

BAB I PENDAHULUAN

1.1 Latar Belakang

Pneumonia atau radang paru adalah infeksi saluran pernapasan akut yang menyebabkan kantung udara paru (sel alveoli) terisi oleh nanah atau cairan. Penyakit ini terjadi ketika kuman, seperti bakteri, virus atau jamur masuk ke dalam paru-paru. Kondisi ini dipengaruhi oleh pemberian ASI eksklusif, riwayat pemberian vitamin A, berat badan lahir, penduduk miskin, imunisasi campak, dan faktor lingkungan seperti sumber air minum bersih (Kemenkes, 2023; *Siloam Hospitals*, 2024).

Pneumonia merupakan penyebab utama kematian pada balita. *World Health Organization* (WHO) melaporkan sebanyak 14% kematian anak usia balita disebabkan oleh pneumonia. Pada tahun 2019 tercatat sebanyak 740.180 anak meninggal karena pneumonia. Penyakit ini menjadi penyebab kematian balita terbesar di Indonesia. Berdasarkan data Riskesdas (Riset Kesehatan Dasar) 2018, prevalensi pneumonia di Indonesia yang merupakan salah satu ukuran epidemiologi untuk menggambarkan jumlah kasus suatu penyakit atau kondisi kesehatan tertentu dalam populasi pada waktu atau periode tertentu, mencapai 2,1% dengan prevalensi tertinggi pada kelompok anak usia 12 hingga 23 bulan (Centauri, 2022). Melihat jumlah kematian akibat pneumonia yang tinggi perlu dilakukan upaya dalam menangani kasus tersebut seperti analisis regresi. Analisis regresi tidak secara langsung mengatasi masalah pneumonia pada balita, tetapi dapat memberikan pertimbangan dalam pencegahan masalah ini.

Analisis regresi merupakan metode statistika yang digunakan untuk memodelkan hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor (Drapper, 1992). Terdapat tiga pendekatan dalam analisis regresi untuk mengetahui hubungan antara variabel respon dengan variabel prediktor yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik (Sihotang dkk., 2021). Menurut Wulandari dan Budiantara (2014) regresi parametrik digunakan ketika kurva regresi membentuk pola tertentu seperti kubik, linier atau kuadratik. Regresi nonparametrik digunakan apabila kurva regresi tidak membentuk pola tertentu dan regresi semiparametrik digunakan jika dalam suatu model regresi ditemukan komponen parametrik dan nonparametrik (Dewi & Budiantara, 2018).

Analisis regresi nonparametrik memiliki fleksibilitas yang tinggi dalam mengestimasi kurva regresi. Salah satu model regresi nonparametrik yang sering digunakan adalah *Multivariate Adaptive Regression Spline* (MARS). Friedman pertama kali mengembangkan metode *Multivariate Adaptive Regression Spline* (MARS) pada tahun 1991 sebagai suatu pendekatan regresi nonparametrik yang bertujuan untuk memprediksi suatu nilai dari variabel respon dari sekumpulan/set variabel prediktor tanpa asumsi mengenai hubungan fungsi dasar antara variabel respon dengan variabel prediktor (Diana dan Rory, 2020). Model MARS disusun untuk mengatur koefisien beberapa fungsi basis yang dikendalikan oleh data regresi. Model ini memiliki keunggulan dalam menangani data berdimensi tinggi atau yang dikenal sebagai *curse of dimensionality*, serta mampu menghasilkan prediksi variabel

respon dengan tingkat akurasi yang tinggi. MARS merupakan pengembangan dari pendekatan *Recursive Partition Regression* (RPR) yang dikombinasikan dengan metode spline sehingga model yang dihasilkan kontinu pada titik knot (Sita & Otok, 2014).

Data kasus pneumonia pada balita memiliki distribusi Poisson karena termasuk data diskrit dengan peluang kejadian yang jarang terjadi dalam rentang waktu tertentu sehingga data ini dapat dimodelkan menggunakan regresi Poisson. Regresi poisson merupakan metode regresi nonlinier yang digunakan untuk menganalisis data dengan variabel responnya data diskrit dan berasal dari distribusi poisson. Salah satu asumsi yang penting dalam regresi poisson adalah ekuidispersi yaitu nilai variansi dan rata-rata dari variabel respon tersebut sama (McCullagh dan Nelder, 1989). Namun, terkadang terjadi pelanggaran asumsi seperti nilai variansinya lebih besar dari nilai rata-rata (overdispersi) atau lebih kecil dari nilai rata-rata (underdispersi). Pelanggaran tersebut dapat ditangani dengan model *Generalized Poisson Regression* (GPR) yang merupakan perluasan dari regresi Poisson (Masfian dkk., 2016).

Metode yang menggabungkan GPR dengan MARS dikenal sebagai *Multivariate Adaptive Generalized Poisson Regression Spline* (MAGPRS) (Prastika dkk., 2021). Menurut Otok dkk (2019), metode ini digunakan dalam model regresi nonparametrik untuk menangani data dengan distribusi poisson yang menghadapi permasalahan overdispersi atau underdispersi. Selain itu, metode ini digunakan pada data yang berdimensi tinggi dengan jumlah variabel prediktor sebanyak 3 sampai 20 variabel (Irvandi dkk., 2023). Pemodelan MAGPRS memperhatikan tiga komponen yaitu *Basis Function* (BF), *Maximum Interaction* (MI) dan *Minimum Observation* (MO), sementara pemilihan model terbaik dilakukan dengan melihat nilai *Generalized Cross Validation* (GCV) terkecil dan nilai R^2 terbesar. Pada penerapannya, maksimum jumlah BF yang disarankan adalah 2 hingga 4 kali variabel prediktor, sementara nilai MI yang diperbolehkan adalah 1,2,3 dan nilai MO yang disarankan adalah 0,1,2,3.

Pemodelan MAGPRS telah dilakukan oleh beberapa peneliti, diantaranya Kartini dan Ishlahuddin (2022) menggunakan *Multivariate Adaptive Generalized Poisson Regression Spline* untuk pengembangan model prediksi produksi padi di Kabupaten Bojonegoro dan memperoleh faktor-faktor yang memengaruhi produksi padi dan prediksi produksi padi untuk beberapa waktu yang akan datang. Prastika, dkk (2021) memodelkan *Multivariate Adaptive Generalized Poisson Regression Spline* pada kasus jumlah kematian ibu di Provinsi Jawa Timur dan menghasilkan faktor berpengaruh secara signifikan terhadap jumlah kematian ibu. Irvandi dkk (2023) memodelkan *Multivariate Adaptive Generalized Poisson Regression Spline* pada kasus *multi-drug resistant tuberculosis* di Kalimantan Barat menghasilkan faktor yang memengaruhi kasus *multi-drug resistant tuberculosis*.

Berdasarkan uraian tersebut, penelitian ini merupakan salah satu upaya pencegahan jumlah kasus pneumonia pada balita yang semakin tinggi dengan memodelkan kasus pneumonia pada balita dengan faktor-faktor yang diduga memengaruhi sehingga dapat diketahui faktor yang berpengaruh signifikan terhadap

jumlah kasus pneumonia pada balita, sehingga penulis tertarik untuk membahas tentang pemodelan *Multivariate Adaptive Generalized Poisson Regression Spline* pada Data Penderita Pneumonia Usia Balita di Provinsi Sulawesi Selatan tahun 2021.

1.2 Tujuan dan Manfaat

Tujuan dilakukannya penelitian ini adalah:

1. Memperoleh model data jumlah kasus pneumonia usia balita di Provinsi Sulawesi Selatan tahun 2021 berdasarkan *Multivariate Adaptive Generalized Poisson Regression Spline* dengan menggunakan kombinasi *Basis Function* (BF), *Maximum Interaction* (MI) dan *Minimum Observation* (MO) yang telah ditentukan.
2. Mengetahui faktor yang memengaruhi pneumonia usia balita di Provinsi Sulawesi Selatan tahun 2021.

Manfaat yang ingin dicapai penelitian ini antara lain:

1. Memberikan wawasan yang lebih khusus kepada penulis tentang pemodelan *Multivariate Adaptive Generalized Poisson Regression Spline* (MAGPRS)
2. Mengkaji faktor penyebab kasus pneumonia pada balita di Provinsi Sulawesi Selatan sehingga diharapkan dapat menjadi bahan pertimbangan bagi pemerintah untuk mengambil kebijakan.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan adalah jumlah kasus pneumonia pada balita di Provinsi Sulawesi Selatan tahun 2021 yang diperinci berdasarkan 24 kabupaten/kota.
2. Kombinasi *Basis Function* (BF) yang digunakan adalah 12, 18 dan 24.
3. *Maximum Interaction* (MI) yang digunakan adalah 1, 2 dan 3.
4. *Minimum Observation* (MO) yang digunakan adalah 0, 1, 2, dan 3.

1.4 Teori

1.4.1 Analisis Regresi

Analisis regresi merupakan salah satu analisis dalam statistika yang dapat digunakan untuk menyelidiki pola hubungan antara variabel respon dan variabel prediktor. Tujuan utama dalam melakukan analisis regresi adalah untuk mencari bentuk estimasi kurva regresi yang tepat (Budiantara, 2019). Secara umum, model regresi yang paling sederhana memiliki persamaan yang ditunjukkan pada Persamaan (1).

$$y = f(x_i) + \varepsilon \quad (1)$$

Eubank (1999) menyatakan bahwa terdapat tiga pendekatan dalam melakukan estimasi kurva regresi, yaitu pendekatan regresi parameterik, regresi semiparametrik dan regresi nonparametrik. Pendekatan regresi parameterik dapat digunakan apabila bentuk kurva regresi diketahui. Pendekatan regresi semiparametrik dapat digunakan apabila bentuk kurva regresi sebagian diketahui dan sebagian tidak diketahui, sedangkan pada pendekatan regresi nonparametrik

bentuk kurva regresi diasumsikan tidak diketahui. Pendekatan regresi nonparametrik tidak bergantung pada asumsi bentuk kurva regresi tertentu, sehingga memiliki fleksibilitas yang tinggi (Dani dkk, 2021).

1.4.2 Regresi Nonparametrik

Regresi nonparametrik merupakan metode statistika yang digunakan apabila bentuk pola hubungan antara variabel respon dengan variabel prediktor tidak diketahui bentuk fungsi regresinya. Pendekatan ini hanya mengasumsikan bahwa kurva regresi bersifat mulus (*smooth*), yang berarti termuat dalam ruang fungsi tertentu sehingga memiliki sifat fleksibilitas yang tinggi. Model regresi nonparametrik secara umum adalah sebagai berikut:

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

dengan

- y_i : variabel respon pengamatan ke- i
- $f(x_i)$: fungsi regresi nonparametrik yang tidak diketahui
- x_i : variabel prediktor pengamatan ke- i
- ε_i : *error* acak pada pengamatan ke- i yang diasumsikan identik, independen dan berdistribusi $N(0, \sigma^2)$

Persamaan (2) dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon} \quad (3)$$

Matriks dan vektor pada Persamaan (3) dapat ditulis (Hardle, 1994 dalam Salam, 2013):

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

1.4.3 Generalized Linear Model

Generalized Linear Model (GLM) adalah analisis regresi yang variabel responnya mengikuti distribusi keluarga eksponensial yaitu normal, binomial, Poisson, geometrik, binomial negatif, eskponensial, dan gamma. Ada tiga komponen utama dalam analisis GLM, yaitu:

1. Komponen acak, terdiri dari variabel respon Y dengan observasi bebas ($y_1, \dots, y_n$) dari sebuah distribusi dalam keluarga eksponensial. Bentuk fungsi densitas probabilitas dari distribusi keluarga eksponensial, yaitu sebagai berikut.

$$f(y_i; \theta_i, \phi) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{a} + c(y_i; \phi) \right) \quad (4)$$

2. Komponen sistematis adalah hubungan dari sebuah vektor ($\eta_1, \dots, \eta_n$) untuk menjelaskan variabel-variabel yang berhubungan dalam sebuah model linear

$$\eta_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (5)$$

atau dalam bentuk matriks

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

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_3 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

dengan $\boldsymbol{\eta}$ adalah vektor ($n \times 1$) dari observasi, $\mathbf{X}$ adalah matriks ($n \times c$) dari variabel bebas, $\boldsymbol{\beta}$ adalah matriks ($c \times 1$) dari koefisien regresi dengan $c = p + 1$.

3. Fungsi penghubung (*link function*) adalah fungsi yang menghubungkan ekspektasi variabel respon dengan komponen sistematis. Misalkan fungsi $\mu_i = E(Y_i), i = 1, 2, \dots, n$, maka model *link* μ_i untuk η_i adalah $\eta_i = g(\mu_i)$

$$g(\mu_i) = \eta_i = \sum_{j=1}^p \beta_j x_{ij}, \quad i = 1, 2, \dots, n \quad (6)$$

fungsi *link* $g(\mu_i) = \mu_i$ dinamakan identitas *link* dengan $\eta_i = \mu_i$. Fungsi *link* yang mentransformasikan nilai rata-ratanya ke parameter natural dinamakan *kanonikal link* sehingga $g(\mu_i) = \theta_i$ dan $\theta_i = \sum_{j=1}^p \beta_j x_{ij}$ (Safrida dkk, 2013).

1.4.4 Distribusi Poisson

Distribusi Poisson merupakan suatu distribusi untuk peristiwa yang probabilitas terjadinya kecil yaitu terjadinya bergantung pada interval waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel saling independen. Interval waktu tersebut dapat berapa saja lamanya misalnya semenit, sehari, seminggu, sebulan atau bahkan setahun. Daerah tertentu yang dimaksudkan dapat berupa suatu garis, luas, volume, atau mungkin sepotong bahan (Walpole, 1995).

Menurut Cameron dan Trivedi (1998), suatu variabel random Y yang bertipe diskrit akan mengikuti distribusi Poisson jika μ adalah rata-rata suatu kejadian per unit waktu dan t adalah periode waktu tertentu, maka rata-rata dari y menjadi μt . Jadi, peluang terjadinya kejadian y pada periode waktu ke- t diberikan oleh Persamaan (7) berikut (Darnah, 2011):

$$f(y, \mu) = \frac{e^{-\mu t} (\mu t)^y}{y!}, \quad y = 0, 1, 2, \dots; \mu > 0 \quad (7)$$

dengan:

$f(y, \mu)$: probabilitas kelas sukses

e : konstanta (2,7182...)

μ : rata-rata keberhasilan (np)

y : banyaknya unsur berhasil dalam sampel atau variabel acak diskrit

n : jumlah/ukuran populasi

t : periode waktu

Bila selang waktu kejadian adalah sama, maka fungsi distribusi peluang untuk variabel acak Poisson Y dengan parameter μ diberikan oleh Persamaan (8) berikut (Darnah, 2011):

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots; \mu > 0 \quad (8)$$

dengan μ merupakan parameter rata-rata dari variabel acak y yang diasumsikan untuk semua kejadian y . Kemudian distribusi Poisson mengasumsikan bahwa nilai μ dari variabel acak mempunyai nilai yang sama dengan variansinya.

1.4.5 Regresi Poisson

Regresi Poisson merupakan model regresi yang dapat digunakan pada data yang variabel responnya berdistribusi tidak normal dan berjenis diskrit. Regresi Poisson merupakan penerapan dari *Generalized Linear Model* (GLM) yang menggambarkan hubungan antara variabel Y data diskrit berdistribusi Poisson dengan variabel prediktor X (Simamata, 2011). Distribusi Poisson memberikan suatu model yang realistis untuk berbagai macam kejadian acak yaitu nilai dari peubah acak Poisson berupa bilangan bulat non-negatif. Fungsi massa peluang dari distribusi Poisson diberikan pada Persamaan (8).

Persamaan (8) digunakan untuk menghitung peluang acak Y dengan rata-rata jumlah kejadian ($E[Y] = \mu = \text{var}(Y)$). Model untuk regresi Poisson menyatakan rata-rata dari distribusi diskrit sebagai fungsi dari variabel diskrit. Model regresi Poisson diberikan pada Persamaan (9) berikut:

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, 2, 3, \dots, n \quad (9)$$

dengan y_i adalah jumlah kejadian dan μ_i adalah rata-rata kejadian dengan μ_i diasumsikan tidak berubah dari data ke data.

Model regresi Poisson merupakan penerapan dari *Generalized Linear Model* (GLM). Terdapat sebuah fungsi g yang linear yang menghubungkan rata-rata dari variabel respon dengan variabel prediktor, yaitu:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} = \mathbf{x}_i^T \boldsymbol{\beta} \quad (10)$$

dengan g merupakan fungsi penghubung sedangkan hubungan dengan rata-rata pada penduga linear diberikan pada Persamaan (11):

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (11)$$

Hubungan antara rata-rata dari variabel terikat dengan variabel respon bebas linear untuk fungsi penghubung log adalah:

$$\begin{aligned} \ln \mu_i &= \mathbf{x}_i^T \boldsymbol{\beta} \\ e^{\ln \mu_i} &= e^{\mathbf{x}_i^T \boldsymbol{\beta}} \\ \mu_i &= e^{\mathbf{x}_i^T \boldsymbol{\beta}} \end{aligned} \quad (12)$$

sehingga fungsi penghubung untuk model regresi Poisson mempunyai logaritma sebagai berikut:

$$\begin{aligned} \ln E(Y|x) &= \ln \mu_i = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \\ \mu_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) \end{aligned} \quad (13)$$

dan fungsi massa peluang distribusi Poisson dapat ditulis sebagai berikut:

$$f(y_i, \beta) = \frac{e^{(-\exp(x_i^T \beta))} [\exp(x_i^T \beta)]^{y_i}}{y_i!} \quad (14)$$

Rata-rata dan variansi untuk model regresi Poisson yaitu:

$$\begin{aligned} \ln \mu_i &= x_i^T \beta \\ e^{\ln \mu_i} &= e^{x_i^T \beta} \\ \mu_i &= e^{x_i^T \beta} \\ \mu_i &= \mu(y_i; \beta) = \exp(x_i^T \beta) \\ \text{var}(Y_i) &= \mu(y_i; \beta) = \exp(x_i^T \beta) \end{aligned} \quad (15)$$

sehingga model regresi Poisson dapat ditulis sebagai berikut (Wibawa & Abapihi, 2023):

$$y_i = \exp(x_i^T \beta) + \varepsilon_i \quad (16)$$

dengan:

- y_i : jumlah kejadian ke- i
- μ_i : rata-rata jumlah kejadian
- β : koefisien estimasi
- ε_i : galat *error*

1.4.6 Overdispersi

Ekuidispersi adalah asumsi yang harus dipenuhi dalam regresi poisson yaitu nilai rata-rata dan variansi pada variabel dependen harus bernilai sama. Menurut Harahap R. I. (2020) pada pengujian ekuidispersi kadang terjadi pelanggaran asumsi dalam analisis data (Maulana dkk., 2023). Namun, dalam analisis data *count* seringkali dijumpai data yang variansinya lebih besar dari rata-rata. Keadaan seperti ini disebut dengan overdispersi. Fenomena overdispersi dapat dijelaskan pada Persamaan (17) berikut:

$$\text{Var}(Y) > E(Y) \quad (17)$$

sebaliknya, data yang variansinya lebih kecil dari rata-rata disebut dengan underdispersi (Mc Cullagh & Nelder, 1989).

Overdispersi atau underdispersi dapat menyebabkan taksiran parameter yang diperoleh tidak efisien, walaupun cenderung tetap konsisten. Penggunaan yang tidak tepat dari regresi Poisson (yang mengandung masalah overdispersi maupun underdispersi) dapat berakibat fatal dalam interpretasi model, khususnya parameter model karena dapat menaksir standar *error* yang terlalu rendah dan dapat memberikan kesimpulan yang keliru tentang signifikan atau tidaknya parameter regresi yang terlibat (Darnah, 2011).

Overdispersi dapat dilihat dari nilai *Deviance* atau *Pearson Chi-square* yaitu sebagai berikut:

1. *Deviance*

Nilai *deviance* adalah nilai logaritma dari uji rasio *likelihood*-nya. Uji rasio *likelihood*-nya membandingkan *current* modelnya dan *saturated* modelnya yang dituliskan sebagai berikut:

$$\phi_1 = \frac{D^2}{db}$$

$$D^2 = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{y}_i} \right) - (y_i - \hat{y}_i) \right) \quad (18)$$

2. Pearson Chi-square

Pengukuran lain yang digunakan untuk mendeteksi overdispersi yaitu statistik *pearson chi-square* yang didefinisikan pada persamaan berikut:

$$\phi_2 = \frac{\chi^2}{db}$$

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \quad (19)$$

dengan:

D^2 : Nilai *deviance*

χ^2 : Nilai *pearson chi-square*

y_i : Nilai variabel respon

$\hat{y}_i$: Estimasi regresi Poisson

dengan $db = n - c$ ($n > c$) dengan $c = p + 1$ merupakan banyaknya parameter, n merupakan banyaknya pengamatan. Jika nilai ϕ_1 dan ϕ_2 lebih besar dari 1, ini menunjukkan nilai variansi yang lebih besar daripada rata-rata, maka telah terjadi overdispersi (Fadil dkk, 2024).

1.4.7 Generalized Poisson Regression

Menurut Melliana (2013), penanganan pelanggaran asumsi ekuidispersi pada model regresi Poisson dapat dikembangkan dengan menggunakan model *Generalized Poisson Regression* (GPR). Model GPR hampir mirip dengan model regresi Poisson tetapi diasumsikan komponen acaknya didistribusikan menjadi *Generalized Poisson* (GP). Dengan kata lain, model GPR dapat digunakan untuk data diskrit yang mempunyai distribusi Poisson tanpa adanya asumsi ekuidispersi (Ruliana dkk., 2016). Model *Generalized Poisson Regression* (GPR) memiliki tambahan parameter θ . Misal, $y_i = 0, 1, 2, \dots$ merupakan variabel respon, maka distribusi GP diberikan pada Persamaan (20) berikut:

$$f(y_i; \mu_i; \theta) = \left(\frac{\mu_i}{1 + \theta \mu_i} \right)^{y_i} \frac{(1 + \theta y_i)^{y_i - 1}}{y_i} \exp \left(\frac{-\mu_i (1 + \theta y_i)}{1 + \theta \mu_i} \right) \quad (20)$$

dengan rata-rata dan varians sebagai berikut (Chaniago & Wulandari, 2022):

$$E(y_i) = \mu \quad (21)$$

$$Var(y_i) = \mu(1 + \theta \mu)^2 \quad (22)$$

Model GPR adalah sebagai berikut:

$$\mu_i = \exp(\mathbf{x}^T \boldsymbol{\beta}) \quad (23)$$

Penaksir parameter model GPR dilakukan dengan metode *Maximum Likelihood Estimator* (MLE). Fungsi likelihood untuk model GPR adalah sebagai berikut:

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^n f(y_i; \mu; \theta)$$

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^n \left\{ \left(\frac{\mu_i}{1 + \theta \mu_i} \right)^{y_i} \frac{(1 + \theta y)^{y_i - 1}}{y_i!} \exp \left(\frac{-\mu(1 + \theta y_i)}{1 + \theta \mu} \right) \right\} \quad (24)$$

Persamaan (21) diubah dalam bentuk fungsi logaritma natural menjadi:

$$\ln L(\boldsymbol{\beta}, \theta) = \ln \left[\prod_{i=1}^n \left\{ \left(\frac{\mu_i}{1 + \theta \mu_i} \right)^{y_i} \frac{(1 + \theta y)^{y_i - 1}}{y_i!} \exp \left(\frac{-\mu(1 + \theta y_i)}{1 + \theta \mu} \right) \right\} \right]$$

$$= \sum_{i=1}^n \left\{ y_i \ln(\mu_i) - y_i \ln(1 + \theta \mu_i) + (y_i - 1) \ln(1 + \theta y_i) - \ln(y_i!) - \frac{\mu_i(1 + \theta y_i)}{1 + \theta \mu_i} \right\} \quad (25)$$

substitusi $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ dan memaksimumkan model *log likelihood* sehingga diperoleh penaksir *maximum likelihood* untuk $\boldsymbol{\beta}$ dan θ pada Persamaan (22) dan (23) berikut:

$$\frac{\partial \ln L(\boldsymbol{\beta}, \theta)}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^n \left\{ \left[\left(\frac{y_i \mathbf{x}_i^T - \frac{\theta y_i \mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \theta \exp(\mathbf{x}_i^T \boldsymbol{\beta})} - (1 + \theta y_i) \right) \right] \right\} = \mathbf{0} \quad (26)$$

$$\frac{\partial \ln L(\boldsymbol{\beta}, \theta)}{\partial \theta} = \sum_{i=1}^n \left\{ \left[\left(\frac{-y_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \theta \exp(\mathbf{x}_i^T \boldsymbol{\beta}))} + \frac{y_i(y_i - 1)}{(1 + \theta y_i)} - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \right) \right] \right\} = \mathbf{0} \quad (27)$$

Persamaan (26) dan (27) memperlihatkan bahwa turunan pertama dari model fungsi *log likelihood* pada estimasi parameter menggunakan metode MLE tidak dalam bentuk linier. Oleh karena itu, prosedur estimasi yang dapat dilakukan dengan menggunakan iterasi numerik. Iterasi numerik yang digunakan dalam penelitian ini adalah *Iteratively Reweighted Least Squares* (IRLS) (Arisandi dkk., 2018).

1.4.8 Spline

Spline merupakan model polinom tersegmen atau terpotong-potong yang mulus dan dapat menghasilkan fungsi regresi yang sesuai dengan data. Mengestimasi *spline* tergantung pada titik knot. Titik knot merupakan suatu titik perpaduan yang terjadi karena perubahan pola perilaku dari suatu fungsi pada selang yang berbeda. Fungsi *spline* pada suatu fungsi f dengan orde p dapat dinyatakan sebagai berikut

$$f(x_i) = \sum_{j=0}^p \beta_j x_i^j + \sum_{l=1}^r \beta_{(p+1)} (x_i - k_l)_+^p \quad (28)$$

dengan k menyatakan banyaknya titik knot dan $(x_i - k_l)_+^p$ menyatakan fungsi potongan (*truncated*) yang dapat dijabarkan sebagai berikut.

$$(x_i - k_l)_+^p = \begin{cases} (x_i - k_l)^p, & x_i \geq k_l \\ 0, & x_i < k_l \end{cases} \quad (29)$$

Bentuk matematis dari fungsi *spline* pada Persamaan (29), dapat dinyatakan bahwa *spline* adalah potongan-potongan polinom yang berbeda digabungkan bersama titik knot $k_1, k_2, k_3, \dots, k_r$ untuk menjamin sifat kontinuitasnya (Putra dkk, 2015).

1.4.9 Multivariate Adaptive Regression Spline

Multivariate Adaptive Regression Spline (MARS) merupakan salah satu metode nonparametrik yang diperkenalkan oleh Friedman pada tahun 1991. Model MARS berguna sebagai metode baru yang mengakomodasi pembagunan model-model prediksi yang lebih akurat untuk variabel-variabel kontinu maupun kategorik biner. Model ini mampu mengatasi permasalahan data yang berdimensi tinggi yaitu memiliki 3 hingga 20 variabel prediktor dan jumlah sampel yang diharuskan untuk pendekatan metode ini adalah $50 \leq n \leq 1000$.

Menurut Friedman (1991) ada beberapa istilah yang perlu diperhatikan dalam metode dan pemodelan (MARS) adalah sebagai berikut:

1. Knot merupakan nilai variabel prediktor ketika *slope* (kemiringan) suatu garis regresi mengalami peubah yang dapat didefinisikan sebagai akhir dari sebuah garis regresi dan awal dari garis regresi lainnya. Minimum jarak antar knot atau *Minimum Observation* (MO) antara knot yang digunakan sebesar 0,1,2, dan 3.
2. *Basis Function* (BF) merupakan fungsi parameter yang menjelaskan hubungan antara variabel respon dan variabel prediktor yang dapat didefinisikan pada tiap region. Maksimum fungsi basis yang disarankan adalah 2 hingga 4 kali jumlah variabel prediktor.
3. Interaksi adalah perkalian silang antara variabel yang saling berkorelasi. Jumlah *Maximum Interaction* (MI) yang diperbolehkan adalah 1,2,3. Ketika MI melebihi 3 maka akan menyebabkan model semakin kompleks dan sulit untuk diinterpretasikan serta nilai *Generalized Cross Validation* (GCV) akan meningkat.

MARS merupakan pengembangan dari pendekatan *Recursive Partitoning Regression* (RPR) dan *Generalized Cross Validation* (GCV). Metode MARS juga mengatasi permasalahan *Recursive Partitoning Regression* (RPR) yaitu menghasilkan model yang tidak kontinu pada titik knot dan dapat mengidentifikasi adanya fungsi linier aditif (Damayanti & Sunendiari, 2021). Friedman (1991) memberikan persamaan MARS untuk estimasi fungsi f dalam Persamaan (30) berikut (Prastika dkk., 2021):

$$f(x_i) = \alpha_o + \sum_{m=1}^M \alpha_m B_{mi}(x_i) + \varepsilon_i \quad (30)$$

dengan fungsi basis *Multivariate Adaptive Regression Spline* (MARS):

$$B_{mi}(x_i) = \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)i} - t_{km})] \quad (31)$$

dengan:

i : banyaknya pengamatan (1,2, ..., n)

- α_0 : konstanta regresi fungsi basis
 α_m : koefisien fungsi basis ke- m
 B_{mi} : fungsi basis ke- m untuk observasi ke- i
 M : maksimum fungsi basis (*nonconstant* fungsi basis)
 K_m : maksimum derajat interaksi pada fungsi basis ke- m
 S_{km} : tanda pada titik knotnya (bernilai ± 1)
 $+1$ jika data berada di sebelah kanan titik knot, maka

$$[+(x_{v(k,m)i} - t_{(k,m)})]_+ = \begin{cases} x_{v(k,m)i} - t_{(k,m)}, & \text{jika } x_{v(k,m)i} > t_{(k,m)} \\ 0, & \text{sebaliknya} \end{cases}$$
 -1 jika data berada di sebelah kiri titik knot, maka

$$[-(x_{v(k,m)i} - t_{(k,m)})]_+ = \begin{cases} x_{v(k,m)i} - t_{(k,m)}, & \text{jika } x_{v(k,m)i} < t_{(k,m)} \\ 0, & \text{sebaliknya} \end{cases}$$
 $x_{v(k,m)}$: variabel prediktor ke- v dengan interaksi ke- k dan fungsi basis ke- m
 t_{km} : nilai knot dari interaksi ke- k dan fungsi basis ke- m
 sehingga jika ditulis dalam bentuk matriks dapat dilihat pada Persamaan (32) berikut (Darmawanti dkk., 2014):

$$\mathbf{y} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (32)$$

dengan nilai $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\boldsymbol{\alpha} = (a_0, a_1, a_2, \dots, a_M)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$. $\mathbf{y}$ merupakan variabel respon yang berukuran $(n \times 1)$, $\boldsymbol{\alpha}$ merupakan vektor yang memuat koefisien regresi berukuran $((M + 1) \times 1)$, $\boldsymbol{\varepsilon}$ merupakan vektor *random error* berukuran $(n \times 1)$, dan nilai $\mathbf{B}$ merupakan matriks berukuran $(n \times (M + 1))$ sebagai berikut.

$$\mathbf{B} = \begin{pmatrix} 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{v(k,1)1} - t_{k1})]_+ & \cdots & \prod_{k=1}^{K_M} [s_{kM}(x_{v(k,M)1} - t_{kM})]_+ \\ 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{v(k,1)2} - t_{k1})]_+ & \cdots & \prod_{k=1}^{K_M} [s_{kM}(x_{v(k,M)2} - t_{kM})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{v(k,1)n} - t_{k1})]_+ & \cdots & \prod_{k=1}^{K_M} [s_{kM}(x_{v(k,M)n} - t_{kM})]_+ \end{pmatrix} \quad (33)$$

Pembentukan model MARS meliputi dua tahap, yaitu *forward stepwise* dan *backward stepwise*. Pada *forward stepwise*, dihasilkan fungsi basis hingga M fungsi basis. Fungsi basis yang tidak konstan pada tahap ini selanjutnya dieliminasi menggunakan *backward stepwise* untuk memenuhi konsep parsimoni dengan meminimumkan nilai GCV (Friedman, 1991). Hasil dari *backward stepwise* memberikan kombinasi fungsi basis yang optimal bagi model MARS.

Model yang diperoleh pada *backward stepwise* bukan merupakan model akhir yang akan digunakan, karena pada tahap tersebut masih diperoleh beberapa kombinasi fungsi basis dengan model yang berbeda. Oleh karena itu, dibutuhkan kriteria pemilihan model MARS yang terbaik untuk mewakili data penelitian. (Friedman, 1991) menggunakan GCV untuk memilih model terbaik (Widagdo dkk., 2020).

$$GCV(M) = \frac{MSE}{\left[1 - \frac{\tilde{C}(M)}{n}\right]} = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(\mathbf{x}_i)]^2}{\left[1 - \frac{\tilde{C}(M)}{n}\right]^2} \quad (34)$$

Dengan $\tilde{C}(M) = C(M) + d \cdot M$, $\hat{f}_m(\mathbf{x}_i) = \hat{\alpha}_0 + \sum_{m=1}^M \hat{\alpha}_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)i} - t_{km})] +$
dan $C(M) = \text{tr}[(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$
dengan:

$\hat{f}_M$: nilai taksiran variabel respon pada fungsi basis M

n : jumlah data

$\tilde{C}(M)$: fungsi kompleks

d : derajat interaksi, $2 \leq d \leq 4$

y_i : nilai variabel respon ke- i

$\hat{f}_m(\mathbf{x}_i)$: nilai taksiran variabel respon pada fungsi basis M

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n [y_i - \hat{y}_i]^2}{\sum_{i=1}^n [y_i - \bar{y}_i]^2} \quad (35)$$

Model terbaik MARS adalah model dengan nilai GCV terkecil dan nilai R^2 terbesar (Prastika dkk., 2021).

Analisis kepentingan variabel diperlukan untuk memaksimalkan sistem penyeleksian. Nilai kepentingan variabel prediktor pada *package R earth* mengacu pada tiga kriteria yaitu *nsubset*, GCV dan RSS (*Residual Sum of Squares*).

1. Kriteria *nsubset* dilakukan dengan menghitung jumlah *subsets* model yang menyertakan variabel di dalamnya. Semakin sering suatu variabel muncul dalam subset, semakin besar pengaruhnya terhadap model.
2. Kriteria RSS dilakukan dengan menghitung penurunan RSS untuk setiap *subsets*. Setiap penambahan variabel akan dihitung penurunan RSS-nya. Variabel dengan penurunan RSS terbesar dianggap sebagai variabel yang memberikan kontribusi signifikan dalam mengurangi kesalahan model.
3. Kriteria GCV, apabila penambahan variabel meningkatkan nilai GCV maka variabel tersebut dianggap kurang penting atau memiliki pengaruh yang tidak baik pada model. Variabel yang memberikan penurunan GCV terbesar dianggap paling relevan.

Selanjutnya untuk mempermudah dalam interpretasi, penurunan nilai RSS atau GCV diubah ke dalam skala, di mana penurunan terbesar diberikan skala 100 (Oktora, 2015).

1.4.10 Multivariate Adaptive Generalized Poisson Regression Spline

Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS) adalah metode yang menggabungkan metode *Generalized Poisson Regression* (GPR) dengan metode *Multivariate Adaptive Regression Spline* (MARS) (Prastika dkk., 2021). Metode MAGPRS digunakan untuk data yang berdistribusi Poisson dan memiliki permasalahan overdispersi atau underdispersi, serta data yang digunakan berdimensi tinggi, serta model regresi yang berbentuk nonparametrik. Model MAGPRS dituliskan sebagai berikut:

$$\begin{aligned}
Y_i &\sim GP(\mu_i, \theta) \\
\ln(\mu_i) &= f(x_i) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)i} - t_{km})_+] \\
&= \alpha_0 + \sum_{m=1}^M \alpha_m B_{mi}(\mathbf{x}_i) \\
\mu_i &= \exp\left(\alpha_0 + \sum_{m=1}^M \alpha_m B_{mi}(\mathbf{x}_i)\right) = \exp(\beta_i^T \vec{\alpha}) \\
\mu_i &= \exp(\mathbf{B}\alpha) \\
\ln(\mu_i) &= \mathbf{B}\alpha
\end{aligned} \tag{36}$$

1.4.11 Uji Signifikansi Parameter

1.4.11.1 Uji Simultan

Pengujian secara simultan atau serentak MAGPRS menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dan pengujian parameter secara parsial menggunakan statistik uji T. Hipotesis pengujian serentak adalah sebagai berikut:

$$H_0: \alpha_1 = \dots = \alpha_M = 0$$

$$H_1: \text{paling sedikit ada satu } \alpha_m \neq 0, m = 1, \dots, M$$

Statistik uji:

$$\begin{aligned}
\hat{G}^2 &= -2 \ln \Lambda = 2(\ln L(\hat{\Omega}) - \ln L(\hat{\omega})) \\
&= 2 \left[\frac{\sum_{i=1}^n y_i (\mathbf{B}\hat{\alpha}) - \sum_{i=1}^n y_i \ln(\hat{\alpha}_0) + \sum_{i=1}^n y_i \frac{\ln(1 + \hat{\theta} \exp(\hat{\alpha}_0))}{\ln(1 + \hat{\theta} \exp(\mathbf{B}\hat{\alpha}))}}{\sum_{i=1}^n \frac{\exp(\mathbf{B}\hat{\alpha}) (1 + \hat{\theta} y_i)}{(1 + \hat{\theta} \exp(\mathbf{B}\hat{\alpha}))} + \sum_{i=1}^n \frac{\exp(\hat{\alpha}_0) (1 + \hat{\theta} y_i)}{(1 + \hat{\theta} \exp(\hat{\alpha}_0))}} \right]
\end{aligned} \tag{37}$$

$\hat{G}^2$ berdistribusi χ^2 , dengan $v = n(\Omega) - n(\omega)$. Daerah penolakan hasil dari parameter yang diperoleh merupakan suatu nilai *ln likelihood* dibawa H_0 dan populasi. Kriteria penolakan H_0 jika $\hat{G}^2 > \chi^2_{\alpha, v}$. Jika H_0 ditolak pada pengujian serentak maka Langkah selanjutnya adalah melakukan pengujian secara parsial (Hidayati & Otok, 2019).

1.4.11.2 Uji Parsial

Pengujian secara parsial atau individu dilakukan untuk mengetahui variabel prediktor yang mempunyai pengaruh signifikan terhadap variabel respon pada fungsi basis di dalam model. Hipotesis pengujian parsial adalah sebagai berikut:

$$H_0: \alpha_m = 0$$

$$H_1: \alpha_m \neq 0, m = 1, \dots, M$$

Statistik uji:

$$t_{hitung} = \frac{\hat{\alpha}_m}{se(\hat{\alpha}_m)}; \quad se(\hat{\alpha}_m) = \sqrt{\left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}\right) C_{mm}} \tag{38}$$

C_{mm} adalah elemen-elemen pada diagonal utama matriks $(\mathbf{B}^T \mathbf{B})^{-1}$. Kriteria Penolakan H_0 jika $|t_{hit}| > t_{(\frac{\alpha}{2}, n-k)}$ atau $P - value < \alpha$ artinya bahwa semua fungsi

basis yang terpilih dalam model MAGPRS memiliki pengaruh terhadap variabel respon (Prastika dkk., 2021).

1.4.12 Pneumonia pada Balita

Pneumonia merupakan penyakit pernapasan yang sering terjadi dibandingkan jenis penyakit pernapasan lainnya. Pneumonia terjadi ketika alveolus yaitu kantong udara yang ada di paru-paru terisi oleh nanah atau cairan lain. Akibatnya, pengidap pneumonia kesulitan untuk mendapat asupan oksigen yang cukup. Pneumonia dapat menyerang siapa saja tanpa memandang usia, termasuk balita (Makarim, 2019). Pneumonia merupakan penyebab utama kematian balita di dunia, lebih banyak dibandingkan dengan gabungan penyakit AIDS, malaria, dan campak. WHO (2012) menyatakan penyakit ini lebih banyak di bawah usia lima tahun dan diperkirakan 1,1 juta kematian setiap tahun disebabkan oleh pneumonia. WHO (2013) mengatakan bahwa diperkirakan dua balita meninggal setiap menit disebabkan oleh pneumonia.

Kasus pneumonia di Provinsi Sulawesi Selatan pada tahun 2020 diperkirakan mencapai 33.345 kasus. Namun, hanya ditemukan 2.736 penderita (8,21%). Sedangkan pada tahun 2021 diperkirakan pneumonia pada balita sebesar 33.108 orang. Namun, hanya ditemukan 2.443 kasus (7,4%) (Dinkes Sulsel, 2022). Adapun faktor-faktor yang memengaruhi jumlah kasus pneumonia pada balita adalah pemberian ASI eksklusif, riwayat pemberian vitamin A, berat badan lahir, penduduk miskin, imunisasi campak, dan faktor lingkungan seperti sumber air minum bersih (Kemenkes, 2023; *Siloam Hospitals*, 2024).

1.4.12.1 Pemberian ASI Eksklusif

Air Susu Ibu (ASI) merupakan sumber nutrisi yang sangat istimewa untuk bayi. ASI mengandung semua nutrisi yang dibutuhkan bayi selama enam bulan kehidupannya seperti karbohidrat, elektrolit, mineral serta imunoglobulin. Kementerian Kesehatan Indonesia (2014) membagi pola pemberian ASI menjadi tiga kategori, yaitu menyusui eksklusif, menyusui predominan dan menyusui parsial. Menyusui eksklusif adalah tidak memberikan makanan dan minuman lain pada bayi selain ASI, kecuali pemberian obat atau vitamin yang dibutuhkan bayi. Menyusui predominan adalah pernah memberikan asupan lain selain ASI, seperti air putih dan teh sebagai makanan atau minuman prelakteal sebelum ASI diperbolehkan sedangkan menyusui parsial adalah memberikan makanan atau minuman selain ASI sebelum bayi berusia 6 bulan, seperti susu formula, bubur dan lain-lain.

Penelitian Puspitasari dan Fariani (2015) menemukan bahwa status ASI eksklusif merupakan salah satu faktor risiko terjadinya pneumonia pada balita. Balita yang tidak mendapatkan ASI eksklusif berisiko 7 kali lebih besar terkena pneumonia dibandingkan dengan balita yang mendapatkan ASI eksklusif (Fauziah dkk., 2021).

1.4.12.2 Pemberian Vitamin A

Vitamin A merupakan salah satu zat gizi yang penting bagi bayi dan balita. Vitamin A akan melindungi bayi dan balita dari beberapa risiko penyakit sehingga pertumbuhan dan perkembangan bayi balita lebih optimal. Pemberian vitamin A pada

bayi dan balita memiliki beberapa manfaat seperti meningkatkan daya tahan tubuh terhadap penyakit dan infeksi, mendukung pertumbuhan dan perkembangan balita dan berperan penting dalam fungsi penglihatan (Koesdyahmurti dkk., 2023). Bayi dan balita yang tidak diberikan vitamin A secara rutin berisiko untuk terkena pneumonia dibandingkan dengan bayi dan balita yang diberikan vitamin A secara rutin. Hasil studi ini telah dilakukan oleh beberapa peneliti yaitu Dewi dkk (2020), Frini dkk, (2018), Gustom & Wulandari (2020) yang menemukan bahwa pemberian vitamin A merupakan faktor risiko kejadian pneumonia pada bayi dan balita.

1.4.12.3 Berat Badan Lahir

Berat badan lahir (BBL) menentukan pertumbuhan, perkembangan fisik dan mental pada balita. Berat badan bayi baru lahir dikatakan normal apabila bayi baru lahir memiliki berat badan 2.500 - 4.000 gram, sedangkan dikatakan berat badan lahir rendah (BBLR) apabila bayi baru lahir memiliki berat badan kurang dari 2.500 gram. Bayi dengan berat badan lahir rendah memiliki pembentukan zat anti kekebalan kurang sempurna. Pertumbuhan dan maturasi organ dan alat-alat tubuh belum sempurna akibatnya bayi dengan berat badan lahir rendah lebih berisiko terkena komplikasi dan infeksi, terutama pneumonia dan penyakit pernapasan lainnya (Rigusti dkk, 2019). Penelitian yang dilakukan oleh Nickontara dkk (2024) menemukan bahwa BBLR lebih berisiko terkena pneumonia dibandingkan dengan BBL Normal.

1.4.12.4 Penduduk Miskin

Faktor ekonomi berkontribusi besar terhadap penyakit saluran pernapasan. Penelitian di Amerika Serikat terhadap kematian pneumonia pada balita yang diamati sejak tahun 1939 – 1996 menunjukkan bahwa masyarakat ekonomi rendah lebih banyak mengunjungi pelayanan kesehatan dibandingkan masyarakat ekonomi tinggi. Salah satu program yang dilakukan untuk menurunkan kematian pneumonia pada balita pada tahun 1972 adalah meningkatkan akses penduduk miskin ke fasilitas pelayanan kesehatan dan program *The Women, Infants and Children* (Machmud, 2009). Selain itu, penelitian yang dilakukan oleh Machmud (2009) menemukan bahwa rumah tangga yang dikategorikan miskin memiliki risiko terkena penyakit pneumonia pada balita sebesar 1,73 kali dibandingkan dengan rumah tangga tidak miskin.

1.4.12.5 Imunisasi Campak

Imunisasi campak merupakan imunisasi yang digunakan untuk mencegah penyakit campak pada anak yang termasuk penyakit menular. Balita yang telah mendapatkan imunisasi campak diharapkan dapat terhindar dari penyakit campak dan pneumonia yang merupakan komplikasi paling sering terjadi pada anak yang mengalami campak. Menurut UNICEF-WHO pemberian imunisasi ini dapat mencegah infeksi yang dapat menyebabkan pneumonia sebagai komplikasi dari penyakit ini (Rigustia dkk, 2019). Penelitian yang dilakukan oleh Hartati (2011) menemukan bahwa balita yang tidak mendapatkan imunisasi campak berisiko 3,21 kali terkena penyakit pneumonia dibandingkan dengan balita yang mendapatkan imunisasi campak.

1.4.12.6 Sumber Air Minum Bersih

Air adalah salah satu elemen utama di bumi yang menjadi bagian tidak terpisahkan bagi seluruh manusia. Studi Kualitas Air Minum Rumah Tangga (SKAMRT) Kementerian Kesehatan pada tahun 2020 menyebutkan bahwa sebanyak 7 dari 10 rumah tangga Indonesia mengonsumsi air minum dari infrastruktur yang terkontaminasi oleh bakteri *E. coli*, sementara hanya 11,9% rumah tangga yang memiliki akses terhadap air yang aman untuk dikonsumsi. Air yang terkontaminasi oleh bakteri dapat menyebabkan penyakit seperti pneumonia (Munthe dkk., 2024).

BAB II METODOLOGI PENELITIAN

2.1 Sumber Data

Penelitian ini menggunakan data sekunder yang berasal dari Dinas Kesehatan yakni publikasi Profil Kesehatan Provinsi Sulawesi Selatan dan Badan Pusat Statistik Provinsi Sulawesi Selatan Tahun 2021. Unit penelitian yang digunakan adalah 24 Kabupaten/Kota di Provinsi Sulawesi Selatan. Variabel respon yang digunakan adalah jumlah kasus penyakit pneumonia pada balita sedangkan variabel prediktor yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Variabel Respon dan Prediktor

Variabel	Keterangan
y	Jumlah Kasus Penyakit Pneumonia pada Balita
x_1	Persentase Pemberian Asi Eksklusif
x_2	Persentase Balita yang Mendapatkan Vitamin A
x_3	Persentase Berat Badan Lahir Rendah (BBLR)
x_4	Persentase Penduduk Miskin
x_5	Persentase Balita (0-59 Bulan) yang Mendapatkan Imunisasi Campak
x_6	Persentase Rumah Tangga dengan Sumber Air Minum Bersih

Banyak faktor yang memengaruhi penyakit pneumonia pada balita, namun pada penelitian ini hanya digunakan 6 faktor yang memengaruhi penyakit ini dikarenakan keterbatasan dalam ketersediaan data.

2.2 Tahapan Penelitian

Metode analisis yang digunakan dalam penelitian ini adalah *Multivariate Adaptive Generalized Poisson Regression Splines* (MAGPRS). Pengolahan data dilakukan dengan menggunakan *software* R-Studio. Adapun tahapan analisis yang dilakukan dalam penelitian ini adalah:

1. Melakukan analisis deskriptif pada data.
2. Membuat *scatter plot* untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor.
3. Melakukan pengujian equidispersi menggunakan Persamaan (18) atau Persamaan (19).
4. Melakukan pemodelan MAGPRS dengan mengkombinasikan besarnya *Basis Function* (BF) yaitu 12, 18 dan 24, *Maximum Interaction* (MI) yaitu 1,2 dan 3, dan *Minimum Observation* (MO) yaitu 0,1,2, dan 3.
5. Menentukan model terbaik dari MAGPRS berdasarkan nilai GCV terkecil menggunakan Persamaan (34).
6. Melakukan pengujian simultan dan parsial model MAGPRS untuk mengetahui pengaruh variabel prediktor terhadap variabel respon menggunakan Persamaan (37) dan (38).

7. Melakukan interpretasi model MAGPRS.
8. Melakukan analisis kepentingan variabel prediktor terhadap variabel respon dengan menggunakan GCV.
9. Membuat kesimpulan dari hasil analisis