

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi merupakan suatu metode analisis statistik yang menjelaskan hubungan antara satu atau lebih variabel independen terhadap satu variabel dependen (Zulkarnain dkk., 2020). Variabel dependen pada analisis regresi dapat berupa kualitatif dan kuantitatif, jika variabel dependen bersifat kualitatif maka digunakan model regresi logistik (Enus dkk., 2022). Variabel dependen pada model regresi logistik dapat berupa biner, ordinal, maupun multinomial. Variabel yang terdiri atas tiga kategori atau lebih dan memiliki tingkatan berbentuk ordinal, variabel yang terdiri atas tiga kategori atau lebih dan tidak memiliki tingkatan berbentuk multinomial, sedangkan yang paling sederhana adalah variabel dependen berbentuk biner karena hanya terdiri dua kategori, yaitu bernilai 0 atau 1 (Saragih dkk., 2020).

Pemodelan regresi logistik biner telah mengalami perkembangan, salah satunya dengan menggabungkannya dengan pendekatan regresi nonparametrik. Pendekatan ini memiliki kemampuan dalam mengikuti variasi data sehingga dapat menambah akurasi pada pemodelan. Para peneliti telah mengembangkan pendekatan ini menjadi beberapa estimator seperti *spline truncated*, *kernel*, dan *deret fourier*. *Spline truncated* adalah estimator yang paling sering digunakan yang dicirikan dengan potongan polinomial tersegmen dan kontinu. Dalam teori dan aplikasi regresi, polinomial tersegmen ini memiliki peranan penting, yaitu memberikan sifat fleksibilitas yang lebih baik dari polinomial sederhana dan dapat memungkinkan untuk penyesuaian bentuk kurva sehingga dapat mengikuti karakteristik data (Dani & Adrianingsih, 2021).

Penggabungan metode regresi logistik biner dengan *spline truncated* telah dilakukan oleh beberapa peneliti diantaranya penelitian Islamiyati dkk. (2023) yang menggunakan pemodelan regresi logistik *spline truncated* pada data status gizi balita di Kabupaten Barru. Penelitian lainnya dilakukan oleh Arifin dkk. (2023) yang menggunakan pemodelan regresi logistik *spline truncated* pada data status gizi balita di Kabupaten Gowa. Adapun penelitian yang dilakukan oleh Chen dkk. (2024) menggunakan pemodelan regresi logistik *restricted cubic spline* untuk menganalisis hubungan antara Indeks *Immune-Inflamasi Sistemik* (SII) dan risiko malnutrisi di kalangan pasien yang dirawat di rumah sakit China. Namun, masalah multikolinieritas belum dipertimbangkan pada penelitian tersebut.

Multikolinieritas merupakan salah satu asumsi yang harus diperhatikan dalam pemodelan regresi logistik. Pelanggaran asumsi ini terjadi ketika antar variabel independen memiliki korelasi yang tinggi (Azhari & Fitriani, 2022). Dampak dari pelanggaran asumsi ini yaitu tidak efisiennya hasil dari estimasi parameter regresi karena mempunyai bias dan variansi yang besar. Analisis komponen utama (AKU) merupakan salah satu pendekatan yang efektif untuk mengatasi masalah multikolinieritas. Pendekatan ini merupakan pembentukan komponen utama yang tidak saling berkorelasi berasal dari kombinasi linier dari variabel independen. Kelebihan dari pendekatan ini adalah mereduksi dimensi data, yaitu mengubah

dimensi data yang banyak menjadi yang lebih sedikit namun tetap menyerap sebagian besar jumlah varian (Harahap dkk., 2023).

Penelitian terdahulu yang menggunakan AKU diantaranya Pratama dkk. (2023) menganalisis setimen kendaraan listrik menggunakan media sosial twitter memperoleh hasil akurasi dengan menggunakan regresi logistik sebesar 87,9%, sedangkan dengan menggunakan regresi logistik AKU akurasi meningkat menjadi 90%. Penelitian lainnya dilakukan oleh Azhari & Fitriani (2022) yang memprediksi risiko pasien terkena penyakit jantung memperoleh hasil akurasi dengan menggunakan regresi logistik sebesar 83% sedangkan dengan regresi logistik AKU akurasi meningkat menjadi 85%. Selain pada regresi logistik, penggunaan AKU juga telah dilakukan dengan *spline truncated* seperti yang dilakukan oleh Arifin dkk. (2020) untuk menganalisis data diabetes melitus tahun 2014-2018. Pada penelitian tersebut belum menggunakan pemodelan regresi logistik komponen utama dengan *spline truncated* secara bersamaan yang dapat diterapkan di berbagai bidang termasuk kesehatan.

Masalah gizi merupakan salah satu masalah yang urgen dalam bidang kesehatan di Indonesia. *Wasting* yang merujuk pada orang yang terlalu kurus untuk tinggi badannya dan *stunting* yang merujuk pada orang yang terlalu pendek untuk usianya merupakan dua bentuk masalah gizi yang memiliki risiko kematian yang tinggi. Kedua kondisi ini telah menjadi perhatian pemerintahan Indonesia karena tingkat prevalensi *wasting* dan *stunting* di Indonesia secara beruntun 8,50% dan 21,16% yang belum mencapai sesuai pada Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024, pemerintah menargetkan tingkat prevalensi *wasting* sebesar 7% dan tingkat prevalensi *stunting* sebesar 14% (Kementerian Kesehatan, 2021). Kabupaten Gowa merupakan salah satu kabupaten di Indonesia yang memiliki tingkat prevalensi *wasting* dan *stunting* yang hampir serupa dengan kondisi di Indonesia. Hal ini terlihat pada tahun 2023, tingkat prevalensi *wasting* dan *stunting* di Kabupaten Gowa secara berurut sebesar 8,42% dan 21,10% yang belum mencapai dan perlu penurunan yang cukup jauh dari target Indonesia (Hijrah, 2024).

Dengan demikian, penelitian ini akan mengklasifikasikan data *wasting* pada balita *stunting* di Kabupaten Gowa dengan menggunakan pemodelan regresi logistik biner komponen utama untuk kasus multikolinieritas dengan *spline truncated*. Dalam penelitian ini, variabel dependen yang digunakan adalah *wasting* yang berbentuk biner. Data tersebut diklasifikasikan guna mengetahui balita yang mengalami *wasting* dan *stunting* secara bersamaan karena kondisi ini memiliki risiko kematian yang sangat tinggi. Klasifikasi dilakukan dengan menggunakan variabel independen yang diduga memengaruhi *wasting*. Penelitian sebelumnya oleh Hawa Nur Rohmah & Rahayu Nadhiroh (2024) telah menggunakan berat badan lahir dan tinggi badan lahir sebagai variabel independennya. Adapun penelitian oleh Arifin dkk. (2023) telah menggunakan usia, berat badan saat pengukuran, dan tinggi badan saat pengukuran sebagai variabel independennya. Maka dari itu, penelitian ini diharapkan dapat memberikan hasil klasifikasi yang akurat berdasarkan variabel independen yang digunakan sehingga dapat mengurangi tingkat prevalensi *wasting* maupun *stunting*.

## 1.2 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan, maka diperoleh tujuan penelitian berikut:

1. Mendapatkan estimasi parameter model regresi logistik biner komponen utama pada data multikolinieritas dengan *spline truncated*.
2. Mendapatkan model regresi logistik biner komponen utama pada data multikolinieritas dengan *spline truncated* menggunakan data status *wasting* balita *stunting* di Kabupaten Gowa pada tahun 2023.

## 1.3 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini yaitu :

1. Menambah wawasan dan pengetahuan dalam bidang ilmu statistika khususnya dalam regresi logistik, *spline truncated*, dan analisis komponen utama.
2. Menyediakan model klasifikasi data status *wasting* balita *stunting* yang akurat.
3. Dapat menjadi bahan rujukan bagi pemerintah untuk lebih memperhatikan status gizi balita.

## 1.4 Batasan Masalah

Agar ruang lingkup permasalahan menjadi terarah, penelitian ini memberikan batasan masalah sebagai berikut:

1. Jumlah variabel independen yang digunakan hanya lima variabel.
2. Kategori yang digunakan pada variabel dependen adalah normal dan *wasting*.
3. Pemodelan *spline truncated* hanya dibatasi pada orde satu (linier), dua (kuadrat), dan tiga (kubik).
4. Metode estimasi parameter yang digunakan adalah *Maximum Likelihood Estimation* (MLE).
5. Pemilihan titik *knot* optimal berdasarkan metode *Generalized Cross-Validation* (GCV).
6. Jumlah titik *knot* yang digunakan satu hingga tiga titik *knot*.

## 1.5 Teori

### 1.5.1 Multikolinieritas

Multikolinieritas adalah salah satu asumsi dalam regresi yang digunakan untuk menguji terdapat atau tidaknya hubungan atau korelasi variabel independen dalam suatu model regresi. Pelanggaran asumsi ini terjadi ketika antar variabel independen memiliki korelasi yang tinggi. Pelanggaran asumsi ini tidak ditemukan pada model regresi yang baik. Dampak dari pelanggaran asumsi ini yaitu tidak efisiennya hasil dari estimasi parameter regresi karena mempunyai bias dan variansi yang besar (Alwi dkk., 2023).

Multikolinieritas dapat dideteksi dengan menggunakan korelasi *pearson*. Metode ini adalah salah satu perhitungan koefisien korelasi yang digunakan untuk melihat hubungan diantara variabel. Nilai korelasi *pearson* antar variabel independen dapat dihitung seperti pada Persamaan (1) (Pendi, 2021).

$$r_{ab} = \frac{n \sum x_a x_b - (\sum x_a)(\sum x_b)}{\sqrt{(n \sum x_a^2 - (\sum x_a)^2)} \sqrt{(n \sum x_b^2 - (\sum x_b)^2)}} ; a = b = 1, 2, \dots, p \quad (1)$$

dengan  $r_{ab}$  adalah nilai koefisien korelasi *Pearson*,  $x_a$  dan  $x_b$  adalah variabel independen yang diuji,  $n$  adalah jumlah amatan data, dan  $p$  adalah jumlah variabel independen dalam model.

Kriteria pengujian multikolinieritas dengan menggunakan korelasi, yaitu nilai korelasi  $\geq 0,80$  berarti antar variabel independen tersebut saling berhubungan sehingga terjadi multikolinieritas, sedangkan nilai korelasi  $< 0,80$  berarti antar variabel independen tersebut tidak ada hubungan maupun multikolinieritas (Purwita & Sari, 2023). Setelah perhitungan korelasi antara variabel independen dilakukan, maka dibentuklah matriks korelasi seperti berikut (Numpacharoen & Atsawarungruangkit, 2012).

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix}$$

dengan  $R$  merupakan matriks korelasi berorde  $p \times p$ .

### 1.5.2 Analisis Komponen Utama

Analisis komponen utama (AKU) merupakan suatu teknik statistik untuk mengubah variabel independen saling berkorelasi satu dengan yang lainnya menjadi komponen utama yang diperoleh berdasarkan pada kombinasi linier dari variabel independen. Penggunaan AKU memiliki kelebihan untuk membentuk satu set variabel baru (komponen utama optimal) dengan dimensi yang lebih kecil. Sehingga dengan menggunakan AKU variabel akan tereduksi dimensinya namun tetap menyerap sebagian besar jumlah variansnya (Harahap dkk., 2023).

Nilai eigen dan vektor eigen diperlukan untuk membentuk komponen utama. Dalam konsep regresi, nilai eigen dan vektor eigen merupakan metode untuk menganalisis matriks korelasi. Nilai eigen dapat diperoleh dari matriks korelasi dengan sifat jika dan hanya jika  $R - \lambda I$  adalah matriks singular. Persamaan (2) menunjukkan perhitungan nilai eigen ( $\lambda$ ) dengan menggunakan matriks korelasi (Strang, 2006).

$$\det(R - \lambda I) = 0 \quad (2)$$

dengan  $R$  merupakan matriks korelasi berorde  $p \times p$ ,  $\lambda$  merupakan nilai eigen, dan  $I$  adalah matriks identitas berorde  $p \times p$ .

Persamaan (2) dinamakan dengan persamaan karakteristik yang digunakan untuk memperoleh nilai eigen. Adapun vektor eigen yang dapat diperoleh setelah perhitungan nilai eigen. Vektor eigen dapat diperoleh dengan menggunakan sifat *null space* dari  $R - \lambda I$ . Persamaan (3) menunjukkan perhitungan vektor eigen ( $V$ ) dengan menggunakan matriks korelasi (Strang, 2006).

$$(R - \lambda I)V = 0 \quad (3)$$

dengan  $V$  merupakan matriks vektor eigen berorde  $p \times p$ .

Pendefinisian komponen utama dilakukan dengan menggunakan pasangan nilai eigen dan vektor eigen secara ortonormal yaitu  $(\lambda_1, \mathbf{v}_1), (\lambda_2, \mathbf{v}_2), \dots, (\lambda_p, \mathbf{v}_p)$  dengan  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . Persamaan (4) menunjukkan pedefinisian Komponen utama ke- $a$  ( $\mathbf{W}_a$ ).

$$\mathbf{W}_a = \mathbf{V}'\mathbf{X} = v_{1j}\mathbf{x}_1 + v_{2j}\mathbf{x}_2 + \dots + v_{pa}\mathbf{x}_a, \quad a = 1, 2, \dots, p \quad (4)$$

Persamaan (4) dapat dinyatakan ke dalam bentuk matriks seperti yang ditunjukkan oleh Persamaan (5).

$$\mathbf{W}_a = \mathbf{V}'\mathbf{X} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1p} \\ v_{21} & v_{22} & \dots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \dots & v_{pp} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{bmatrix} \quad (5)$$

Komponen pertama ( $\mathbf{w}_1$ ) dengan memaksimalkan nilai  $\mathbf{v}'_1\mathbf{R}\mathbf{v}_1 = \lambda_1$ , sedangkan komponen kedua ( $\mathbf{w}_2$ ) memenuhi sisa keragaman selain  $\mathbf{w}_1$  dengan memaksimalkan nilai  $\mathbf{v}'_2\mathbf{R}\mathbf{v}_2 = \lambda_2$ , dan seterusnya hingga komponen ke- $p$  ( $\mathbf{w}_p$ ) memenuhi sisa keragaman selain  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{p-1}$  dengan memaksimalkan nilai  $\mathbf{v}'_p\mathbf{R}\mathbf{v}_p = \lambda_p$  (Jolliffe & Cadima, 2016). Kriteria dalam pemilihan komponen utama optimal dapat digunakan visualisasi *scree plot* yang terjadi perubahan nilai eigen yang besar. Sehingga dengan melakukan pemilihan komponen utama optimal, maka dimensi akan menjadi lebih kecil (Pendi, 2021).

### 1.5.3 Regresi Logistik Biner

Regresi logistik biner adalah metode yang digunakan untuk mencari hubungan antara variabel independen dan variabel dependen yang bersifat *dichotomous* (Safitri dkk., 2019). Sifat ini berarti bahwa variabel dependen pada regresi logistik biner terdiri dua kategori, misalkan  $y = 1$  menyatakan hasil yang diperoleh 'sukses' dan  $y = 0$  menyatakan hasil yang diperoleh 'gagal'. Model regresi logistik biner termasuk ke dalam distribusi Bernoulli, yaitu distribusi dari variabel dependen yang hanya mempunyai dua kategori yaitu 0 dan 1. Fungsi peluang regresi logistik biner seperti pada Persamaan (6) (M. Utami dkk., 2024).

$$f(y_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad ; \quad y_i = 0, 1 \quad (6)$$

dengan  $\pi(x_i)$  merupakan probabilitas yang bergantung pada  $x_i$ ,  $i$  merupakan indeks yang mengacu pada data ke- $i$ ,  $y_i$  merupakan variabel dependen ke- $i$  yang menunjukkan hasil 0 atau 1.

Hosmer & Lemeshow (2000) berpendapat bahwa peluang kejadian  $\pi(x_i)$  dapat dinyatakan seperti pada Persamaan (7).

$$\pi(x_i) = \frac{\exp(\beta_0 + \sum_{a=1}^p \beta_a x_{ai})}{1 + \exp(\beta_0 + \sum_{a=1}^p \beta_a x_{ai})} \quad (7)$$

Hubungan variabel dependen dan independen dapat diketahui dengan memperoleh fungsi linier terlebih dahulu. Fungsi linier diperoleh dengan melakukan transformasi  $\pi(x_i)$  ke dalam bentuk logit. Adapun bentuk logit pada  $\pi(x_i)$  seperti pada Persamaan (8).

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \sum_{a=1}^p \beta_a x_{ai} \quad (8)$$

atau dapat dijabarkan menjadi  $g(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$ .

#### 1.5.4 Maximum Likelihood Estimation

*Maximum Likelihood Estimation* (MLE) atau estimasi maksimum likelihood digunakan untuk mengestimasi parameter model  $\beta$  dengan cara memaksimalkan kemungkinan (*likelihood*) bahwa model yang dibangun sesuai dengan data yang diamati yang berdistribusi probabilitas tertentu. Fungsi *likelihood* merupakan perkalian fungsi peluang. Variabel independen  $x_a (x_1, x_2, \dots, x_p)$  memiliki Fungsi *likelihood*  $L(\beta)$  yang dapat ditunjukkan pada Persamaan (9) (Utami dkk., 2022).

$$L(\beta) = \prod_{i=1}^n f(y_i) \quad (9)$$

Fungsi peluang yang digunakan berdistribusi bernouli sehingga fungsi *likelihood*  $L(\beta)$  dapat dinyatakan dalam Persamaan (10).

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (10)$$

Fungsi *likelihood* pada Persamaan (10) akan lebih mudah dalam perhitungan dengan memaksimumkan ke bentuk fungsi *log-likelihood* ( $\ell(\beta)$ ) seperti yang ditunjukkan oleh Persamaan (11).

$$\begin{aligned} \ell(\beta) &= \ln(L(\beta)) = \ln\left[\prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}\right] \\ \ell(\beta) &= \sum_{i=1}^n [y_i \ln \pi(x_i) + (1 - y_i) \ln(1 - \pi(x_i))] \end{aligned} \quad (11)$$

Persamaan (11) diturunkan terhadap parameter  $\beta$  dan menyamakannya dengan nol dengan tujuan untuk menemukan estimasi parameter  $\beta$  yang memaksimalkan fungsi *likelihood*. Persamaan (12) menyatakan Turunan pertama terhadap parameter  $\beta$  dari Persamaan (11), yaitu:

$$\begin{aligned} \frac{\partial \ell(\beta)}{\partial \beta_a} &= \sum_{i=1}^n \left[ \frac{y_i}{\pi(x_i)} \frac{\partial \pi(x_i)}{\partial \beta_a} - \frac{1 - y_i}{1 - \pi(x_i)} \frac{\partial (1 - \pi(x_i))}{\partial \beta_a} \right] \\ \frac{\partial \ell(\beta)}{\partial \beta_a} &= \sum_{i=1}^n (y_i - \pi(x_i)) x_{ai} = 0 \end{aligned} \quad (12)$$

Hasil turunan pertama belum bisa diselesaikan secara analitik sehingga untuk memaksimalkan fungsi *log-likelihood* digunakan metode numerik seperti algoritma *newton raphson*. Algoritma ini memperbarui parameter secara iteratif dengan menggunakan turunan pertama dari fungsi *log-likelihood* dan matriks *hessian*

yang merupakan turunan kedua dari fungsi *log-likelihood*. Proses iteratif ini dimulai dengan tebakan awal untuk parameter  $\beta$ , kemudian menghitung turunan pertama dan keduanya untuk memperbarui estimasi parameter dengan menggunakan Persamaan (13), yaitu:

$$\beta^{(t+1)} = \beta^{(t)} - H^{-1} \ell'(\beta^{(t)}) \quad (13)$$

dengan  $\beta^{(t)}$  adalah estimasi parameter pada iterasi ke- $t$ ,  $H$  adalah matriks *hessian* atau turunan kedua dari fungsi *log-likelihood*, dan  $\ell'(\beta^{(t)})$  adalah vektor turunan pertama dari fungsi *log-likelihood*. Proses ini berlanjut hingga perbedaan antara  $\beta^{(t+1)}$  dan  $\beta^{(t)}$  sudah tidak signifikan atau konvergen. Sehingga nilai  $\beta$  yang memaksimalkan nilai *log-likelihood* memberikan estimasi terbaik untuk parameter model yang digunakan (Purba, 2020).

### 1.5.5 Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan yang dapat digunakan dalam mengestimasi pola kurva regresi. Jika data tidak terkait dengan asumsi bentuk kurva yang berpola sehingga digunakan pendekatan regresi nonparametrik (Pratiwi, 2020). Fleksibilitas pendekatan ini adalah mencari bentuk pola kurva regresinya dan tidak dipengaruhi oleh pemikiran subjektif dari peneliti sehingga modelnya dapat berbentuk fungsi linier maupun nonlinier merupakan ciri dari pemodelan regresi nonparametrik. Pendekatan ini telah berkembang menjadi beberapa estimator seperti *spline truncated*, *kernel*, dan *deret fourier* (Dani & Adrianingsih, 2021). Secara umum, Persamaan (14) menunjukkan pemodelan regresi nonparametrik.

$$y_i = f(x_i) + \varepsilon_i ; i = 1, 2, \dots, n \quad (14)$$

dengan  $y_i$  adalah variabel dependen,  $f(x_i)$  fungsi atau model yang menjelaskan hubungan antara  $x_i$  dan  $y_i$  yang akan diestimasi,  $x_i$  variabel independen, dan  $\varepsilon_i$  adalah eror.

### 1.5.6 Spline Truncated

Estimator *spline truncated* merupakan bagian dari pendekatan dalam regresi nonparametrik yang dicirikan dengan potongan polinomial tersegmen dan kontinu. Dalam teori dan aplikasi bidang statistika, polinomial tersegmen ini memiliki peranan penting di antaranya, memberikan sifat yang lebih baik dari polinomial sesederhana dan fleksibilitas sehingga sifat ini dapat menunjukkan kemungkinan penyesuaian bentuk kurva lebih mengikuti karakteristik data. Titik-titik *knot* tidak lepas dari pendekatan yaitu titik kombinasi bersama yang mengindikasikan terlihat hubungan pola perilaku kurva data yang berubah ketika sub interval tertentu. Jika sembarang fungsi  $f$  dalam ruang *spline* dengan orde  $q$  pada titik *knot*  $k_1, k_2, \dots, k_h$  maka persamaannya dapat ditunjukkan oleh Persamaan (15) (Purnaraga dkk., 2020).

$$y(x) = \beta_0 + \sum_{k=1}^q \beta_k x_i^k + \sum_{h=1}^r \beta_{q+h} (x_i - k_h)_+^q + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (15)$$

Fungsi *truncated* didefinisikan sebagai berikut :

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

dengan koefisien polinomial dinotasikan dengan  $\beta_k$ , koefisien komponen *truncated* dinotasikan dengan  $\beta_{q+h}$ , orde dinotasikan dengan  $q$ , jumlah titik *knot* dinotasikan dengan  $r$ , dan posisi titik *knot* dinotasikan dengan  $k_h$ . Persamaan (15) dapat dinyatakan ke dalam Persamaan (16) untuk setiap data pengamatan  $i = 1, 2, \dots, n$ .

$$y(x) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_q x_i^q + \beta_{q+1} (x_i - k_1)_+^q + \beta_{q+2} (x_i - k_2)_+^q + \dots + \beta_{q+r} (x_i - k_r)_+^q + \varepsilon_i \quad (16)$$

### 1.5.7 Pemilihan Titik *Knot* Optimal

Penentuan model pendekatan regresi nonparametrik dengan estimator *spline truncated* penting untuk dilakukan penentuan titik *knot* optimal karena dapat memengaruhi baik atau tidaknya suatu model. Titik *knot* adalah perpaduan dari dua interval yang terjadi perubahan garis perilaku pola. *Cross Validation* (CV) dan *Generalized Cross Validation* (GCV) merupakan metode yang dapat digunakan dalam penentuan titik *knot* dengan pendekatan model *spline* (Islamiyati, 2017). Metode GCV merupakan metode penentuan yang paling sering digunakan dalam penentuan titik *knot* terbaik atau optimal, metode ini adalah hasil modifikasi dari metode CV. Nilai GCV terendah atau minimum mengindikasikan titik *knot* optimal. Persamaan GCV ditunjukkan pada Persamaan (17) (Nurhuda dkk., 2022).

$$GCV(\bar{k}) = \frac{MSE}{[n^{-1} \text{trace}[\mathbf{I} - \mathbf{A}]]^2} \quad (17)$$

dengan  $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ ,  $y_i$  adalah variabel dependen,  $\hat{y}_i$  adalah nilai estimasi variabel dependen,  $n$  adalah jumlah amatan,  $\mathbf{I}$  adalah matriks identitas, dan  $\mathbf{A}$  adalah matriks estimasi dengan  $\mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ .

### 1.5.8 Uji Signifikansi

Uji signifikansi bertujuan untuk menguji signifikansi parameter yang digunakan. Uji signifikansi dapat berupa uji simultan dan uji parsial. Uji simultan adalah pengujian signifikansi parameter yang digunakan untuk mengetahui pengaruh secara simultan variabel independen terhadap variabel dependen dengan menggunakan uji rasio *likelihood* (Harahap dkk., 2023).

Hipotesis yang digunakan, yaitu:

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

$$H_1: \text{Minimal ada satu } \beta_a \neq 0 \text{ untuk } a = 1, 2, \dots, p$$

Statistik uji ditunjukkan oleh Persamaan (18).

$$G_{hitung} = -2 \ln \left( \frac{L_o}{L_a} \right) \quad (18)$$

dengan  $L_o$  adalah nilai *likelihood* tanpa variabel independen dan  $L_a$  adalah nilai *likelihood* dengan variabel independen.

Daerah penolakan/Kriteria pengujian:

Jika nilai  $G_{hitung} \geq \chi_{\alpha,p}^2$  atau  $p - \text{value} \leq \alpha$  maka  $H_0$  ditolak dan  $H_1$  diterima

Jika nilai  $G_{hitung} < \chi_{\alpha,p}^2$  atau  $p - \text{value} > \alpha$  maka  $H_0$  diterima dan  $H_1$  ditolak

Adapun uji parsial adalah pengujian signifikansi parameter yang digunakan untuk mengetahui variabel independen ke- $a$  yang memiliki peranan atau hubungan yang lebih kuat terhadap variabel dependen dengan menggunakan uji *wald*.

Hipotesis yang digunakan, yaitu:

$$H_0: \beta_a = 0, \quad a = 1, 2, \dots, p$$

$$H_1: \beta_a \neq 0, \quad a = 1, 2, \dots, p$$

Statistik uji yang ditunjukkan oleh Persamaan (19).

$$Z_{hitung} = \frac{\hat{\beta}_a}{SE(\hat{\beta}_a)} \quad (19)$$

dengan  $\hat{\beta}_a$  adalah penduga estimator  $\beta_a$  dan  $SE(\hat{\beta}_a)$  adalah penduga standar error dari estimator  $\beta_a$ .

Daerah penolakan/Kriteria pengujian:

Jika nilai  $|Z_{hitung}| \geq \chi_{\alpha,1}^2$  atau  $p - value \leq \alpha$  maka  $H_0$  ditolak dan  $H_1$  diterima

Jika nilai  $|Z_{hitung}| < \chi_{\alpha,1}^2$  atau  $p - value > \alpha$  maka  $H_0$  diterima dan  $H_1$  ditolak

### 1.5.9 Evaluasi Model

Tingkat ketepatan klasifikasi yang dilakukan pada pemodelan regresi logistik dapat dievaluasi menggunakan *confusion matrix*. Dalam kasus biner, terdapat empat komponen yang digunakan pada *confusion matrix*, yaitu *true positif* (TP) menunjukkan hasil klasifikasi yang benar diidentifikasi sebagai positif, *true negatif* (TN) menunjukkan hasil klasifikasi yang benar diidentifikasi sebagai negatif, *false positif* (FP) menunjukkan hasil klasifikasi yang salah diidentifikasi sebagai positif, dan *false negatif* (FN) menunjukkan hasil klasifikasi yang salah diidentifikasi sebagai negatif (Jayidan dkk., 2024). Tabel 1 menunjukkan *confusion matrix* dalam kasus biner.

**Tabel 1.** *Confusion Matrix* Biner

Data sebenarnya	Hasil Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

*Confusion matrix* digunakan untuk mengukur kinerja model yaitu dengan menghitung nilai akurasi. Perhitungan nilai akurasi dilakukan dengan jumlah prediksi yang benar (TP dan TN) dibagi dengan banyaknya data. Akurasi dapat dikatakan sudah baik jika mencapai angka 80% (Bakti Kominfo, 2019). Persamaan (20) menunjukkan perhitungan akurasi (Jayidan dkk., 2024).

$$Akurasi = \frac{\text{Jumlah prediksi yang benar}}{\text{Banyaknya data}} = \frac{TP+TN}{TP+TN+FP+FN} \quad (20)$$

### 1.5.10 Wasting dan Stunting pada Balita

*Wasting* dan *stunting* merupakan masalah gizi yang memiliki risiko yang kematian yang tinggi. *Wasting* yang merujuk pada orang yang terlalu kurus untuk tinggi badannya, kondisi ini mempunyai risiko kematian yang lebih tinggi namun dapat diterapi. *Stunting* yang merujuk pada orang yang terlalu pendek untuk usianya,

kondisi ini mengakibatkan gangguan fisik dan kognitif yang tidak dapat diperbaiki sehingga menyertai hambatan pertumbuhan linier. Risiko kematian akan meningkat ketika balita yang mengalami *wasting* dan *stunting* secara bersamaan. Oleh karena itu, perlu tindakan untuk mengatasi masalah kekurangan gizi tersebut (Kementerian Kesehatan, 2021).

Pemerintah Indonesia telah berkomitmen untuk mengatasi masalah kekurangan gizi, yaitu dengan menetapkan target untuk menurunkan prevalensi *stunting* dan *wasting*. Target yang ditetapkan pemerintah Indonesia, yaitu mengakhiri segala bentuk masalah gizi termasuk mencapai target *wasting* yang telah disepakati secara internasional <5% pada tahun 2025 dan Tujuan Pembangunan Berkelanjutan (TPB) atau *Sustainable Development Goals* (SDGs) <3% pada tahun 2030, menurunkan prevalensi *stunting* menjadi 14% dan prevalensi *wasting* sebesar 7% yang tercantum dalam Rencana Pembangunan Jangka Menengah (RPJMN) 2020 – 2024, serta memberikan tata laksana bagi 90% anak gizi buruk di tahun 2024 sebagaimana tercantum dalam Peraturan Presiden No 72 Tahun 2021 (Kementerian Kesehatan, 2021). Pengklasifikasian balita yang bergizi buruk dapat dilakukan dengan pemantauan pertumbuhan dan perkembangan balita.

Pemantauan pertumbuhan dan perkembangan balita dapat dilakukan dengan penilaian status gizi, salah satunya dengan antropometri. Antropometri adalah metode yang digunakan untuk pengukuran tubuh manusia berupa ukuran, proporsi, dan komposisi tubuh. Berdasarkan pengukuran tersebut, dapat timbul beberapa masalah status gizi pada balita yaitu *underweight* (berat badan kurang) dan *overweight* (berat badan lebih) yang berdasarkan pada berat badan dan umur (BB/U), *stunting* (pendek) dan tinggi yang berdasarkan pada tinggi badan dan umur (TB/U), serta *wasting* (gizi kurang) dan *obese* (obesitas) yang berdasarkan pada berat badan dan tinggi badan (BB/TB). Berdasarkan pada Peraturan Menteri Kesehatan No 2 Tahun 2020 terdapat empat kategori status gizi yang berdasarkan pada tinggi badan dan umur (TB/U) pada balita yaitu sangat pendek (*severely stunted*), pendek (*stunted*), normal, dan tinggi. Tabel 2 menunjukkan kategori dan ambang menurut tinggi badan dan umur pada balita.

**Tabel 2.** Kategori dan Ambang Batas Menurut Tinggi Badan dan Umur

Indeks Pengukuran	Kategori	Ambang Batas
Status gizi yang berdasarkan pada tinggi badan dan umur pada balita (TB/U)	Sangat Pendek	$z - score < -3 SD$
	Pendek	$-3 SD \leq z - score < -2 SD$
	Normal	$-2 SD \leq z - score \leq 3 SD$
	Tinggi	$z - score > 3 SD$

Indeks TB/U menggambarkan pertumbuhan tinggi badan balita berdasarkan umurnya. Balita yang mengalami kesesuaian pertumbuhan tinggi badan berdasarkan umurnya tergolong pada kategori normal dengan nilai ambang batas pada  $-2 SD \leq z - score \leq 3 SD$ . Indeks ini juga dapat mengidentifikasi balita yang *stunting* (pendek atau sangat pendek) yaitu dengan nilai ambang batas pada  $z - score < -2 SD$ . Kondisi ini dicirikan dengan balita yang terlihat pendek jika dibandingkan dengan balita seusianya yang disebabkan oleh gizi kurang dalam waktu lama atau sering sakit. Adapun balita yang tergolong tinggi menurut umurnya juga dapat diidentifikasi dengan nilai ambang batas pada  $z - score > 3 SD$ . Kondisi ini dicirikan dengan balita

yang terlihat tinggi jika dibandingkan dengan balita seusianya yang disebabkan oleh gangguan endokrin yang jarang terjadi di Indonesia.

Berdasarkan pada Peraturan Menteri Kesehatan No 2 Tahun 2020 terdapat enam kategori status gizi yang berdasarkan pada berat badan dan tinggi badan (BB/TB) pada balita yaitu gizi buruk (*severely wasted*), gizi kurang (*wasted*), gizi baik, berisiko gizi lebih (*possible risk of overweight*), gizi lebih (*overweight*), dan obesitas (*obese*). Tabel 3 menunjukkan kategori dan ambang menurut berat badan dan tinggi badan pada balita.

**Tabel 3.** Kategori dan Ambang Batas Menurut Berat Badan dan Tinggi Badan

Indeks Pengukuran	Kategori	Ambang Batas
Status gizi yang berdasarkan pada berat badan dan tinggi badan pada balita (BB/TB)	Gizi buruk	$z - score < -3 SD$
	Gizi kurang	$-3 SD \leq z - score < -2 SD$
	Gizi baik	$-2 SD \leq z - score \leq 1 SD$
	Berisiko gizi lebih	$1 SD < z - score \leq 2 SD$
	Gizi lebih	$2 SD < z - score \leq 3 SD$
	Obesitas	$z - score > 3 SD$

Indeks BB/TB ini menggambarkan kesesuaian berat badan balita terhadap pertumbuhan tinggi badannya. Balita yang mengalami kesesuaian berat badan terhadap pertumbuhan tinggi badannya tergolong pada kategori gizi baik dengan nilai ambang batas pada  $-2 SD \leq z - score \leq 1 SD$ . Indeks ini juga dapat mengidentifikasi balita yang *obese* yaitu tergolong kategori berisiko gizi lebih, gizi lebih, atau obesitas dengan nilai ambang batas pada  $z - score > 3 SD$ . Kondisi ini dicirikan dengan balita yang terlihat gemuk. Selain itu, Indeks ini juga dapat digunakan untuk mengidentifikasi balita yang *wasting* yaitu tergolong kategori gizi kurang atau gizi buruk dengan nilai ambang batas pada  $z - score < -2 SD$ . Kondisi ini dicirikan dengan balita terlihat sangat kurus yang disebabkan oleh penyakit dan kekurangan asupan gizi yang baru saja terjadi maupun yang telah lama terjadi. Dari semua bentuk masalah gizi pada balita, *wasting* memiliki risiko kematian tertinggi terutama gizi buruk berisiko kematian 12 kali lebih tinggi dibandingkan dengan balita gizi baik (Unicef, 2023).

## BAB II METODE PENELITIAN

### 2.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data status *wasting* balita *stunting* di Kabupaten Gowa pada bulan Agustus 2023. Data tersebut merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kabupaten Gowa. Adapun surat pengajuan data dapat dilihat pada Lampiran 1. Data status *wasting* balita yang digunakan terdiri atas 2094 amatan dari 18 kecamatan di Kabupaten Gowa yaitu Bajeng, Bajeng Barat, Barombong, Biringbulu, Bontolempangan, Bontomarannu, Bontonompo, Bontonompo Selatan, Bungaya, Manuju, Pallangga, Parangloe, Parigi, Pattallassang, Somba Opu, Tinggimoncong, Tombolo Pao, dan Tompobulu. Data yang digunakan pada penelitian ini dapat dilihat pada Lampiran 2.

### 2.2 Deskripsi Variabel

Variabel yang digunakan pada penelitian ini terdiri atas satu variabel dependen dan lima variabel independen. Variabel dependen yang digunakan merupakan status *wasting* balita (BB/TB) yang berupa variabel biner, yaitu normal (gizi baik) yang diberi kode "0" dan *wasting* (gizi kurang dan gizi buruk) yang diberi kode "1". Adapun lima variabel independen, yaitu berat badan lahir ( $x_1$ ), tinggi badan lahir ( $x_2$ ), usia ( $x_3$ ), berat badan saat pengukuran ( $x_4$ ), dan tinggi badan saat pengukuran ( $x_5$ ). Variabel yang digunakan pada penelitian ini seperti pada Tabel 4.

**Tabel 4.** Variabel Penelitian

Variabel Penelitian	Simbol	Nama Variabel	Definisi Operasional	Skala Pengukuran
Variabel Dependen	$y$	Status <i>Wasting</i> Balita	Status <i>wasting</i> berdasarkan indeks (BB/TB) dengan kategori normal (0) ketika hasil pengukuran $-2 SD \leq z - score \leq 1 SD$ dan kategori <i>wasting</i> (1) ketika hasil pengukuran $z - score < -2 SD$	Nominal
Variabel Independen	$x_1$	Berat Badan Lahir	Berat balita saat baru lahir dengan satuan kilogram (kg)	Rasio
	$x_2$	Tinggi Badan Lahir	Panjang tubuh saat baru lahir yang diukur dengan satuan sentimeter (cm)	Rasio
	$x_3$	Usia	Jarak waktu saat baru lahir hingga saat pengukuran dalam satuan bulan	Rasio
	$x_4$	Berat Badan Saat Pengukuran	Berat balita saat pengambilan data dengan satuan kilogram (kg)	Rasio

---

$x_5$	Tinggi Badan Saat Pengukuran	Panjang tubuh pengambilan data dengan satuan sentimeter (cm)	saat dengan	Rasio
-------	------------------------------------	--	----------------	-------

---

### 2.3 Metode Analisis

Tahapan metode analisis data yang dilakukan dalam penelitian ini terdiri atas 2 bagian, yaitu:

1. Mengestimasi parameter model regresi logistik biner AKU dengan estimator *spline truncated*.
2. Melakukan pemodelan dengan menggunakan regresi logistik biner komponen utama pada data multikolinieritas dengan estimator *spline truncated* menggunakan data status gizi balita *stunting* di Kabupaten Gowa pada bulan Agustus 2023. Adapun tahapan pemodelan ini sebagai berikut :
  - a. Membuat statistik deskriptif dengan tujuan untuk menggambarkan data.
  - b. Membentuk matriks korelasi dengan menggunakan korelasi *pearson* pada Persamaan (1) untuk menguji adanya multikolinieritas pada variabel independen
  - c. Menghitung nilai eigen sesuai pada Persamaan (2) dan vektor eigen sesuai pada Persamaan (3) dari matriks korelasi yang telah diperoleh pada poin a.
  - d. Menerapkan AKU dan membentuk komponen-komponen ( $w_1, w_2, \dots, w_m$ ) untuk mengatasi multikolinieritas pada variabel independen.
  - e. Memilih komponen utama optimal ( $w_j$ ) berdasarkan pada *scree plot*.
  - f. Memodelkan komponen utama optimal ( $w_j$ ) dengan estimator *spline truncated* orde 1 *knot* pada orde 1 (linier), 2 (kuadratik), dan 3 (kubik).
  - g. Memilih model terbaik berdasarkan titik *knot* optimal dengan menggunakan nilai *Generalized Cross Validation* (GCV) paling minimum yang diperoleh berdasarkan pada Persamaan (17).
  - h. Melakukan pemodelan pada 2 dan 3 *knot* dengan mengulang tahapan poin e dan f.
  - i. Melakukan uji signifikansi secara simultan dengan menggunakan Persamaan (18) dan parsial dengan menggunakan Persamaan (19).
  - j. Mengukur ketepatan klasifikasi model dengan menggunakan akurasi sesuai pada Persamaan (20).
  - k. Menginterpretasi model dan menarik kesimpulan.