BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi adalah metode statistik yang digunakan untuk menguji dan menjelaskan hubungan antara variabel berkaitan, yaitu variabel independen dan variabel dependen. Pada penerapannya, analisis regresi sering digunakan untuk variabel dependen kuantitatif (kontinu), namun pada kasus tertentu dapat dijumpai variabel dependen kualitatif (kategorik). Metode yang mampu menyelesaikan data dengan variabel dependen kategorik adalah metode regresi logistik (Fitri dkk., 2024). Metode regresi logistik merupakan metode yang digunakan untuk menentukan hubungan antar variabel dependen kualitatif dengan variabel independent yang bersifat kontinu maupun kategorik. Berdasarkan variabel dependen, regresi logistik terbagi menjadi, regresi logistik biner untuk variabel dependen dua kategori, regresi logistik multinomial untuk variabel dependen lebih dari dua kategori tidak terurut, dan regresi logistik ordinal untuk variabel dependen lebih dari dua kategori terurut (Annas dkk., 2022). Regresi logistik yang umum digunakan yakni regresi logistik biner, terutama pada bidang kesehatan, sosial, dan ekonomi untuk memprediksi kemungkinan dengan dua kategori seperti, ya atau tidak (Gadrich dkk., 2022).

Regresi logistik biner merupakan metode statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen bersifat biner dengan beberapa variabel independen (Ronny Susetyoko dkk., 2022). Metode ini efektif untuk memprediksi probabilitas suatu kejadian berdasarkan variabel prediktor kuantitatif maupun kualitatif. Bentuk penelitian mengenai model regresi logistik biner yaitu McEligot dkk (2020) melakukan penelitian dengan menekankan pentingnya regresi logistik biner dalam memahami faktor risiko terkait diet dan kesehatan, serta memberikan wawasan tentang pengaruh faktor diet terhadap diagnosis kanker payudara. Meskipun demikian, metode ini terbatas dalam memenuhi asumsi hubungan parametrik antara variabel independen dan dependen. Hal ini menjadi masalah ketika pola data tidak mengikuti bentuk parametrik, sehingga model sering menghasilkan hasil yang kurang akurat dan bias dalam estimasinya. Oleh karena itu, pendekatan regresi nonparametrik dapat menjadi solusi karena memberikan fleksibilitas lebih dalam menangkap variasi data, sehingga model dapat beradaptasi dengan pola data yang kompleks dan dinamis (Dani dkk., 2021).

Pendekatan regresi nonparametrik digunakan ketika pola hubungan antara variabel tidak diketahui, sehingga kurva regresi diasumsikan bersifat mulus. Metode ini menawarkan fleksibilitas tinggi karena memungkinkan data untuk menentukan sendiri bentuk estimasi kurva regresinya tanpa pengaruh subjektivitas dari peneliti. Dalam pendekatan ini, tidak ada spesifikasi pola fungsi regresi tertentu, sehingga dapat digunakan berbagai estimator untuk memperkirakan fungsi regresi dalam model. Beberapa estimator yang umum digunakan termasuk spline, kernel, dan linear lokal. Estimator spline merupakan estimator dalam regresi nonparametrik yang mampu mencari sendiri estimasi data berdasarkan pola pergerakan datanya,

sehingga disebut sebagai model yang paling fleksibel (Dani & Adrianingsih, 2021). Penggunaan estimator dalam analisis regresi nonparametrik juga meningkatkan kualitas hasil dan interpretasi model.

Salah satu estimator yang sering digunakan dalam pendekatan regresi nonparametrik adalah estimator *spline truncated* (Nurhuda dkk., 2022) yang juga telah dikembangkan untuk diterapkan dalam regresi logistik. Regresi logistik *spline truncated* merupakan pendekatan non-parametrik yang digunakan untuk mengatasi keterbatasan model regresi logistik, khususnya dalam menangkap pola non-linear yang signifikan dalam data. Metode ini memanfaatkan fungsi *spline truncated* yaitu fungsi polinomial tersegmen, sehingga model lebih fleksibel dalam menyesuaikan karakter lokal dari data. Penerapan regresi logistik ordinal spline menggunakan tiga titik knot pada data gizi balita di Gowa ditemukan akurasi 92,25% (Arifin dkk., 2023). Selain itu, terdapat penerapan regresi logistik spline binary untuk menganalisis status gizi anak di Kabupaten Barru, Sulawesi Selatan, dengan akurasi klasifikasi sebesar 87,50% (Islamiyati dkk., 2023). Dengan menggunakan *spline truncated*, analisis dapat menangkap variasi kompleks dan dinamis dalam hubungan antara variabel independen dan dependen, sehingga meningkatkan akurasi prediksi dan mengurangi *Mean Squared Error* (MSE) secara signifikan (Nur Fadhilah, 2016). Namun, model ini memiliki tantangan terkait multikolinearitas.

Multikolinearitas adalah kondisi ketika dua atau lebih variabel independen memiliki hubungan yang sangat kuat, sehingga sulit dalam memperkirakan pengaruh masing-masing variabel terhadap variabel dependen. Ketika multikolinearitas tidak ditangani, hal ini dapat menyebabkan estimasi koefisien regresi menjadi tidak stabil dan sulit diinterpretasikan. Salah satu metode yang efektif untuk mengatasi masalah tersebut adalah *Least Absolute Shrinkage and Selection Operator* (LASSO). Metode ini mampu menyusutkan koefisien variabel yang berkorelasi tinggi menjadi nol, sehingga memungkinkan pemilihan variabel yang lebih baik dan meningkatkan stabilitas model (Robbani dkk., 2019). Penelitian dengan LASSO menunjukkan nilai *Mean Squared Error Prediction* lebih baik dalam menangani multikolinearitas dibandingkan PLS dan OLS (Dewi, 2010) dan meningkatkan stabilitas model (Robbani dkk., 2019).

Penelitian oleh Padhillah & Herrhyanto (2024) menggunakan regresi logistik biner dan LASSO pada data indeks pembangunan manusia di Jawa Barat. Hasil yang diperoleh nilai λ optimal sebesar 0,0167 dari validasi silang. Temuan ini menunjukkan efektivitas pendekatan regresi fleksibel dalam menangani data dengan multikolinearitas dan pola non-linear. Oleh karena itu, penelitian selanjutnya akan difokuskan pada status gizi balita stunting dengan pengembangan metode yang ada.

Gizi adalah makanan penting untuk perkembangan, pertumbuhan, dan kesehatan tubuh. Status gizi yang baik sangat penting untuk kesehatan optimal, sementara kekurangan gizi dapat menyebabkan masalah perkembangan, energi rendah, penurunan kekebalan, dan gangguan fungsi otak, terutama pada balita usia 0-59 bulan (Handayani & Charis Fauzan, 2024). Pada tahun 2022, tingkat stunting di Kabupaten Gowa mencapai 33%, namun berhasil turun menjadi 21,1% pada tahun 2023, menjadikannya daerah dengan penurunan stunting tertinggi kedua di Sulawesi

Selatan setelah Luwu Utara. Meski beberapa daerah sudah mencapai prevalensi di bawah 20%, target nasional 14% pada tahun 2024 belum tercapai. Oleh karena itu, analisis faktor-faktor yang mempengaruhi status gizi balita diperlukan untuk menekan angka stunting di tahun-tahun berikutnya. Pendekatan ini akan memanfaatkan temuan dari penelitian sebelumnya untuk mengidentifikasi faktor risiko dan pola pertumbuhan anak, meningkatkan efektivitas analisis data, dan memungkinkan intervensi yang lebih efisien. Penelitian ini menerapkan regresi logistik biner menggunakan penalti LASSO dengan estimator spline truncated pada data status gizi balita. Metode ini diharapkan memberikan wawasan mendalam tentang faktor-faktor yang memengaruhi status gizi balita dan mendukung intervensi gizi yang lebih tepat sasaran, serta dapat meningkatkan stabilitas model dan memperbaiki interpretabilitas hasil analisis, menjadikannya efektif untuk data berdimensi tinggi dan kompleks.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Pemodelan *spline truncated* dalam regresi logistik biner hanya dibatasi pada orde satu (linear).
2. Metode penaksiran parameter yang digunakan adalah *Maximum Likelihood Estimation* (MLE).
3. Pemilihan titik knot optimal yang dilakukan hanya sampai pada tiga titik knot yang dipilih berdasarkan metode *Generalized Cross-Validation* (GCV).

1.3 Tujuan Penelitian dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengestimasi parameter model regresi logistik biner menggunakan penalti *Least Absolute Shrinkage and Selection Operator* dengan estimator *spline truncated*.
2. Memodelkan hubungan antara Status Gizi Balita dengan faktor-faktor yang memengaruhinya di Kabupaten Gowa pada tahun 2023, dengan pendekatan regresi logistik biner *Least Absolute Shrinkage and Selection Operator* dengan estimator *spline truncated*.

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Sebagai sumber pengetahuan mengenai tahapan estimasi parameter model regresi logistik biner estimator *spline truncated* dengan penalti *Least Absolute Shrinkage and Selection Operator*.
2. Sebagai sumber pengetahuan dan informasi mengenai model status gizi balita yang dihasilkan melalui regresi logistik biner estimator *spline truncated* dengan penalti *Least Absolute Shrinkage and Selection Operator*.

1.5 Teori

1.5.1 Regresi Nonparametrik *Spline Truncated*

Regresi nonparametrik adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan independen tanpa harus mengasumsikan bentuk tertentu dari hubungan tersebut. Menurut Wasserman (2006), metode ini

memberikan fleksibilitas dalam menggambarkan hubungan antar variabel. Berbeda dengan regresi parametrik yang mengharuskan penetapan bentuk model tertentu, regresi nonparametrik tidak terikat pada asumsi matematis yang ketat. Hal ini memungkinkan regresi nonparametrik untuk menangkap pola yang lebih kompleks, seperti hubungan non-linear yang sering ditemukan dalam data keuangan atau lingkungan (Copeland, 1997). Secara matematis, model regresi nonparametrik dapat dinyatakan dengan persamaan berikut :

$$y_i = f(x_i) + \epsilon_i \quad (1)$$

Fungsi $f(x_i)$ dalam Persamaan (1) dinyatakan sebagai fungsi *spline truncated* yang berbentuk polinomial dengan sifat tersegmen, $i = 1,2, \dots, n$

Spline memiliki fleksibilitas yang lebih tinggi dibandingkan polinomial biasa. Spline dapat menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi data, atau dengan kata lain, spline dapat menghasilkan fungsi regresi yang sesuai dengan data (Widyastuti dkk., 2021). Fungsi $f(x_i)$ adalah fungsi spline berorde m dengan titik knot $k_1, k_2, \dots, k_r$ yang diberikan oleh persamaan:

$$f(x_i) = \sum_{g=0}^{m} \beta_g x_i^g + \sum_{h=1}^{r} \beta_{m+h}(x_i - k_h)_+^m \quad (2)$$

dengan $\beta_0, \beta_1, \dots, \beta_q, \beta_{m+1}, \dots, \beta_{m+h}$ adalah parameter regresi, k_h merupakan titik knot ke-h, $(h = 1,2, \dots, r)$. Apabila Persamaan (2) disubtitusikan ke dalam Persamaan (1) maka diperoleh model regresi nonparametric spline truncated sebagai berikut:

$$y_i = \sum_{g=0}^{m} \beta_g x_i^g + \sum_{h=1}^{r} \beta_{m+h}(x_i - k_h)_+^m + \varepsilon_i \quad (3)$$

Jika data memuat variabel independent sebanyak p , maka model regresi nonparametrik *spline truncated* dapat dinyatakan dalam bentuk sebagai berikut:

$$y_i = f(x_{1i}) + f(x_{2i}) + \cdots + f(x_{pi}) + \varepsilon_i$$

$$y_i = \sum_{j=1}^{p} f(x_{ji}) + \varepsilon_i$$

Dengan $f(x_i) = \sum_{g=0}^{m} \beta_{jg} x_{ji}^g + \sum_{h=1}^{r} \beta_{j(m+h)} (x_{ji} - k_{jh})_+^m$

Keterangan:

y_i : Variabel dependen,

x_i : Variabel independen,

β_{jg} : Parameter *polynomial* pada variabel independen ke-j dan orde ke-g,

k_{jh} : Nilai titik knot pada prediktor ke-j dan titik knot ke-h,

r : Banyak titik knot,

m : Orde *polynomial spline Truncated*,

$\beta_{j(m+h)}$: Parameter *Truncated* pada variabel ke-j dan titik knot ke-$m + h$,

ε_i : Nilai titik knot pada prediktor ke-j dan titik knot ke-h

Persamaan (3) sebagai model regresi nonparametrik *spline truncated* juga dapat ditulis bentuk matriks sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m & (x_1-k_1)_+^m & \cdots & (x_1-k_r)_+^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m & (x_1-k_1)_+^m & \cdots & (x_2-k_r)_+^m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m & (x_1-k_1)_+^m & \cdots & (x_n-k_r)_+^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \\ \beta_{(m+1)} \\ \vdots \\ \beta_{m+r} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Pemilihan titik knot yang meminimalkan nilai *Generalized Cross Validation* (GCV) adalah kunci untuk menentukan solusi optimal dari model ini (Ratnasari dkk., 2021). Dengan menggunakan pendekatan *spline truncated*, penyesuaian lokal terhadap data dapat dilakukan dengan lebih baik, sehingga metode ini efektif dalam menangani pola data yang kompleks dan dinamis.

1.5.2 Regresi Logistik Biner

Regresi logistik biner adalah teknik analisis statistik yang digunakan untuk memeriksa hubungan antara variabel independen dan variabel dependen yang bersifat dikotomik (memiliki dua kategori, yaitu 0 dan 1) (Tampil dkk., 2017). Variabel dependen ini diasumsikan mengikuti distribusi Bernoulli, yang berarti setiap pengamatan merupakan hasil dari variabel acak dengan dua kemungkinan nilai, yaitu $y = 1$ yang menyatakan 'sukses' dan y = 0 yang menyatakan 'gagal'. Jika y_i merupakan variabel dependen untuk pengamatan ke-i dan $x_{i1}, x_{i2}, \dots, x_{ip}$ adalah variabel independent, maka probabilitas terjadinya keberhasilan ($y = 1$) untuk pengamatan ke-i dinyatakan sebagai :

$$P\left(y_i = 1 \middle| x_{i1}, x_{i2}, \dots, x_{ip}\right) = \pi(x_i) \tag{4}$$

Adapun fungsi probabilitas untuk distribusi bernoulli dinyatakan sebagai :

$$P(y_i|x_i) = \pi(x_i)^{y_i}\left(1 - \pi(x_i)\right)^{1-y_i} \tag{5}$$

Menurut Hosmer dan Lemeshow (2000) model regresi logistik biner dengan variabel prediktor memiliki bentuk persamaan :

$$\pi(x_i) = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}}$$

Fungsi $\pi(x_i)$ dalam regresi logistik bersifat non-linear, sehingga perlu diubah ke dalam bentuk logit untuk menjadikannya fungsi linear. Transformasi ini memudahkan analisis hubungan antara variabel respon dan variabel prediktor. Transformasi logit dari $\pi(x)$ adalah:

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$$

Maka model regresi logistik dapat dituliskan dalam bentuk persamaan :

$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}} \tag{6}$$

Regresi logistik biner menggunakan metode *Maximum Likelihood Estimation* (MLE) untuk mengestimasi parameter model β dengan memaksimalkan fungsi *likelihood*. Fungsi *likelihood* menunjukkan probabilitas data yang diamati

berdasarkan parameter model yang diusulkan. MLE digunakan untuk menemukan nilai parameter yang paling mungkin menghasilkan data yang diamati (Liang & Du, 2012). Metode ini berguna dalam penelitian yang melibatkan variabel respons dikotomik, seperti 'sukses' atau 'gagal', dan sering digunakan dalam pemodelan risiko penyakit serta pengambilan keputusan berbasis data biner (Utami dkk., 2024). Untuk data independen, fungsi *likelihood* dapat dirumuskan pada Persamaan (7):

$$L(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}\left(1 - \pi(x_i)\right)^{1-y_i} \tag{7}$$

Untuk mempermudah proses estimasi, maka fungsi *likelihood* Persamaan (7) dimaksimumkan, sehingga menghasilkan fungsi *log-likelihood*:

$$\ln[L(\beta)] = \sum_{i=1}^{n} [y_i \ln \pi(x_i) + (1 - y_i) \ln (1 - \pi(x_i))] \tag{8}$$

dengan mensubtitusikan $\pi(x_i)$ pada Persamaan (6) ke Persamaan (8) maka:

$$\ln L(\beta) = \sum_{i=1}^{n} [y_i \ln \left(\frac{e^{g(x_i)}}{1 + e^{g(x_i)}}\right) + (1 - y_i) \ln \left(1 - \frac{e^{g(x_i)}}{1 + e^{g(x_i)}}\right)]$$

Sehingga, didapatkan :

$$\ln L(\beta) = \sum_{i=1}^{n} \left[y_i g(x_i) - \ln\left(1 + e^{g(x_i)}\right)\right] \tag{9}$$

Log-likelihood lebih stabil secara numerik dan lebih mudah untuk didiferensiasi dibandingkan dengan fungsi *likelihood*. *Log-likelihood* juga menunjukkan sejauh mana parameter β dapat menjelaskan data yang diamati. Untuk menemukan nilai parameter β yang memaksimalkan *log-likelihood*, bentuk ln pada persamaan (9) diturunkan terhadap β_j dan hasil dari turunan disamakan dengan nol, sehingga dinyatakan persamaan :

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij}(y_i - \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}}) \tag{10}$$

Persamaan (10) merupakan persamaan yang tidak linear pada parameter β sehingga digunakan metode *fisher scoring* untuk mendapatkan nilai β. Untuk memudahkan perhitungan, digunakan metode iterasi dengan program computer yang akan berhenti ketika memperoleh nilai konvergen pada iterasi yang merupakan estimator yang memaksimumkan fungsi *likelihood*.

1.5.3 Multikolinearitas

Multikolinearitas terjadi ketika terdapat hubungan linear yang sangat kuat antara kolom-kolom dalam matriks $\boldsymbol{X}$. Jika hubungan linear tersebut sempurna, nilai determinan dari $\boldsymbol{X^T X}$ akan menjadi nol, yang dikenal sebagai multikolinearitas sempurna. Keberadaan multikolinearitas dapat mempengaruhi varians koefisien regresi, yang pada gilirannya mengurangi keandalan hasil regresi. Ketika variabel-variabel independen memiliki korelasi tinggi, interpretasi pengaruh masing-masing variabel terhadap variabel dependen menjadi kabur (Shrestha, 2020).

Metode yang paling sederhana dan umum untuk mendeteksi multikolinearitas adalah melalui analisis matriks korelasi. Metode ini menilai tingkat

hubungan antara variabel independen dengan menggunakan koefisien korelasi Pearson, yang nilainya berkisar antara -1 hingga 1. Koefisien korelasi Pearson r_{ij} dapat dihitung menggunakan Persamaan (11):

$$r_{ij} = \frac{n\left(\sum x_i x_j\right) - \left(\sum x_i\right)\left(\sum x_j\right)}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]\left[n\sum x_j^2 - \left(\sum x_j\right)^2\right]}} \tag{11}$$

dengan n adalah jumlah observasi, x_i adalah variabel independen ke-i, dan x_j adalah variabel independen ke-j. Setelah semua koefisien korelasi dihitung, hasilnya dapat disusun dalam bentuk matriks korelasi $\boldsymbol{R}$, yang merepresentasikan hubungan antar variabel independen dalam model regresi. Matriks korelasi dirumuskan sebagai berikut:

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

Matriks $\boldsymbol{R}$ memiliki sifat bahwa elemen diagonal utamanya r_{ii} selalu bernilai 1, karena setiap variabel memiliki korelasi sempurna dengan dirinya sendiri. Dalam praktiknya, nilai koefisien korelasi r_{ij} yang melebihi 0,8 atau kurang dari -0,8 dianggap menunjukkan adanya korelasi yang sangat tinggi antara variabel independen x_i dan x_j. Korelasi tinggi ini dapat menjadi indikasi adanya masalah multikolinearitas, yang dapat mengurangi akurasi model regresi (Kim, 2019).

1.5.4 *Least Absolute Shrinkage and Selection Operator*

Least Absolute Shrinkage and Selection Operator (LASSO) merupakan metode pengembangan regresi yang mampu menyelesaikan model regresi yang terdapat multikolinearitas. LASSO merupakan metode yang digunakan untuk *shrinkage* yaitu menyusutkan koefisien taksiran mendekati angka nol dan *selection operator* yaitu menyeleksi variabel independen sehingga menghasilkan model dengan variabel terbaik. Selain itu, regresi LASSO juga digunakan untuk data yang kontinu dan memerlukan variabel independent yang berdistribusi normal baku.

Penaksir koefisien pada regresi LASSO ($\widehat{\beta}_J^{\,LASSO}$) diperoleh dengan cara meminimumkan Persamaan (12) berikut:

$$\widehat{\beta}_J^{\,LASSO} = \sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{k}\left(x_{ij}\beta_j\right)^2\right) \tag{12}$$

dengan fungsi kendala $\sum_{j=1}^{p}\left|\beta_j\right| \le t$. Kemudian dapat ditulis dengan persamaan Lagrange berikut:

$$\widehat{\beta}_J^{\,LASSO} = \arg\min\left(\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right) \tag{13}$$

dengan Nilai t merupakan suatu besaran yang mengontrol besarnya penyusutan pada koefisien regresi LASSO. Dan λ disebut sebagai parameter tuning yang

berkorespondensi satu-satu dengan t. Artinya, untuk setiap nilai $t \geq 0$ yang menghasilkan solusi $\widehat{\beta_J}^{LASSO}$, terdapat $\lambda \geq 0$ sedemikian sehingga menghasilkan solusi $\widehat{\beta_J}^{LASSO}$ juga. Solusi regresi LASSO tidak memiliki solusi eksplisit karena pada fungsi kendala regresi LASSO berbentuk fungsi mutlak yang tidak dapat diturunkan pada titik beloknya (Robbani dkk., 2019).

Misalkan diketahui $\widehat{\beta_j}$ merupakan penaksir *Ordinary Least Square* (OLS), dengan nilai t_0 didefinisikan sebagai $t_0 = \sum_{j=1}^{p} |\widehat{\beta_J}|$, maka

1. Jika nilai $t > t_0$, maka koefisien OLS akan menyusut ke arah nol, dan memungkinkan untuk menjadi tepat nol.
2. Jika nilai $t > t_0$, maka koefisien regresi LASSO memberikan hasil yang sama dengan koefisien OLS.

Algoritma Least Angle Regression (LAR) merupakan sebuah algoritma untuk menghasilkan model linier yang ditemukan tahun 2002. Algoritma LAR membutuhkan beberapa langkah untuk mendapatkan koefisien taksiran OLS. Dengan memodifikasi algoritma LAR dapat memberikan koefisien taksiran metode LASSO. Algoritma yang dimodifikasi ini memiliki langkah yang lebih efisien dibanding metode LASSO itu sendiri. Algoritma LAR yang dimodifikasi ini sering disebut juga sebagai algoritma LARS. Algoritma LARS memberikan jalan yang efisien dalam menyelesaikan regresi LASSO. Algoritma ini dimulai dengan semua koefisien β sama dengan nol. Algoritma LAR asli adalah sebagai berikut:

1. Bakukan variabel independen sehingga memiliki nilai tengah nol dan varians satu. Mulai dengan residual $r = y - \bar{y}, \beta 1, \beta 2, \ldots, \beta p = 0$.
2. Cari variabel independen x_j yang paling berkorelasi dengan r.
3. Ubah nilai β_j dari 0 bergerak menuju koefisien kuadrat terkecil (x_j, r), sampai kompetitor lain x_k memiliki korelasi sebesar korelasi x_j dengan sisaaan sekarang.
4. Ubah nilai β_j dan β_k bergerak dalam arah yang didefinisikan oleh koefisien kuadrat terkecil bersama dari sisaan sekarang dalam (x_j, x_k) sampai kompetitor x_l lain memiliki korelasi dengan sisaan sekarang dengan besaran yang sama.
5. Teruskan cara ini sampai semua pp variabel bebas telah masuk. Setelah min$(N-1, p)$ langkah, solusi model untuk OLS diperoleh.

LAR selalu mengambil p langkah untuk mendapatkan penaksir OLS secara penuh, sedangkan modifikasi LAR untuk metode LASSO dapat memiliki lebih dari p langkah untuk mendapatkannya. Algoritma LASSO dengan memodifikasi LAR adalah suatu cara yang efisien dalam komputasi solusi masalah LASSO khususnya ketika $p > N$. Pada hasil algoritma LAR, akan muncul $plot$ pergerakan variabel-variabel independen dengan parameter tuning bentuk standar (s). Menurut nilai parameter tuning s dapat diperoleh dengan Persamaan (14):

$$s = \frac{\mathrm{t}}{\sum_{\mathrm{j}=1}^{\mathrm{p}} \left|\widehat{\beta}_{\mathrm{j}}^{OLS}\right|} \tag{14}$$

Jika nilai $s = 1$, maka solusi regresi LASSO akan sama dengan solusi OLS. Nilai s yang optimal dalam penelitian ini akan diperoleh melalui validasi silang lipat-10.

1.5.5 Pemilihan Titik Knot Optimal

Titik knot merupakan perpaduan Bersama yang menunjukkan perubahan perilaku pada data. Model regresi spline terbaik tergantung pada titik knot optimal. Metode yang sering digunakan untuk mencari titik knot optimal adalah *Generalized Cross Validation* (GCV). Nilai GCV yang memberikan titik knot optimal adalah nilai GCV minimum (Davala dkk., 2024). Rumus GCV dapat dinyatakan pada Persamaan (15) berikut:

$$GCV(k) = \frac{MSE(k)}{(n^{-1}trace[\boldsymbol{I} - \boldsymbol{A}(k)])^2} \tag{15}$$

dengan $MSE(k) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2, k = k_1, k_2, \dots, k_r$ sebagai titik knot, $\boldsymbol{I}$ sebagai matriks identitas, $\boldsymbol{A}[k] = \boldsymbol{X}[k](\boldsymbol{X}'[k]\boldsymbol{X}[k])^{-1}\boldsymbol{X}'[k]$ dan $k = [k_1, k_2, \dots, k_r]$,dan n sebagai jumlah pengamatan.

1.5.6 Ketepatan Klasifikasi Model

Klasifikasi merupakan penilaian objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Sistem klasifikasi diharapkan mampu melakukan klasifikasi dengan benar pada set data, namun tidak dipungkiri bahwa kesalahan akan terjadi dalam proses pengklasifikasian tersebut sehingga perlu dilakukan dengan *confusion matrix* (Riehl dkk., 2023). Matriks konfusi merupakan table pencatat hasil kerja klasifikasi. *Confusion matrix* untuk data biner dituliskan pada Tabel 1.

Tabel 1. *Confusion Matrix*

Aktual	**Prediksi**	
	Positive	***Negative***
Positive	*True Positive*	*False Negative*
Negative	*False Positive*	*True Negative*

Nilai akurasi menunjukkan tingkat keakuratan model secara keseluruhan, semakin tinggi nilai akurasi semakin tinggi keakuratan suatu model. Berikut merupakan perhitungan nilai akurasi menggunakan Persamaan (16):

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

1.5.7 Status Gizi Balita

Status gizi balita adalah indikator penting yang mencerminkan kesehatan dan perkembangan anak. Penilaian status gizi biasanya dilakukan melalui metode antropometri, yang melibatkan pengukuran berat badan, tinggi badan, dan lingkar

kepala. Status gizi balita dapat dikategorikan ke dalam dua kelompok yaitu, normal dan abnormal. Anak-anak dengan status gizi sehat memiliki *z-score* antara -2 hingga +1 SD, sementara kondisi abnormal meliputi gizi buruk, gizi kurang, berisiko gizi lebih, gizi lebih, dan obesitas (Ferreira, 2020).

Pada Kabupaten Gowa, status gizi balita menghadapi tantangan besar. Pada tahun 2017, sekitar satu dari lima balita mengalami gizi kurang, dengan prevalensi mencapai 22,7%. Selain itu, prevalensi stunting mencapai 33% pada tahun 2022. Stunting adalah kondisi di mana anak mengalami pertumbuhan yang terhambat akibat kekurangan gizi kronis. Namun, berkat berbagai intervensi gizi yang dilakukan oleh pemerintah dan berbagai pihak terkait, angka stunting berhasil turun menjadi 21,1% pada tahun 2023. Penurunan ini menunjukkan bahwa upaya yang dilakukan mulai membuahkan hasil, meskipun masih banyak pekerjaan yang harus dilakukan untuk mencapai status gizi yang optimal bagi semua balita.

Kekurangan nutrisi pada masa balita dapat menyebabkan berbagai masalah jangka panjang. Anak-anak yang mengalami stunting cenderung memiliki kemampuan belajar yang lebih rendah dan risiko lebih tinggi terhadap penyakit kronis di masa dewasa. Oleh karena itu, intervensi nutrisi yang tepat selama periode emas ini sangat penting untuk meningkatkan kualitas hidup anak. Kolaborasi seluruh pihak dalam meningkatkan status gizi balita di Kabupaten Gowa sangatlah penting. Pemerintah, Lembaga Kesehatan, dan Masyarakat perlu bekerjasama untuk memastikan bahwa setiap anak mendapatkan nutrisi yang baik sebagai Upaya mengatasi tantangan gizi untuk generasi yang akan datang.

BAB II
METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Dinas Kesehatan Kabupaten Gowa. Data ini merupakan data status gizi balita Kabupaten Gowa tahun 2023 dengan 2091 data yang terlampir pada Lampiran 1.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri atas variabel dependen Status Gizi Balita dan lima variabel independen yang diduga memengaruhi Status Gizi Balita di Kabupaten Gowa. Variabel yang digunakan tercantum pada Tabel 2.

Tabel 2. Variabel Penelitian

Variabel	Keterangan	Kategori	Satuan
y	Status Gizi Balita	0 = Normal & 1 = Tidak Normal	-
x_1	Berat Badan	Rasio	Kilogram
x_2	Tinggi Badan	Rasio	Sentimeter
x_3	Berat Badan Lahir	Rasio	Kilogram
x_4	Tinggi Badan Lahir	Rasio	Sentimeter
x_5	Usia	Rasio	Bulan

2.3 Metode Analisis

Langkah-langkah analisis data yang dilakukan dalam penelitian ini sebagai berikut:

1. Mengestimasi parameter model regresi logistik biner dengan estimator *spline truncated* LASSO melalui tahapan
   a. Mengubah variabel independent kategorik ke dalam bentuk numerik kontinu X^* melalui kuantifikasi optimal (*optimal scalling*)
   b. Membentuk matrik korelasi untuk memeriksa adanya multikolinearitas antar variabel independent.
   c. Menyatakan variabel independent X^* ke dalam model regresi logistik biner *spline truncated* LASSO linear sebagai berikut:

$$g(X_i^*) = logit[\pi(X_i^*)] = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji}^*$$

   d. Mengestimasi parameter β_0, β_j menggunakan penalti L1 (LASSO) untuk menyaring variabel signifikan dan menekan multikolinearitas.
2. Memodelkan Status Gizi Balita menggunakan Regresi Logistik Biner *Spline Truncated* LASSO:
   a. Melakukan analisis deskriptif untuk variabel dependen dan Independen.
   b. Menguji multikolinearitas dengan matriks korelasi antar variabel independent dan VIF.
   c. Melakukan pemodelan regresi logistik biner *spline truncated* LASSO linear.
   d. Memilih model terbaik berdasarkan nilai *Generalized Cross-Validation* (GCV)
   e. Menguji signifikansi parameter menggunakan metode *likelihood ratio*.

f. Mengevaluasi performa model dengan akurasi, sensitivitas, dan spesifisitas.
g. Menarik Kesimpulan berdasarkan hasil analisis dan interpretasi efektivitas model.