokasi pengamatan dipengaruhi oleh kondisi geografis yang berbeda-beda dan saling berpengaruh.

Pengembangan analisis regresi dengan menggunakan data spasial adalah *Geographically Weighted Regression* (GWR). Menurut Caraka & Yasin (2017), GWR adalah metode statistika yang digunakan untuk menganalisis heterogenitas spasial. Heterogenitas spasial terjadi apabila suatu variabel yang sama memberikan respon yang tidak sama pada lokasi yang berbeda dalam satu wilayah pengamatan. Model GWR menggunakan unsur matriks pembobot spasial yang besarnya tergantung pada kedekatan lokasi. Semakin dekat suatu lokasi, bobot pengaruhnya akan semakin besar (Haryanto & Andriani, 2021).

Berbicara tentang analisis regresi khususnya model GWR, seringkali dijumpai kondisi dimana suatu variabel prediktor yang saling berkorelasi. Kondisi tersebut disebut sebagai multikolinearitas. Multikolinearitas adalah kondisi dimana prediktor yang saling berkorelasi yang menyebabkan estimasi parameter dari model regresi yang dihasilkan memiliki varian yang sangat besar yang berpotensi menyebabkan kesalahan dalam menginterpretasi parameter (Yulita dkk, 2015). Metode alternatif yang dapat mengatasi permasalahan multikolinearitas dalam data adalah *principal component analysis* (PCA). Menurut Zhu, dkk (2021), "*PCA was used to eliminate the multi-collinearity of independent variables, so the global model of PCA-OLS was established*". Yang berarti bahwa PCA mampu mengeliminasi multikolinearitas pada variabel prediktor sehingga terbentuk komponen baru dari hasil PCA. *Geographically Weighted Regression Principal Components Analysis* (GWRPCA) merupakan model yang menggabungkan antara PCA dan GWR apabila pada model GWR terdapat masalah multikolinearitas. Menurut (Charlton dkk, 2010), "*PCA has been shown to be useful in dealing with colinearity in the data for global regression models; similarly GWRPCA should be useful in dealing with colinearity in GWR*".

Model GWR mampu mengatasi masalah analisis regresi spasial dengan baik jika data atau kurvanya diketahui dan memiliki pola linear. Namun kenyataannya, tidak semua kasus yang dihadapi dalam analisisnya memiliki pola linear. Seringkali kita menjumpai data yang dianalisis tidak memiliki pola parametrik, seperti pola

linear, kuadratik maupun kubik. Sehingga diperlukan pendekatan nonparametrik pada model GWR. Salah satu pendekatan nonparametrik adalah spline (Ente, dkk, 2021). Pada pendekatan spline, model cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak karena terdapat titik-titik knot, yaitu titik-perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Budiantara, 2009). Pendekatan spline terdiri dari beberapa estimator, yaitu spline *smoothing* (Islamiyati dkk, 2020), *penalized* spline (Islamiyati dkk, 2020) dan *truncated* spline (Islamiyati dkk, 2022). Pada penelitian ini digunakan spline *truncated*. Estimator ini efektif untuk data yang tidak mengikuti pola parametrik, sehingga cocok untuk data dengan perubahan mendadak atau pola yang kompleks (Nurhuda dkk, 2022). Spline *truncated* mampu menangkap perubahan pola data pada subinterval tertentu dengan menggunakan fungsi basis yang terpotong pada titik-titik knot, sehingga menghasilkan model yang kompleks dan adaptif.

Pengembangan model GWR dengan pendekatan nonparametrik spline telah dilakukan oleh beberapa peneliti. Pada penelitian yang dilakukan oleh Ilmi, dkk (2021), menggunakan *geographically weighted spline nonparametric regression* untuk mengidentifikasi faktor – faktor yang mempengaruhi tingkat pengangguran terbuka di Kalimantan. Kemudian penelitian yang dilakukan oleh Sifriyani (2019) mengkaji tentang uji statistik dan distribusi untuk uji hipotesis simultan dengan menggunakan *multivariable nonparametric Geographically Weighted Regression* dengan pendekatan spline *truncated*. Namun, beberapa penelitian terdahulu yang serupa, belum mempertimbangkan adanya kasus multikolinearitas yang mungkin terjadi pada data. Sehingga diperlukan pengembangan penelitian dengan melakukan pendekatan nonparametrik pada model GWRPCA.

Diabetes Melitus (DM) merupakan penyakit tidak menular yang ditandai dengan tingginya kadar gula dalam tubuh yang disebabkan oleh kurang responsifnya sel sel tubuh terhadap insulin. Selain faktor genetik, usia atau kondisi medis tertentu, gaya hidup yang buruk juga merupakan faktor yang sangat penting sebagai faktor resiko diabetes. Menurut penelitian yang dilakukan oleh Widiyari, dkk (2021), dijelaskan bahwa obesitas, hipertensi, aktivitas fisik yang kurang, mengonsumsi makanan yang tidak sehat, mengandung tinggi glukosa dan rendah serat dapat memberikan peluang tinggi untuk menderita intoleransi insulin. Anri (2022), juga menyimpulkan bahwa, faktor yang berhubungan dengan kejadian diabetes adalah indeks massa tubuh, pola makan, dan aktivitas fisik. Diabetes sejauh ini masih tercatat sebagai kasus penyebab kematian yang tinggi dengan jumlah kasus yang selalu meningkat. Sejauh ini, sebanyak 6,7 juta kasus kematian akibat diabetes tercatat di *International Diabetes Federation* (2021) dengan jumlah kasus diabetes sebanyak 536 juta dan diperkirakan akan meningkat sebanyak 643 juta kasus pada tahun 2030 dan 784 juta kasus pada tahun 2045. Di Indonesia, pada tahun 2021 tercatat sebanyak 236 ribu kasus kematian akibat diabetes dan menempati peringkat keenam dengan jumlah kematian terbanyak. Pada tahun 2022, di provinsi Sulawesi Selatan sebanyak 54.007 kasus.

Pemetaan distribusi data kasus diabetes mellitus di Sulawesi Selatan merupakan permasalahan yang memperhatikan faktor geografis, sehingga dapat dianalisis menggunakan metode GWR. Hasil eksplorasi data kasus diabetes mellitus dengan faktor – faktor yang mempengaruhinya, menunjukkan bahwa pola persebaran data tidak membentuk pola parametrik dan berubah – ubah pada interval tertentu sehingga pendekatan spline dengan fungsi *truncated* dapat digunakan. Faktor – faktor yang dianggap mempengaruhi kasus diabetes mellitus di Sulawesi Selatan adalah indeks massa tubuh (normal, berat badan berlebih dan obesitas), pola makan (dalam hal ini konsumsi gula, garam atau lemak berlebih), serta aktivitas fisik. Hal yang menarik pada data tersebut adalah beberapa diantaranya memiliki multikolinearitas yang tinggi sehingga untuk mengatasi masalah tersebut digunakan metode GWRPCA. Dengan demikian, pendekatan spline *truncated* dalam model GWRPCA dapat digunakan pada data kasus diabetes mellitus di Sulawesi Selatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi model GWRPCA menggunakan estimator spline *truncated*?
2. Bagaimana model GWRPCA dengan spline *truncated* pada kasus diabetes mellitus di Sulawesi Selatan?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut:

1. Mengestimasi model GWRPCA menggunakan estimator spline *truncated*.
2. Mendapatkan model GWRPCA dengan spline *truncated* pada kasus diabetes mellitus di Sulawesi Selatan.

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat untuk berbagai pihak, diantaranya adalah sebagai berikut:

1. Bagi Peneliti
Menambah ilmu dan pengetahuan tentang penerapan spline *Truncated* pada model GWRPCA.
2. Bagi Pembaca
Memperoleh informasi terkait hasil analisis kasus diabetes mellitus di Sulawesi Selatan agar dapat dijadikan salah satu acuan untuk mengurangi kasus tersebut. Penelitian ini juga diharapkan mampu menjadi referensi penelitian dalam meneliti suatu masalah yang serupa.

1.5 Batasan Masalah

Berdasarkan pada rumusan masalah, penelitian ini dibatasi pada beberapa hal yaitu sebagai berikut:

1. Penafsiran parameter menggunakan metode *Weighted Least Square* (WLS).

2. Estimator spline *truncated* yang digunakan dibatasi hanya menggunakan tiga titik knot pada setiap variabel prediktor.
3. Pemilihan titik knot optimal menggunakan *Generalized Cross Validation* (GCV).
4. Pemilihan *bandwidth* optimal menggunakan metode *Cross Validation* (CV) dengan matriks pembobot *fixed kernel* dan *adaptive kernel*.

1.6 Kajian Teori

Pada bagian ini diuraikan teori pendukung yang akan digunakan pada hasil hasil pembahasan yang diambil dari Caraka & Yasin (2017), Johson & Wichern (2007), Fotheringham, dkk (2002) dan beberapa jurnal yang terkait.

1.6.1 Analisis Regresi

Analisis regresi digunakan digunakan untuk menjelaskan hubungan antara variabel respon y dengan satu atau lebih variabel prediktor x . Model regresi dapat dituliskan sebagai berikut:

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

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

Dimana

- y_i : Variabel respon ke- i
 x_{ij} : Variabel prediktor ke- i
 β_0 : Intersep model
 β_j : Koefisien regresi
 ε_i : Error pada pengamatan ke- i

Model regresi dapat ditulis dalam bentuk matriks jika Persamaan (1) diatas diuraikan seperti berikut:

$$\begin{aligned}
 y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_j x_{1k} + e_1 \\
 y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_j x_{2k} + e_2 \\
 &\vdots \\
 y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_j x_{nk} + e_n
 \end{aligned}$$

Dari sistem Persamaan di atas, dapat diperoleh bentuk matriks seperti berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Dari bentuk matriks di atas, dapat ditulis menjadi

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

Dimana

- Y** : Vektor ($n \times 1$) dari variabel respon
X : Vektor $n \times (k + 1)$ dari variabel prediktor
 β : Vektor $(k + 1) \times 1$ dari koefisien regresi
 ε : Vektor ($n \times 1$) dari error pengamatan

Dalam analisis regresi, metode yang digunakan untuk menaksir parameter β adalah metode kuadrat terkecil atau sering dikenal sebagai *Ordinary Least Square* (OLS). OLS bekerja dengan cara meminimumkan jumlah kuadrat error dari Persamaan (2) yaitu,

$$\begin{aligned}
 \varepsilon^T \varepsilon &= (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) \\
 &= \mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{X}\beta + \beta^T \mathbf{X}^T \mathbf{X}\beta
 \end{aligned}$$

Maka, diperoleh estimasi parameter β sebagai berikut.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (3)$$

1.6.2 Regresi Nonparametrik

Regresi nonparametrik adalah metode yang digunakan jika pola hubungan antara variabel respon dengan prediktornya tidak diketahui bentuk fungsi regresinya. Sehingga model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linear ataupun nonlinear. Kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi. Secara umum model regresi nonparametrik dituliskan sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i \quad (4)$$

Dimana

- y_i : Variabel respon pada pengamatan ke- i
 $f(x_i)$: Fungsi regresi yang tidak diketahui bentuk atau polanya
 ε_i : Error pada pengamatan ke- i

1.6.3 Pendekatan Regresi Nonparametrik Spline *Truncated*

Pada regresi nonparametrik bentuk kurva regresinya tidak diketahui, sehingga perlu dilakukan pendekatan spline. Spline adalah potongan – potongan polynomial yang memiliki sifat tersegment. Sifat tersegment ini memberikan fleksibilitas lebih dari polynomial biasa, sehingga dapat menyesuaikan diri secara lebih efektif terhadap karakteristik dari suatu fungsi atau data. Menurut Islamiyati (2017), spline mampu menangani data yang memiliki perilaku yang berubah – ubah pada sub – sub interval tertentu dan cenderung mencari sendiri estimasi data kemanapun pola data bergerak. Salah satu jenis spline yang dapat digunakan dalam regresi nonparametrik adalah spline *truncated*. Pada spline *truncated* digunakan titik knot. Knot adalah nilai variabel bebas yang menjadi batas antar segmen. Knot diartikan sebagai suatu titik fokus dalam fungsi spline, sehingga kurva yang dibentuk tersegment pada titik tersebut (yozza & Afrimayani, 2019). Fungsi spline dengan orde m dan titik – titik knot $K_1, K_2, \dots, K_j$ yang didefinisikan sebagai sebarang fungsi f disajikan sebagai berikut:

$$f(x_i) = \sum_{k=0}^m \beta_k x_i^k + \sum_{p=1}^q \beta_{q+l} (x_i - K_p)_+^m \quad (5)$$

Dimana β_k dan β_{q+l} adalah parameter regresi dan K_p adalah titik knot. Sedangkan fungsi $(x_i - K_p)_+^m$ adalah polynomial *truncated* yang dijabarkan seperti berikut:

$$(x_i - K_p)_+^m = \begin{cases} (x_i - K_p)^m, & \text{jika } x_i \geq K_p \\ 0, & \text{jika } x_i < K_p \end{cases} \quad (6)$$

Jika $m = 1$ maka disebut spline *truncated* linear, $m = 2$ disebut spline *truncated* kuadratik, dan $m = 3$ disebut spline *truncated* kubik (Rodriguez, 2001).

1.6.4 Pemilihan Titik Knot Optimum

Penentuan model spline terbaik pada dasarnya adalah menentukan titik knot optimum dan pemilihan lokasi titik – titik knot tersebut. Penentuan spline terbaik juga mencakup pemilihan orde dari fungsi polynomial yang digunakan pada setiap segmen (Yozza & Afrimayani, 2019). Salah satu metode yang dapat digunakan adalah *Generalized Cross Validation* (GCV). GCV menjadi metode yang sangat sering digunakan dalam memilih titik knot yang optimal. Metode tersebut merupakan bentuk modifikasi dari metode *Cross Validation* (CV) (Nurhuda dkk, 2022). Titik knot optimal dipilih berdasarkan nilai GCV yang paling minimum (Nurdiani dkk, 2017). Semakin kecil nilai GCV, maka error dari model spline akan kecil. Semakin kecil error suatu model maka akan lebih representatif. Sehingga titik knot yang optimal akan memberikan model yang terbaik. Adapun Persamaan GCV dapat dituliskan seperti berikut:

$$GCV(K_1, K_2, \dots, K_j) = \frac{MSE(K_1, K_2, \dots, K_j)}{\left(\frac{1}{n} \text{trace}[\mathbf{I} - \mathbf{A}(K_1, K_2, \dots, K_j)]\right)^2} \quad (7)$$

Dengan $MSE(K_1, K_2, \dots, K_j) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$. $\mathbf{A}(K_1, K_2, \dots, K_j)$ diperoleh dari hubungan $\hat{\mathbf{y}} = \mathbf{A}(K_1, K_2, \dots, K_j)\mathbf{y}$ sehingga nilai $\mathbf{A}(K_1, K_2, \dots, K_j) = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ dan $\mathbf{I}$ matriks identitas.

1.6.5 Heterogenitas Spasial

Data spasial adalah data yang mengandung informasi lokasi atau letak geografis suatu daerah dan diperoleh dari hasil pengukuran. Data spasial memiliki karakteristik yang berbeda pada setiap lokasi pengamatan tetapi saling berhubungan antara data di lokasi pengamatan yang berdekatan yang dinamakan efek spasial. Efek spasial yang terjadi antara wilayah pengamatan dikategorikan menjadi efek spasial tipe wilayah (autokorelasi spasial), dan efek spasial titik (heterogenitas spasial). Heterogenitas spasial adalah apabila suatu peubah bebas yang sama memberikan

respon yang tidak sama pada lokasi yang berbeda dalam satu wilayah penelitian. Heterogenitas spasial terjadi karena variansi dari residual suatu pengamatan dengan pengamatannya berbeda.

Heterogenitas spasial dapat diuji dengan menggunakan statistik *Breusch – Pagan Test*. Bentuk hipotesisnya adalah sebagai berikut:

$$H_0 : \sigma_{(u_1, v_1)}^2 = \sigma_{(u_2, v_2)}^2 = \dots = \sigma_{(u_n, v_n)}^2 = \sigma^2 \text{ dimana } \sigma_{(u_i, v_i)}^2 = \sigma^2$$

(Tidak ada pengaruh heterogenitas spasial)

$$H_0 : \text{Paling tidak ada satu } \sigma_{(u_i, v_i)}^2 \neq \sigma^2 \text{ (terdapat heterogenitas spasial)}$$

Dengan statistik uji:

$$BP = \left(\frac{1}{2}\right) \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} + \left(\frac{1}{T}\right) \left[\frac{\mathbf{e}^T \mathbf{W} \mathbf{e}}{\sigma^2} \right]^2 \sim X_{(k+1)}^2 \quad (8)$$

$$\text{Dimana } \mathbf{f} = (f_1, f_2, \dots, f_n), f_i = \frac{e_i}{\sigma^2} - 1, \quad T = \text{Trace}(\mathbf{W}^T \mathbf{W} + \mathbf{W}^2), \quad \mathbf{W}$$

merupakan matriks pembobot antara lokasi i dan j , dan $\mathbf{Z}$ adalah matriks berukuran $n \times (k + 1)$ yang berisi vektor yang sudah dinormal-standarkan (Z) untuk setiap obsrvasi dengan k merupakan banyaknya peubah prediktor. Kriteria pengambilan keputusan tolak H_0 jika nilai $BP > X_k^2$ atau nilai $p - \text{value} < \alpha$.

1.6.6 Multikolinearitas

Multikolinearitas adalah kondisi dimana variabel prediktor yang saling berkorelasi yang menyebabkan estimasi parameter dari model regresi global memiliki error yang sangat besar (Yulita dkk, 2015). Menurut O'Brien (2007), multikolinearitas dapat dideteksi dengan mencari nilai VIF (*Variance Inflation Faktor*) dengan matriks pembobot sebagai lokasi pendeteksi kolinearitas pada model GWR. Variabel prediktor dikatakan saling berkorelasi jika nilai VIF nya lebih dari 10. Pada model GWR, nilai VIF memungkinkan untuk dihitung untuk setiap variabel prediktornya. Nilai VIF dapat dituliskan sebagai berikut:

$$VIF_k(u_i, v_i) = \frac{1}{1 - R_k^2} \quad (9)$$

Dimana R_k^2 sebagai koefisien determinasi antara X_k dengan variabel prediktor.

1.6.7 Principal Component Analysis

Principal Component Analysis (PCA) merupakan metode statistika yang menjelaskan struktur varian kovarian dari sekelompok variabel melalui kombinasi linear yang tidak lebih banyak dari variabel tersebut (Johnson & Wichern, 2007). Tujuan Penggunaan PCA adalah untuk mereduksi dimensi data awal yang dilakukan dengan cara mengeliminasi korelasi antara variabel prediktor melalui transformasi variabel prediktor awal ke variabel prediktor baru yang sudah tidak berkorelasi lagi (Jolliffe, 2002).

Menurut Johnson & Wichern (2007), komponen utama yang dimodelkan sebagai bentuk kombinasi linear dari variabel yang dibakukan yaitu $\mathbf{Z} = [Z_{ij1}, Z_{ij}, \dots, Z_{ijp}]$ dengan pasangan nilai eigen dan vektor eigen yakni

$(\lambda_1, \alpha_1), (\lambda_2, \alpha_2), \dots, (\lambda_p, \alpha_p)$ dimana $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p$ dituliskan seperti Persamaan berikut:

$$\begin{aligned} W_{ij1} &= \alpha'_1 Z = \alpha_{11}Z_{ij1} + \dots + \alpha_{1p}Z_{ijp} \\ W_{ij} &= \alpha'_2 Z = \alpha_{21}Z_{ij1} + \dots + \alpha_{2p}Z_{ijp} \\ &\vdots \\ W_{ijp} &= \alpha'_p Z = \alpha_{p1}Z_{ij1} + \dots + \alpha_{pp}Z_{ijp} \end{aligned} \quad (10)$$

Terdapat tiga metode dalam menentukan jumlah komponen utama, yaitu dengan melihat persentase varian kumulatif yang terpilih yakni mampu menjelaskan total varian data sekitar 70% sampai 80%. Selanjutnya, yaitu dengan melihat pola penurunan nilai eigen atau sudut pada *scree plot*. *Scree plot* itu sendiri adalah plot antara nilai eigen λ_k dengan komponen utama ke- k . Jika terjadi penurunan yang tajam pada *scree plot*, maka mengindikasikan perubahan nilai eigen yang besar. Selain itu, jumlah komponen utama juga dapat dipilih berdasarkan nilai eigen yang lebih dari satu ($\lambda_k > 1$). Nilai eigen yang hampir nol dianggap tidak memberikan pengaruh signifikan.

1.6.8 Geographically Weighted Regression

Geographically Weighted Regression (GWR) merupakan pengembangan dari model regresi. Dalam GWR, setiap parameternya dicari berdasarkan lokasi pengamatan, sehingga setiap lokasi pengamatan memiliki parameter tersendiri yang berbeda dengan lokasi lainnya yang sifatnya lokal (Fotheringham dkk, 2002). Variabel respon y pada model GWR diregresikan dengan variabel prediktornya dimana koefisien regresinya bergantung pada lokasi data pengamatan. Secara umum, model GWR dapat dituliskan seperti berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (11)$$

Menurut Fotheringham dkk (2002), estimasi parameter dalam model GWR dilakukan dengan menggunakan metode *Weighted Least Square* (WLS) yaitu menggunakan pembobot yang berbeda untuk setiap lokasi data yang diamati. Menurut Chen, dkk (2016), pada model GWR terdapat tiga metode dalam mengestimasi parameter bandwidth dari fungsi pembobot: *nearest neighbor bandwidth parameter method*, *Cross Validation* (CV), dan *Akaike Information Criterion* (AIC). Diasumsikan bahwa daerah yang dekat dengan lokasi ke- i memiliki pengaruh yang besar terhadap estimasi parameternya dibanding dengan daerah yang lebih jauh. Berikut estimator parameter dalam model GWR untuk setiap lokasi nya ((Fotheringham dkk, 2002).

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (12)$$

Pembobot sangat berperan penting dalam model GWR karena mampu mewakili letak data pengamatan antara satu dengan yang lainnya. Matriks pembobot yang berdimensi $n \times n$ adalah sebagai berikut:

$$W(u_i, v_i) = \begin{bmatrix} W_{i1} & 0 & \cdots & 0 \\ 0 & W_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{in} \end{bmatrix}$$

Dimana W_{in} adalah bobot untuk data pada titik ke- n pada pengujian model di sekitar titik i .

1.6.9 Geographically Weighted Regression Principal Component Analysis

Geographically Weighted Regression Principal Component Analysis (GWRPCA) merupakan penggabungan antara dua metode yaitu metode PCA dengan GWR. GWRPCA merupakan pengembangan dari analisis regresi yang menggunakan dasar dari PCA dimana datanya mengandung efek spasial (Yulita dkk, 2015). Menurut Fotheringham dkk (2002), cara kerja GWRPCA pada dasarnya sama dengan PCA secara umum. Perbedaan antara GWRPCA dan PCA yaitu GWRPCA menggunakan lokasi terbobot sedangkan PCA tidak menggunakan lokasi terbobot. Pada model GWRPCA, cara untuk mengatasi masalah multikolinearitas adalah dengan membentuk komponen utama pada lokasi terbobot, sehingga dapat dinyatakan kedalam bentuk:

$$\begin{aligned} \sum (u_i, v_i) &= X^T W(u_i, v_i) X \\ \left| \sum (u_i, v_i) - \lambda_i I \right| &= 0 \\ \left(\sum (u_i, v_i) - \lambda_i I \right) \alpha_i &= 0 \end{aligned}$$

Dimana $\sum (u_i, v_i)$ merupakan matriks varian kovarian dari lokasi terbobot (u_i, v_i) .

Langkah pertama yang dilakukan pada GWRPCA adalah melakukan uji heterogenitas spasial pada data untuk melihat apakah metode GWR dapat digunakan. Kemudian, jika terdapat kasus multikolinearitas lokal pada model GWR, maka dilakukan analisis komponen utama untuk menyederhanakan variabel – variabel prediktor yang saling berkorelasi dengan menggunakan PCA. Selanjutnya, dilakukan regresi antara variabel respon dengan komponen utama yang terbentuk dengan menambahkan *bandwith*. Model GWRPCA dapat dituliskan dalam bentuk seperti berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^k \beta_j(u_i, v_i) PC_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (13)$$

Dimana

- y_i : Nilai observasi variabel respon ke- i
- PC_{ij} : Nilai observasi komponen utama ke- j pada lokasi pengamatan ke- i
- $\beta_0(u_i, v_i)$: Konstanta atau intercept pada pengamatan ke- i
- (u_i, v_i) : Menyatakan titik kordinat (*latitude*, *longitude*) lokasi ke- i
- $\beta_j(u_i, v_i)$: Koefisien regresi variabel prediktor ke- j pada lokasi pengamatan ke- i
- ε_i : Error pengamatan ke- i

1.6.10 Model Regresi Nonparametrik Spline dalam GWRPCA

Regresi spline *truncated* yang dipengaruhi oleh masalah heterogenitas spasial merupakan pengembangan dari regresi nonparametrik untuk data spasial dengan estimasi parameter lokal untuk setiap lokasi pengamatan. Pendekatan spline *truncated* digunakan untuk mengatasi masalah analisis spasial dimana kurva regresinya tidak diketahui. Asumsi yang digunakan dari model regresi *error* berdistribusi normal dengan rata-rata nol dan nilai variansi $\sigma^2(u_i, v_i)$ pada setiap lokasi (u_i, v_i) . Koordinat lokasi (u_i, v_i) merupakan faktor yang sangat penting dalam menentukan pembobot yang digunakan untuk mengestimasi parameter dari model. Jika diberikan data $(x_{1i}, x_{2i}, \dots, x_{li}, y_i)$, dan diasumsikan mengikuti model regresi nonparametrik, maka Persamaannya dapat dituliskan sebagai berikut:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{li}) + \varepsilon_i \quad (14)$$

Jika $f(x_{1i}, x_{2i}, \dots, x_{li})$ dilakukan pendekatan fungsi spline *truncated*. Secara matematis, hubungan antara variabel respon dan prediktor pada lokasi ke- i dapat dinyatakan kedalam model GWR nonparametrik dengan pendekatan spline *truncated* sebagai berikut (Sifriyani, 2018).

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pi}^k + \sum_{p=1}^l \sum_{h=1}^r \beta_{p,m+h}(u_i, v_i) (x_{pi} - K_{ph})_+^m + \varepsilon_i \quad (15)$$

Selanjutnya, pada Persamaan (15) merupakan model GWR spline nonparametrik yang diasumsikan bahwa pada data $(x_{1i}, x_{2i}, \dots, x_{li}, y_i)$ tidak terjadi korelasi yang kuat antara variabel prediktornya, dalam hal ini disebut multikolinearitas. Lantas, bagaimana bila terjadi korelasi yang kuat antara variabel prediktornya. Untuk menangani hal tersebut, model yang cocok adalah model GWRPCA dengan pendekatan spline *truncated*. Sehingga Persamaan (15) dapat dinyatakan sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{q=1}^s \sum_{k=1}^m \beta_{qk}(u_i, v_i) w_{qi}^k + \sum_{q=1}^s \sum_{h=1}^r \beta_{q,m+h}(u_i, v_i) (w_{qi} - K_{qh})_+^m + \varepsilon_i \quad (16)$$

Dimana:

- w_{pi} : Nilai observasi komponen utama ke- q pada lokasi ke- i
- K_{qh} : Titik knot ke- h pada observasi komponen utama ke- q
- $\beta_{qk}(u_i, v_i)$: Parameter regresi ke- k untuk nilai observasi komponen utama ke- q pada lokasi ke- i
- $\beta_{q,m+h}(u_i, v_i)$: Parameter regresi fungsi *truncated* ke- $m + h$ pada titik knot ke- h dan nilai observasi komponen utama ke- q pada lokasi ke- i

1.6.11 Weighted Least Square

Metode penaksir pada model GWR adalah menggunakan *Weighted Least Square* (WLS) yaitu dengan memberikan bobot yang berbeda untuk setiap lokasi dimana data tersebut diobservasi. Misalkan pembobot untuk setiap lokasi ke- i adalah

$w_j(u_i, v_i)$ $j = 1, 2, \dots, n$, maka parameter lokasi (u_i, v_i) diestimasi dengan menambahkan unsur pembobot dan kemudian meminimumkan jumlah kuadrat error berikut ini:

$$\sum_{j=1}^n w_j(u_i, v_i) \varepsilon^2 = \sum_{j=1}^n w_j(u_i, v_i) (y_j - \beta_0(u_i, v_i) - \beta_1(u_i, v_i)x_{j1} - \beta_2(u_i, v_i)x_{j2} - \dots - \beta_p(u_i, v_i)x_{jp})^2$$

Misalkan

$$\mathbf{X} = \begin{bmatrix} 1x_{11} & x_{12} & \dots & x_{1p} \\ 1x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1x_{p1} & x_{p2} & \dots & x_{np} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\beta}(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_j(u_i, v_i) \end{bmatrix}$$

Memiliki ordo $\mathbf{X}(n \times (p + 1))$, $\mathbf{Y}(n \times 1)$, $\boldsymbol{\beta}((p + 1) \times 1)$

Dan memiliki Persamaan GWR dalam bentuk matriks:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{W}(u_i, v_i) = \text{diag}[w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i)]$$

Dan $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$

Penyelesaian Persamaan di atas dalam bentuk matriks adalah:

$$\begin{aligned} \boldsymbol{\varepsilon}^T \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon} &= [\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}(u_i, v_i)]^T \mathbf{W}(u_i, v_i) [\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}(u_i, v_i)] \\ &= \mathbf{Y}^T \mathbf{W}(u_i, v_i) \mathbf{Y} - \mathbf{Y}^T \mathbf{W}(u_i, v_i) \mathbf{X}\boldsymbol{\beta}(u_i, v_i) - \boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \\ &\quad + \boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\boldsymbol{\beta}(u_i, v_i) \end{aligned}$$

Karena $\mathbf{X}\boldsymbol{\beta}(u_i, v_i) = \boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T$ maka Persamaan di atas menjadi:

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

Jika Persamaan di atas didiferensialkan terhadap matriks $\boldsymbol{\beta}^T(u_i, v_i)$ dan hasilnya disamakan dengan nol maka didapat:

$$\begin{aligned} -2\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} + 2\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\boldsymbol{\beta}(u_i, v_i) &= 0 \\ \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\boldsymbol{\beta}(u_i, v_i) &= \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \\ (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\boldsymbol{\beta}(u_i, v_i) &= (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \\ \boldsymbol{\beta}(u_i, v_i) &= (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \end{aligned}$$

Sehingga, bentuk penaksir parameter dari model GWR untuk setiap lokasi adalah:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \quad (17)$$

1.6.12 Fungsi Pembobot

Peran pembobot pada model GWR sangat penting karena nilai pembobot mampu mewakili letak data observasi yang satu dengan lainnya. Estimasi parameter pada suatu lokasi pengamatan ke- i lebih dipengaruhi oleh titik – titik yang lebih dekat dibandingkan dengan titik yang berada lebih jauh dari lokasi pengamatan tersebut. Hasil estimasi akan berbeda karena bergantung pada jarak antar lokasi pengamatan. Jarak Euclidean antara lokasi i dengan kordinat (u_i, v_i) ke lokasi j dengan kordinat (u_j, v_j) dapat dihitung dengan menggunakan Persamaan berikut (Fotheringham dkk, 2002).

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (18)$$

Dimana d_{ij} merupakan jarak *Eucliden*.

Ada beberapa literatur yang dapat digunakan dalam menentukan besarnya pembobot untuk masing – masing lokasi yang berbeda pada model GWR, diantaranya dengan menggunakan fungsi kernel. Fungsi kernel dinotasikan $K(u)$ merupakan suatu fungsi yang kontinu, simetris, terbatas, dan $\int_{-\infty}^{\infty} K(u)du = 1$ (Caraka & Yasin, 2017). Secara umum, fungsi kernel terbagi atas dua yaitu *fixed kernel* dan *adaptive kernel*. *Fixed kernel* memiliki *bandwidth* yang sama untuk semua lokasi pengamatan sedangkan *adaptive kernel* memiliki *bandwidth* yang berbeda untuk setiap titik lokasi pengamatan (Pamungkas, 2016). Adapun fungsi *fixed kernel* terbagi atas tiga yaitu sebagai berikut:

1. *Fixed Gaussian*

$$w_j(u_i, v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \quad (19)$$

2. *Fixed Bisquare*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases} \quad (20)$$

3. *Fixed Tricube*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases} \quad (21)$$

Sedangkan, untuk fungsi *adaptive kernel* juga terbagi tiga yaitu sebagai berikut:

1. *Adaptive Gaussian*

$$w_j(u_i, v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_{i(q)}}\right)^2\right) \quad (22)$$

2. *Adaptive Bisquare*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_{i(q)}}\right)^2\right)^2, & \text{untuk } d_{ij} \leq h_{i(q)} \\ 0, & \text{untuk } d_{ij} > h_{i(q)} \end{cases} \quad (23)$$

3. *Adaptive Tricube*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_{i(q)}}\right)^3\right)^3, & \text{untuk } d_{ij} \leq h_{i(q)} \\ 0, & \text{untuk } d_{ij} > h_{i(q)} \end{cases} \quad (24)$$

Dengan:

$w_j(u_i, v_i)$: Pembobot lokasi j pada lokasi i

h : Parameter *bandwidth* ke- i

q : Jarak tetangga terdekat dari lokasi i

1.6.13 Bandwidth Optimum

Bandwidth merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan data. Metode pemilihan *bandwidth* sangat penting digunakan untuk pendugaan fungsi kernel yang tepat (Ariyanto, 2017). Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satu diantaranya adalah metode *Cross Validation* (CV) yang secara matematis didefinisikan sebagai berikut (Caraka & Yasin, 2017).

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad (25)$$

Dengan $\hat{y}_{\neq i}(h)$ adalah penaksir y_i , dimana pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi. Dengan meminimumkan Persamaan (25), maka akan diperoleh nilai *bandwidth* yang optimum (Ariyanto, 2017).

1.6.14 Uji Signifikansi Parameter

Uji signifikansi parameter dilakukan dua tahap, yaitu secara simultan dan secara parsial.

a. Uji Simultan

Pengujian parameter secara simultan bertujuan untuk mengetahui apakah variabel prediktor dari model regresi secara bersama – sama atau simultan mempengaruhi variabel responnya secara signifikan. Bentuk hipotesisnya adalah sebagai berikut (Sifriyani dkk, 2018):

$$H_0 : \beta_{11}(u_i, v_i) = \beta_{12}(u_i, v_i) = \dots = \beta_{lm}(u_i, v_i) = \beta_{1,m+1}(u_i, v_i) = \beta_{1,m+2}(u_i, v_i) = \dots = \beta_{l,m+r}(u_i, v_i)$$

(Tidak ada pengaruh variabel prediktor secara simultan terhadap variabel respon)

$$H_0 : \text{Paling tidak, ada satu } \beta_{pk}(u_i, v_i) \neq 0 \text{ atau } \beta_{p,m+h}(u_i, v_i) \neq 0; p = 1, \dots, l; k = 1, \dots, j; h = 1, \dots, r; i = 1, \dots, n$$

Dengan statistik uji:

$$V^* = \frac{\left(\frac{\mathbf{y}^T \mathbf{M}(u_i, v_i) \mathbf{y}}{\text{tr}((\mathbf{I} - \mathbf{B}_\omega)^T (\mathbf{I} - \mathbf{B}_\omega))} \right)}{\left(\frac{\mathbf{y}^T \mathbf{D}(u_i, v_i) \mathbf{y}}{\text{tr}((\mathbf{I} - \mathbf{D})^T (\mathbf{I} - \mathbf{D}))} \right)} \quad (26)$$

Dimana $\mathbf{M}(u_i, v_i) = ((\mathbf{I} - \mathbf{B}_\omega)^T (\mathbf{I} - \mathbf{B}_\omega))$ dengan $\mathbf{y}$ adalah vektor variabel respon dan $\mathbf{I}$ adalah matriks identitas. Kriteria penolakan H_0 adalah tolak H_0 jika $F_{\text{hitung}}(V^*) > F_{(a; db_1; db_2)}$ atau $p - \text{value} < \alpha$.

b. Uji Parsial

Pengujian parameter secara parsial bertujuan untuk mengetahui variabel prediktor apa saja yang mempengaruhi secara individu dari model regresi terhadap responnya secara signifikan. Bentuk hipotesisnya adalah sebagai berikut (Sifriyani dkk, 2018):

$$H_0 : \beta_{pk}(u_i, v_i) = \beta_{p,m+h}(u_i, v_i) = 0$$

(Tidak ada pengaruh variabel prediktor secara parsial terhadap variabel respon)

$$H_0 : \text{Paling tidak ada satu } \beta_{pk}(u_i, v_i) \neq 0 \text{ atau } \beta_{p,m+h}(u_i, v_i) \neq 0; p = 1, \dots, l; k = 1, \dots, j; \\ h = 1, \dots, r; i = 1, \dots, n$$

Dengan statistik uji:

$$t = \frac{\hat{\eta}(u_i, v_i)}{SE(\hat{\eta}(u_i, v_i))} \quad (27)$$

$\hat{\eta}(u_i, v_i)$ adalah estimator dari $\eta(u_i, v_i)$ dan $SE(\hat{\eta}(u_i, v_i))$ adalah elemen diagonal ke $k + 1$ dari matriks $(Q^T W(u_i, v_i) Q)^{-1} \hat{\sigma}^2(u_i, v_i)$, Q adalah matriks ukuran $n \times 1[1 + (d \times p) + (d \times r)]$ yang memuat variabel respon dan variabel prediktor yang mempunyai komponen spline *truncated*. Kriteria penolakan H_0 adalah tolak H_0 jika $|t_{Hitung}| > t_{(\frac{\alpha}{2}, db)}$ atau $p - value < \alpha$.

1.6.15 Koefisien Determinasi

Koefisien Determinasi merupakan suatu pengukuran yang penting dalam mengestimasi baik atau tidaknya model regresi yang terbentuk. Nilai koefisien determinasi mampu menggambarkan seberapa besar variasi dari variabel respon dapat dijelaskan oleh variabel prediktornya. Sehingga, baik atau tidaknya suatu model regresi ditentukan oleh nilai R^2 nya dengan rentang nilai nol sampai satu. Persamaan R^2 dituliskan seperti berikut:

$$R^2 = 1 - \frac{SSE}{SST} \quad (28)$$

dengan:

$$SSE \text{ (Sum of Summary Error)} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SST \text{ (Sum of Summary Total)} = \sum_{i=1}^n (y_i - \bar{y})^2$$

dimana:

y_i : Pengamatan variabel respon ke- i

$\hat{y}_i$: Estimasi variabel respon ke- i

$\bar{y}$: Rata – rata variabel respon

1.6.16 Diabetes Mellitus

Diabetes mellitus (DM) didefinisikan sebagai suatu penyakit atau gangguan metabolisme kronis dengan multi etiologi yang ditandai dengan tingginya kadar gula darah disertai dengan gangguan metabolisme karbohidrat, lipid, dan protein sebagai akibat insufisiensi fungsi insulin. Insufisiensi fungsi insulin disebabkan karena kurang responsifnya sel – sel tubuh terhadap insulin.

Diabetes mellitus adalah penyakit kronis yang mengganggu kemampuan tubuh dalam mengolah gula (glukosa) darah menjadi energi. Saat jumlah hormone insulin dalam tubuh tidak mencukupi untuk mengubah glukosa menjadi energy, maka glukosa menetap di dalam darah. Sel tubuh yang kebal terhadap insulin atau

resistensi insulin juga menjadi penyebab diabetes. Penyakit diabetes dapat disebabkan oleh berbagai hal, mulai dari faktor keturunan, pengaruh lingkungan, hingga gaya hidup yang tidak sehat (Dinas Kesehatan Provinsi Sulawesi Selatan, 2020).

Berdasarkan penelitian yang dilakukan Anri (2020), disimpulkan bahwa, indeks massa tubuh, aktivitas fisik, dan pola makan berhubungan dengan kejadian diabetes mellitus. Orang yang obesitas berisiko 3,1 kali menderita DM dibandingkan dengan orang yang tidak obesitas. Orang yang pola makannya tidak seimbang berisiko 3,8 kali menderita DM dibandingkan dengan orang yang pola makannya seimbang. Dan orang yang kurang melakukan aktivitas fisik berisiko 2,9 kali menderita DM dibandingkan dengan orang yang cukup melakukan aktivitas sehari. Penjelasan mengenai faktor – faktor yang mempengaruhi kejadian diabetes mellitus sebagai berikut:

1. Indeks Massa Tubuh

Indeks massa tubuh (IMT) merupakan salah cara pengukuran berat badan yang disesuaikan dengan tinggi badan. IMT digunakan dalam mengklasifikasikan kelebihan berat badan dan obesitas pada orang dewasa. Berikut klasifikasi perhitungan IMT disajikan pada tabel berikut:

Tabel 1. Klasifikasi kriteria IMT Asia Pasifik

Klasifikasi	Indeks Massa Tubuh
<i>Underweight</i> (Berat badan kurang)	< 18,5
Normal	18,5 – 22,9
<i>Overweight</i> (Berat badan berlebih)	≥ 23
Beresiko	23 – 24,9
Obes I	25 – 29,9
Obes II	≥ 30

Sumber: Sugondo 2009 (dalam Rasyid, 2021)

Menurut Sugondo (2009) (dalam Rasyid, 2021), hasil perhitungan IMT diklasifikasikan berdasarkan klasifikasi menurut Kriteria Asia Pasifik menjadi *underweight*, normal, *overweight*.

2. Aktivitas Fisik

Aktivitas fisik dapat mengontrol kadar gula darah. Glukosa akan diubah menjadi energi pada saat beraktivitas fisik. Aktivitas fisik mengakibatkan insulin semakin meningkat sehingga kadar gula dalam darah akan berkurang. Pada orang yang kurang melakukan aktivitas fisik, zat makanan yang masuk ke dalam tubuh tidak dibakar tetapi ditimbun dalam tubuh sebagai lemak dan gula dalam bentuk glukosa (Anri, 2020). Menurut kemenkes (2018), kurang aktivitas fisik adalah kegiatan kumulatif kurang dari 150 menit seminggu. Semakin berat aktivitas yang dilakukan, maka semakin sedikit kemungkinan terkena DM (Veridiana & Nurjana, 2019).

3. Pola Makan

Pola makan merupakan suatu cara dalam pengaturan jumlah dan jenis makanan dengan maksud tertentu seperti mempertahankan kesehatan, status nutrisi, mencegah atau membantu penyembuhan penyakit. Pola makan sehari – hari merupakan pola makan seseorang yang berhubungan dengan kebiasaan makan setiap harinya (Anri, 2020). Pola konsumsi makan yang tidak sehat yaitu tinggi gula, tinggi garam, dan tinggi lemak merupakan salah satu faktor penyebab terjadinya penyakit tidak menular seperti diabetes mellitus (Triadhini dkk, 2018). Batas konsumsi gula, garam dan lemak per orang perharinya di sajikan pada tabel berikut:

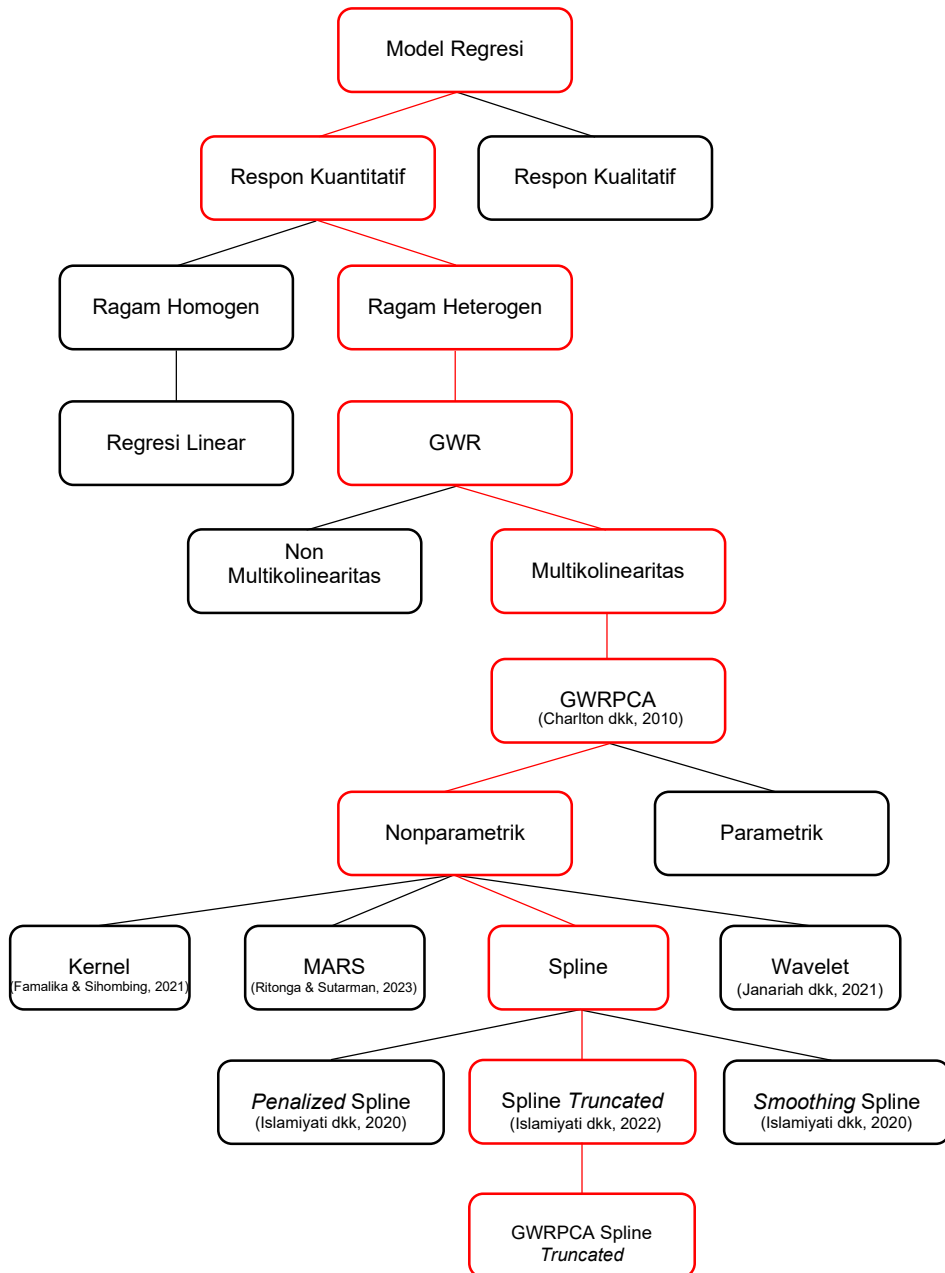
Tabel 2. Batas Konsumsi Gula, Garam, dan Lemak

Jenis Makronutrisi	Batas Konsumsi Perhari
Gula	$\leq 50 \text{ gram}$
Garam	$\leq 2000 \text{ miligram}$
Lemak	$\leq 67 \text{ gram}$

Sumber: Permenkes No. 30 tahun 2013

1.6.17 Kerangka Konsep

Salah satu bagian penting dalam penelitian adalah menyusun kerangka konsep. Kerangka konsep merupakan cara yang digunakan untuk menjelaskan kaitan atau hubungan antara variabel yang akan diteliti. Adapun gambaran kerangka konsep pada penelitian ini adalah seperti pada gambar berikut:



Gambar 1. Kerangka Konsep

BAB II METODE PENELITIAN

2.1 Sumber Data

Penelitian ini menggunakan data sekunder, yaitu data kasus diabetes mellitus di Sulawesi Selatan tahun 2022. Cakupan data meliputi seluruh Kabupaten/Kota di Sulawesi Selatan. Data diperoleh dari Dinas Kesehatan Provinsi Sulawesi Selatan.

2.2 Variabel Penelitian

Pada Penelitian ini, Variabel yang akan digunakan adalah satu variabel respon (y) dan 7 variabel prediktor (x). Adapun variabel – variabel yang akan digunakan disajikan dalam tabel berikut.

Tabel 3. Variabel penelitian

Jenis Variabel	Kode	Keterangan	Satuan
Respon	y	Kasus Diabetes Mellitus	Jiwa
Prediktor	x_1	Aktivitas Fisik Kurang	Jiwa
	x_2	Indeks Massa Tubuh Normal	Jiwa
	x_3	Indeks Massa Tubuh Berat Badan Berlebih	Jiwa
	x_4	Indeks Massa Tubuh Obesitas	Jiwa
	x_5	Konsumsi Gula Berlebih	Jiwa
	x_6	Konsumsi Garam Berlebih	Jiwa
	x_7	Konsumsi Lemak Berlebih	Jiwa

2.3 Langkah – Langkah Penelitian

Pada penelitian ini, *software* yang akan digunakan untuk analisis data adalah *RStudio*. Adapun langkah – langkah yang akan dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan analisis deskriptif pada data untuk mengetahui gambaran umum kasus diabetes mellitus di Sulawesi Selatan beserta variabel – variabel prediktornya.
2. Membuat *scatter plot* untuk melihat pola hubungan antara variabel respon dengan masing - masing variabel prediktor.
3. Melakukan uji asumsi heterogenitas menggunakan uji *Breusch Pagan* untuk mengetahui adanya efek spasial pada data menggunakan Persamaan (8).
4. Melakukan pendeteksian multikolinearitas untuk mengetahui adanya korelasi yang kuat antara variabel prediktor menggunakan nilai VIF pada Persamaan (9).
5. Melakukan analisis PCA untuk membentuk komponen utama yang akan digunakan pada analisis selanjutnya.

6. Menghitung jarak *euclidean* antara lokasi lokasi ke- i terhadap lokasi ke- j menggunakan Persamaan (18).
7. Menentukan fungsi pembobot dengan memperhatikan nilai *bandwith* dan CV minimum
8. Menentukan titik knot optimum dengan metode GCV sesuai dengan Persamaan (7)
9. Menyusun model GWRPCA spline dengan titik knot optimal dan fungsi pembobot
10. Interpretasi model
11. Penarikan kesimpulan