ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 15, Nomor 1, Tahun 2026, Halaman 67 - 76
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PEMODELAN *CLUSTERWISE LINEAR REGRESSION* UNTUK IDENTIFIKASI FAKTOR YANG MEMENGARUHI PREVALENSI STUNTING DI JAWA TENGAH

Berliana Ercha Pratiwi[1], Anton Saputro[2], Afifa Nur Mila[3], Moch. Abdul Mukid[4*], Masithoh Yessi Rochayani[5]
[1,2,3,4,5]Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail**: mamukid@live.undip.ac.id

DOI: 10.14710/j.gauss.15.1.67-76

Article Info:
Received: 2025-08-06
Accepted: 2026-04-06
Available Online: 2026-04-29

Keywords:
Stunting; *Sustainable Development Goals; Clusterwise Linear Regression;* Kesehatan

Abstract: Sustainable development in the Sustainable Development Goals (SDGs) emphasizes health as a key pillar, including in overcoming malnutrition that causes stunting. Central Java Province recorded a stunting prevalence rate of 20.7% in 2023, so it is necessary to analyze the factors that influence this condition. This study uses the Clusterwise Linear Regression (CLR) method to identify factors that contribute to the prevalence of stunting based on regional characteristics. The variables analyzed include the percentage of low birth weight babies (LBW), mothers who exclusively breastfeed less than six months, women who marry at an early age, households with proper sanitation, households with clean water sources, and households that have a Prosperous Family Card (KKS). The results showed that there were 3 optimal clusters. The coefficient of determination for each cluster was 99.52% for cluster 1, 99.76% for cluster 2, and 98.26% for cluster 3.

1. PENDAHULUAN

Pembangunan berkelanjutan yang tercermin dalam *Sustainable Development Goals* (SDGs) menempatkan kesehatan sebagai salah satu pilar utama, mengingat perannya yang krusial dalam mendukung produktivitas sosial dan ekonomi. Kesehatan merupakan kondisi tubuh yang sehat secara fisik, mental, spiritual, dan sosial, sehingga setiap individu mampu menjalani kehidupan yang bermanfaat dan produktif baik dalam aspek sosial maupun ekonomi. Menurut Kementerian Kesehatan Republik Indonesia (2016), salah satu ancaman kesehatan di masyarakat adalah malnutrisi yang ditunjukkan dengan stunting.

Stunting merupakan masalah gizi kronis yang terjadi pada balita yang didefinisikan berdasarkan tinggi badan menurut umur (TB/U) di mana tinggi badan balita lebih pendek dari yang seharusnya (Corrêa dkk., 2023). Balita dikatakan stunting apabila nilai perbandingan tinggi badan dengan nilai standar balita berada di bawah normal dari yang seharusnya (Wulandari dan Kurniawan, 2019). Stunting berdampak besar terhadap kualitas sumber daya manusia karena menghambat perkembangan otak, pertumbuhan fisik, dan meningkatkan risiko penyakit kronis. Prevalensi stunting yang tinggi mencerminkan masalah serius dalam pencapaian tujuan SDGs ke-2 (Tanpa Kelaparan) dan ke-3 (Kehidupan Sehat dan Sejahtera).

Provinsi Jawa Tengah memiliki angka prevalensi stunting sebesar 20.7% pada tahun 2023. Beberapa faktor utama yang memengaruhi prevalensi stunting adalah berat lahir rendah (BBLR), pemberian ASI eksklusif selama kurang dari enam bulan, pernikahan usia dini, sanitasi yang buruk, keterbatasan akses air bersih, dan kondisi ekonomi rendah. Pemerintah telah melakukan berbagai upaya, seperti Program Pemberian Makanan Tambahan (PMT) bagi ibu hamil dan balita, Intervensi Gizi Spesifik dan Sensitif, Pendampingan Keluarga Berisiko Stunting, dan Sanitasi Total Berbasis Masyarakat (STBM)

untuk menekan angka stunting. Namun, efektivitas kebijakan sangat bergantung pada faktor penyebab stunting di setiap wilayah.

Berdasarkan penelitian yang dilakukan oleh Hardinata dkk., (2023), menunjukkan bahwa metode regresi linear berganda konvensional belum optimal dalam menganalisis faktor yang memengaruhi stunting di Indonesia karena tidak mempertimbangkan subpopulasi dalam data. Wilayah Jawa Tengah juga memiliki keragaman karakteristik geografis dan sosial-ekonomi yang kontras, mulai dari wilayah pesisir utara, daerah pegunungan di bagian tengah, hingga kawasan industri perkotaan. Penggunaan regresi linier berganda konvensional akan mengakibatkan seluruh wilayah Jawa Tengah memiliki faktor penyebab stunting yang seragam meski memiliki kondisi geografis dan ekonomi yang berbeda. Pendekatan *Clusterwise Linear Regression* (CLR) dinilai lebih tepat karena mampu mengelompokkan data berdasarkan karakteristik wilayah dan menghasilkan model regresi yang lebih akurat untuk setiap *cluster*. Metode ini umumnya diterapkan pada data dengan variabel yang teridentifikasi memiliki pola mengelompok atau membentuk *cluster*. Perbedaan karakteristik antar kelompok tersebut memungkinkan adanya variasi variabel yang berpengaruh pada setiap *cluster*. Oleh karena itu, penerapan CLR diperlukan untuk mengidentifikasi faktor-faktor yang berpengaruh signifikan terhadap stunting beserta besar pengaruhnya di masing-masing *cluster*.

Secara konseptual *Clusterwise Linear Regression* (CLR) merupakan algoritma pembelajaran mesin yang bertujuan meningkatkan akurasi regresi linier berganda konvensional dengan cara membagi ruang pelatihan ke dalam beberapa subruang (Meylisa dkk., 2023). CLR menggabungkan dua pendekatan utama, yaitu teknik klasterisasi dan analisis regresi, dengan parameter yang diestimasi meliputi jumlah *cluster* serta koefisien regresi pada masing-masing *cluster* (Meylisa dkk., 2023). Berdasarkan hal tersebut, penelitian ini bertujuan untuk memodelkan prevalensi stunting di Jawa Tengah berdasarkan variabel-variabel terkait dengan pendekatan CLR, guna mengidentifikasi pola hubungan yang berbeda antar wilayah serta membantu perumusan kebijakan intervensi yang lebih tepat dan efektif untuk setiap wilayah.

2. TINJAUAN PUSTAKA

Menurut Kementerian Kesehatan Republik Indonesia (2016), stunting merupakan kondisi kekurangan gizi kronis karena asupan gizi kurang pada jangka waktu lama akibat pemberian makanan yang tidak tepat dengan kebutuhan gizi. Stunting bisa mulai terjadi saat janin masih terdapat di dalam kandungan dan umumnya terlihat saat anak berusia dua tahun. Stunting dapat menimbulkan dampak jangka pendek dan jangka panjang. Menurut Kementerian Kesehatan Republik Indonesia (2018), dampak jangka pendek yang ditimbulkan akibat stunting adalah peningkatan risiko terhadap gangguan kesehatan dan kematian, terhambatnya perkembangan motorik, kognitif, dan kemampuan verbal, serta peningkatan biaya kesehatan. Stunting juga menimbulkan dampak jangka panjang, yaitu perubahan postur tubuh menjadi tidak proporsional saat dewasa, peningkatan risiko penyakit tidak menular yang lain dan penyakit obesitas, penurunan kualitas kesehatan reproduksi, serta berkurangnya kemampuan dalam belajar dan performa akademik.

Prevalensi stunting menggambarkan seberapa besar masalah stunting di suatu wilayah pada waktu tertentu. Prevalensi stunting dipengaruhi oleh berbagai faktor, diantaranya berat lahir rendah pada bayi yang mencerminkan kondisi gizi ibu saat proses kehamilan, pemberian ASI eksklusif kurang dari enam bulan dapat berdampak pada asupan nutrisi bayi, perempuan yang melakukan pernikahan usia dini yang berkaitan dengan ketidaksiapan fisik dan ekonomi ibu. Selain itu, sanitasi layak dan sumber air bersih juga memengaruhi prevalensi stunting. Apabila sanitasi tidak layak dan sumber air bersih tidak tercukupi dapat

meningkatkan risiko infeksi yang berpengaruh terhadap penyerapan nutrisi dalam tubuh dan melemahkan sistem imun. Kepemilikan Kartu Keluarga Sejahtera juga ikut memengaruhi prevalensi stunting. Rendahnya kesejahteraan keluarga dapat membatasi akses terhadap kebutuhan gizi dan layanan kesehatan. Angka prevalensi stunting yang tinggi menggambarkan ketidakcukupan asupan gizi serta menjadi ancaman serius terhadap pembangunan kualitas sumber daya manusia di masa mendatang sehingga diperlukan metode analisis yang tepat, seperti *Clusterwise Linear Regression* (CLR) untuk memahami hubungan antar faktor yang memengaruhinya.

Menurut Bagirov dkk., (2017), *Clusterwise Linear Regression* (CLR) adalah metode yang mengkombinasikan antara analisis *clustering* dan analisis regresi. Model umum dari dari *Clusterwise Linear Regression* adalah:

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k \tag{1}$$

Keterangan:
$k = 1, 2, \ldots, \mathrm{K}$ (K adalah jumlah *cluster*).
$\boldsymbol{Y}_k =$ vektor variabel dependen pada *cluster* ke-k.
$\boldsymbol{X}_k =$ matriks variabel independent pada *cluster* ke-k.
$\boldsymbol{\beta}_k =$ matriks koefisien regresi pada *cluster* ke-k.
$\boldsymbol{\varepsilon}_k =$ vektor error pada *cluster* ke-k.

Estimasi parameter dari model Persamaan (1) akan dilakukan menggunakan metode *maximum likelihood.* Dalam model CLR diasumsikan y_i mengikuti distribusi campuran dari distribusi normal univariat bersyarat yang dapat dilihat pada Persamaan (2):

$$y_i \sim \sum_{k=1}^{K} \lambda_k (2\pi\sigma_k^2)^{1/2} \exp\left[\frac{-(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2}{2\sigma_k^2}\right] \tag{2}$$

Fungsi likelihood untuk pengambilan sampel acak berukuran n dari variabel bebas dimana λ_k adalah peluang prior yang menunjukkan bahwa pengamatan acak berasal dari *cluster* k. Ketentuan nilai λ_k dapat dilihat pada Persamaan (5).

$$L = \prod_{i=1}^{n}\left[\sum_{k=1}^{K} \lambda_k \left(2\pi\sigma_k^2\right)^{-1/2} \exp\left[\frac{-(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2}{2\sigma_k^2}\right]\right] \tag{3}$$

Kemudian jika fungsi Persamaan (3) dilakukan operasi logaritma natural maka akan diperoleh Persamaan (4),

$$\ln L = \sum_{i=1}^{n} ln\left[\sum_{k=1}^{K} \lambda_k \left(2\pi\sigma_k^2\right)^{-1/2} \exp\left[\frac{-(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2}{2\sigma_k^2}\right]\right] \tag{4}$$

Estimasi parameter dari model di Persamaan (1) didapatkan dengan memaksimalkan persamaan (4) dengan kendala,

$$\sum_{k=1}^{K} \lambda_k = 1; \lambda_k \geq 0, \sigma_k \geq 0, k = 1, 2, \ldots, K; \sigma_k^2 > 0 \tag{5}$$

Algoritma Ekspektasi-Maksimalisasi (EM) digunakan untuk estimasi kemungkinan maksimum dari λ_k, $\boldsymbol{\beta}_k, \sigma_k^2$, dan p_{ik}. Parameter yang diestimasi diperoleh dengan terlebih dahulu membuat fungsi Φ.

$$\Phi = \sum_{i=1}^{n} ln\left[\sum_{k=1}^{K} \lambda_k f_{ik}(y_i \mid X_{ij}, \sigma_k^2, \boldsymbol{\beta}_k)\right] - \mu\left(\sum_{k=1}^{K} \lambda_k - 1\right) \tag{6}$$

Dimana μ adalah Lagrange Multiplier. Fungsi pada Persamaan (6) dimaksimalkan dengan menetapkan turunan parsial orde pertama dari fungsi ln likelihood terhadap parameter $\lambda_k, \boldsymbol{\beta}_k, \sigma_k^2$ menjadi nol,

$$\frac{\partial\,\Phi}{\partial\,\lambda_{\mathrm{k}}}=\sum_{i=1}^{n}\frac{1}{\sum_k \lambda_{\mathrm{k}} f_{ik}(*)}f_{ik}(*)-\mu=0, \tag{7}$$

$$\frac{\partial\,\Phi}{\partial\,\sigma_{\mathrm{k}}^2}=\sum_{i=1}^{n}\frac{1}{\sum_k \lambda_{\mathrm{k}} f_{ik}(*)}\lambda_{\mathrm{k}}\frac{\partial f_{ik}(*)}{\partial\,\sigma_{\mathrm{k}}^2}=0, \tag{8}$$

$$\frac{\partial\,\Phi}{\partial\boldsymbol{\beta}_{\mathrm{k}}}=\sum_{i=1}^{n}\frac{1}{\sum_k \lambda_{\mathrm{k}} f_{ik}(*)}\lambda_{\mathrm{k}}\frac{\partial f_{ik}(*)}{\partial\,\boldsymbol{\beta}_{\mathrm{k}}}=\mathbf{0}. \tag{9}$$

Dimana $f_{ik}(*)$ adalah $f_{ik}(y_i \mid X_{ij}, \sigma_k^2, \boldsymbol{\beta}_k)$.

Diketahui nilai awal untuk λ_k, σ_k^2, dan $\boldsymbol{\beta}_k$ untuk k=1,2,…,K. Maka berikut adalah tahapan iteratif algoritma EM untuk melakukan estimasi parameter CLR Pada iterasi ke-(t+1) dari langkah E, pembaruan untuk kuantitas diperoleh sebagai berikut:

$$p_{ik}^{(t+1)}=\frac{\lambda_{\mathrm{k}} f_{ik}\left(y_i|X_{ij},\widehat{\sigma}_{\mathrm{k}}^{2(t)},\boldsymbol{\beta}_k^{(t)}\right)}{\sum_{k=1}^{K}\lambda_{\mathrm{k}}\, f_{ik}\left(y_i|X_{ij},\widehat{\sigma}_{\mathrm{k}}^{2(t)},\boldsymbol{\beta}_k^{(t)}\right)} \tag{10}$$

Dimana t dinotasikan sebagai indeks iterasi ke-t. Menggunakan p_{ik} dari Step E, Step M terdiri dari perhitungan iteratif sebagai berikut:

$$\lambda_k^{(t+1)}=\frac{\sum_{i=1}^{n} p_{ik}^{(t+1)}}{n} \tag{11}$$

$$\boldsymbol{\beta}_k^{(t+1)}=\left(\sum_{i=1}^{n} p_{ik}^{(t+1)}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{n} p_{ik}^{(t+1)}\mathbf{x}_i y_i\right) \tag{12}$$

$$\sigma_k^{2(t+1)}=\frac{\sum_{i=1}^{n} p_{ik}^{(t+1)}\left(\mathrm{y_i}-\mathbf{x}_{\mathrm{i}}'\boldsymbol{\beta}_k^{(t+1)}\right)^2}{\sum_{i=1}^{n} p_{ik}^{(t+1)}} \tag{13}$$

Setelah parameter diperoleh, alokasikan setiap observasi ke setiap *cluster* dengan menerapkan aturan: alokasikan i ke *cluster* k jika dan hanya jika $\hat{p}_{ik} > \hat{p}_{il}$ untuk semua $l \neq k = l \ldots K$. Proses iterasi berhenti ketika proses konvergen atau iterasi maksikum telah tercapai.

Pengujian hipotesis pada *Clusterwise Linear Regression* (CLR) dilakukan dengan menggunakan metode *maximum likelihood* (MLE). Penduga *maximum likelihood* pada kondisi keteraturan umum secara asimtotik memenuhi kondisi seperti pada Persamaan (14).

$$\sqrt{\mathrm{n}}\left(\widehat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k\right)\xrightarrow{\mathrm{d}}\mathrm{N}\left(\mathbf{0},\lim\left(\frac{\mathrm{R}(\boldsymbol{\beta}_k)}{\mathrm{n}}\right)^{-1}\right) \tag{14}$$

dengan

$$\mathrm{R}(\boldsymbol{\beta}_{\mathrm{k}})=-\mathrm{E}\left[\frac{\partial^2\boldsymbol{\Phi}}{\partial\boldsymbol{\beta}_{\mathrm{k}}\,\partial\boldsymbol{\beta}_{\mathrm{k}}'}\right] \tag{15}$$

$R(\boldsymbol{\beta})$ adalah sebuah matriks informasi. Judge., *et al*. (1985) menjelaskan bahwa mengubah $\lim\left(\frac{R(\boldsymbol{\beta})}{n}\right)$ dengan pendugaan yang konsisten tidak mengubah distribusi asimtotik dan statistik uji maupun selang kepercayaan untuk $\widehat{\boldsymbol{\beta}}_k$. Pendugaan konsisten yang digunakan adalah:

$$\mathbf{F}=\frac{1}{\mathrm{n}}\left[\sum_{\mathrm{i}=1}^{\mathrm{n}}\left[\frac{\partial\boldsymbol{\Phi}}{\partial\boldsymbol{\beta}_{\mathrm{k}}}\right]\left[\frac{\partial\boldsymbol{\Phi}}{\partial\boldsymbol{\beta}_{\mathrm{k}}}\right]'\right]_{\boldsymbol{\beta}_{\mathrm{k}}=\widehat{\boldsymbol{\beta}}_{\mathrm{k}}}, \tag{16}$$

dan varian asimtotik dari $\widehat{\boldsymbol{\beta}}_k$ bisa didefinisikan sebagai elemen diagonal utama $\mathbf{F}^{-1}$ (matriks varians kovarian asimtotik). Dari persamaan (14) - (16), dapat disimpulkan bahwa interval kepercayaan asimtotik $(1-\alpha)100\%$ untuk β_{jk} dapat dilihat pada Persamaan (17).

$$\left(\hat{\beta}_{jk} - Z_{\frac{\alpha}{2}}\sqrt{f^{-1}_{j_k j_k}}, \hat{\beta}_{jk} + Z_{\frac{\alpha}{2}}\sqrt{f^{-1}_{j_k j_k}}\right), \tag{17}$$

untuk j = 1,2,…,J, dengan $Z_{\alpha/2}$ adalah nilai tengah dari distribusi normal dengan rata-rata nol dan varians satu dan merupakan pendugaan asimtotik dari varians $\hat{\beta}_{jk}$. Sehingga nilai dari Z_{hitung} dapat lihat pada dilihat pada Persamaan (18).

$$Z_{hitung} = \frac{\hat{\beta}_{jk} - \beta_{jk}}{\sqrt{f^{-1}_{j_k j_k}}} \tag{18}$$

Berikut adalah pengujian hipotesis dari *Clusterwise Linear Regression* (CLR).
Hipotesis:
H_0: $\beta_{jk} = 0$
H_1: $\beta_{jk} \neq 0$ dengan j = 1, 2, ..., J dan k =1, 2, ..., K
Taraf Signifikansi: 5%
Statistik Uji: Z_{hitung}
Daerah Kritis:
Tolak H_0 jika nilai $|Z| > Z_{tabel}$ $(Z_{\alpha/2})$ atau p-value < α.

Pada metode CLR, salah satu ukuran evaluasi model yang digunakan adalah koefisien determinasi (R^2). Menurut Hair dkk., (2010), koefisien determinasi R^2 merupakan proporsi variansi dari variabel dependen yang dijelaskan oleh variabel independen dalam model regresi. Dalam konteks CLR, R^2 dihitung untuk setiap *cluster* dan juga dapat digunakan sebagai agregat (rata-rata tertimbang) untuk mengevaluasi keseluruhan performa model. Hal ini penting untuk menilai apakah pemisahan *cluster* memang menghasilkan peningkatan kualitas prediksi dibandingkan regresi linier tunggal pada seluruh data.

Koefisien determinasi memiliki interval nilai antara nol sampai satu. Nilai yang kecil menandakan kemampuan variabel-variabel bebas dalam menjelaskan variasi variabel terikat amat terbatas. Sedangkan nilai yang mendekati satu menandakan variabel-variabel bebas memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variasi variabel terikat (Ghozali, 2018). Kategori nilai koefisien determinasi dapat dilihat pada Tabel 1 (Chin, 1998).

Tabel 1. Kategori Koefisien Determinasi

Nilai	Kategori
$0{,}67 > R^2 \geq 1$	Kuat
$0{,}33 > R^2 \geq 0{,}67$	Moderat
$0{,}19 > R^2 \geq 0{,}33$	Lemah
$0 > R^2 \geq 0{,}19$	Sangat Lemah

Pendekatan dalam mengidentifikasi jumlah *cluster* K yang sesuai dengan prosedur pengelompokan campuran melibatkan *Akaike's Information Criterion* (AIC) yang didefinisikan sebagai pada Persamaan (19).

$$AIC(K) = -2\ln[\max L(K)] + 2n(K) \tag{19}$$

dengan $n(K)$ adalah jumlah parameter efektif yang diduga dalam *Clusterwise Linear Regression* seperti pada Persamaan (20)*:*

$$n(K) = JK + 2K - 1 \tag{20}$$

dengan J adalah jumlah variabel bebas dan K adalah jumlah *cluster*.

3. METODE PENELITIAN

Data yang digunakan merupakan data persentase faktor prevalensi stunting yang diperoleh dari Badan pusat Statistik dan Dinas Kesehatan Tahun 2023, dengan objek pengamatan sebanyak 35 kabupaten/kota di Provinsi Jawa Tengah.

Variabel yang digunakan dalam penelitian ini adalah persentase prevalensi stunting (Y), persentase Bayi dengan Berat Lahir Rendah (BBLR) (X_1), persentase ibu yang memberikan ASI eksklusif selama kurang dari enam bulan (X_2), persentase perempuan yang melakukan pernikahan usia dini (X_3), persentase rumah tangga dengan sanitasi layak (X_4), persentase rumah tangga dengan sumber air bersih (X_5), persentase rumah tangga yang memiliki Kartu Keluarga Sejahtera (X_6).

Tahapan analisis data dengan metode *Clusterwise Linier Regression* adalah sebagai berikut:

1. Input data variabel terikat (Y) dan variabel bebas (X).
2. Melakukan analisis statistika deskriptif pada data yang telah diinput.
3. Pemodelan *Clusterwise Linear Regression* dengan tahap awal yaitu menentukan jumlah *cluster* (K).
4. Ambil nilai awal untuk estimasi $\hat{\lambda}_k, \widehat{\boldsymbol{\beta}}_{jk}$ dan σ_k^2 .
5. Hitung nilai dugaan $\hat{p}_{ik}$ dengan menggunakan Persamaan (10).
6. Setelah didapatkan $\hat{p}_{ik}$ sementara, hitung kembali nilai $\hat{\lambda}_k$ dengan menggunakan Persamaan (11).
7. Menghitung nilai $\widehat{\boldsymbol{\beta}}_{jk}$ menggunakan kuadrat terkecil terboboti dengan menggunakan Persamaan (12).
8. Menghitung kembali nilai σ_k^2 dengan menggunakan Persamaan (13).
9. Menghitung nilai log likelihood sesuai dengan Persamaan (4).
10. Langkah 3 sampai dengan 9 dilakukan sampai konvergen.
11. Menentukan jumlah *cluster* yang optimal yang ditunjukan dengan nilai *Akaike Information Criteria* (AIC) yang minimum.
12. Menentukan anggota setiap *cluster* dan pendugaan model regresi di setiap *cluster*.
13. Melakukan pengujian hipotesis dan evaluasi model yang terbentuk.
14. Profilisasi *cluster*.

4. HASIL DAN PEMBAHASAN

Analisis deskriptif digunakan untuk mendapatkan hasil mengenai jumlah data setiap variabel, nilai minimum, maksimum, rata-rata, dan standar deviasi. Hasil analisis deskriptif setiap variabel data ditampilkan pada Tabel 2 berikut:

Tabel 2. Statistika Deskriptif

Variabel	N	Minimum	Maksimum	Mean	Std.deviation
Y	35	1.14	18.24	10.11	4.56
X_1	35	0.87	14.57	5.29	3.12
X_2	35	24.54	82.95	63.66	14.83
X_3	35	16.52	57.71	35.86	11.40
X_4	35	46.09	98.15	85.31	11.89
X_5	35	66.77	98.36	82.28	8.34
X_6	35	2.56	35.59	15.06	7.45

Tahap selanjutnya adalah penentuan jumlah *cluster* optimal. Jumlah *cluster* optimal ditentukan berdasarkan nilai *Akaike Information Criteria* (AIC) terkecil pada jumlah *cluster* yang terbentuk. Penentuan jumlah *cluster* dilakukan dengan rumus berikut:

$$n \geq KJ$$
$$35 \geq K \times 6$$
$$K \leq 5.83$$

Berdasarkan perhitungan tersebut, diperoleh hasil bahwa jumlah *cluster* yang terbentuk adalah K = 2, 3, 4, dan 5. Guna menentukan jumlah *cluster* optimal, diperlukan perhitungan nilai VIF untuk setiap jumlah *cluster*, baik K = 2, 3, 4, maupun 5 dengan hasil dapat dilihat pada Tabel 3.

Tabel 3. Nilai AIC dari Setiap K

K	AIC
2	193.6741
3	150.5967
4	Tidak Dapat Dicari
5	Tidak Dapat Dicari

Berdasarkan Tabel 3, nilai AIC pada K = 4 dan 5 tidak dapat dicari karena terdapat *cluster* yang tidak memiliki anggota sehingga model pada *cluster* tersebut tidak valid. Oleh karena itu, jumlah *cluster* optimal yang terbentuk adalah K = 3 dengan nilai AIC sebesar 150.5967. Pemodelan *Clusterwise Linear Regression* dengan jumlah *cluster* sebanyak 3 diterapkan untuk mengelompokkan kabupaten-kabupaten di Provinsi Jawa Tengah berdasarkan faktor-faktor yang memengaruhi prevalensi stunting.

Cluster 1 memiliki ukuran *cluster* sebanyak 13 kabupaten/kota yang meliputi Banyumas, Batang, Blora, Kebumen, Kota Magelang, Kudus, Magelang, Pati, Pekalongan, Rembang, Semarang, Temanggung, Wonosobo. Model yang terbentuk adalah sebagai berikut.

$\hat{Y} = 6.0081964 + 0.0022879X_1 + 0.1094368X_2 + 0.0085846X_3 - 0.2102204X_4 + 0.1149641X_5 + 0.4082503X_3$

Hipotesis dari uji signifikansi parameter individual dari model yang terbentuk pada *cluster* 1 adalah:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$, dengan j = 1,2,…,J

Taraf signifikansi (α) yang digunakan sebesar 5%, dengan daerah kritis yaitu tolak H_0 jika p-value < α. Diperoleh hasil bahwa variabel X_2, X_4, X_5, dan X_6 masing-masing memiliki nilai p-value sebesar < 2.2e-16, < 2.2e-16, 1.493e-13, < 2.2e-16 sehingga H_0 ditolak untuk variabel X_2, X_4, X_5, dan X_6. Dengan demikian, keempat variabel tersebut dapat disimpulkan berpengaruh signifikan terhadap model. Namun, variabel X_1 dan X_3 memiliki nilai p-value sebesar 0.9565 dan 0.4021 sehingga H_0 gagal ditolak untuk variabel X_1 dan X_3, akibatnya variabel tersebut disimpulkan tidak berpengaruh signifikan terhadap model.

Koefisien determinasi pada *cluster* 1 menunjukkan bahwa nilai R^2 sebesar 0.9952 atau 99.52%. Hal tersebut menjelaskan bahwa pada *cluster* 1, variabel X_1, X_2, X_3, X_4, X_5, dan X_6 memengaruhi variabel Y sebesar 99.52%. Sementara itu, 0.48% dipengaruhi oleh variabel lain.

Cluster 2 memiliki ukuran *cluster* sebanyak 11 kabupaten/kota yang meliputi Boyolali, Cilacap, Demak, Jepara, Kendal, Kota Semarang, Kota Tegal, Sragen, Sukoharjo, Tegal, Wonogiri. Model yang terbentuk adalah sebagai berikut.

$\hat{Y} = -5.883959 + 0.598483X_1 - 0.203445X_2 + 0.850959X_3 + 0.358142X_4 - 0.270691X_5 - 0.948279X_6$

Hipotesis dari uji signifikansi parameter individual dari model yang terbentuk pada *cluster* 2 adalah:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$, dengan j = 1,2,…,J

Taraf signifikansi (α) yang digunakan sebesar 5%, dengan daerah kritis yaitu tolak H_0 jika p-value < α. Diperoleh hasil bahwa semua variabel, yaitu variabel X_1, X_2, X_3, X_4, X_5, dan X_6 memiliki nilai p-value sebesar < 2e-16 sehingga disimpulkan bahwa H_0 ditolak untuk semua variabel dan semua variabel berpengaruh signifikan terhadap model.

Koefisien determinasi pada *cluster* 2 menunjukkan bahwa nilai R^2 sebesar 0.9976 atau 99.76%. Hal tersebut mejelaskan bahwa pada *cluster* 2, variabel X_1, X_2, X_3, X_4, X_5, dan X_6 memengaruhi variabel Y sebesar 99.76%. Sementara itu, 0.24% dipengaruhi oleh variabel lain.

Cluster 3 memiliki ukuran *cluster* sebanyak 11 kabupaten/kota yang meliputi Banjarnegara, Brebes, Grobogan, Karanganyar, Klaten, Kota Pekalongan, Kota Salatiga, Kota Surakarta, Pemalang, Purbalingga, Purworejo. Model yang terbentuk adalah sebagai berikut.

$$\hat{Y} = 70.192562 - 0.663177X_1 + 0.023549X_2 + 0.065655X_3 - 0.113113X_4 - 0.596833X_5 - 0.158488X_6$$

Hipotesis dari uji signifikansi parameter individual dari model yang terbentuk pada *cluster* 3 adalah:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$, dengan j = 1,2,…,J

Taraf signifikansi (α) yang digunakan sebesar 5%, dengan daerah kritis yaitu tolak H_0 jika p-value < α. Diperoleh hasil bahwa variabel X_1, X_3, X_4, X_5, dan X_6 masing-masing memiliki nilai p-value sebesar 7.221e-08, 0.001827, 4.984e-08, < 2.2e-16, dan 8.696e-06 sehingga H_0 ditolak untuk variabel X_1, X_3, X_4, X_5, dan X_6. Dengan demikian, kelima variabel tersebut dapat disimpulkan berpengaruh signifikan terhadap model. Namun, variabel X_2 memiliki nilai p-value sebesar 0.193290 sehingga H_0 gagal ditolak untuk variabel X_2, akibatnya variabel tersebut disimpulkan tidak berpengaruh signifikan terhadap model.

Koefisien determinasi pada *cluster* 3 menunjukkan bahwa nilai R^2 sebesar 0.9826 atau 98.26%. Hal tersebut menjelaskan bahwa pada *cluster* 3, variabel X_1, X_3, X_4, X_5, dan X_6 memengaruhi variabel Y sebesar 98.26%. Sementara itu, 1.74% dipengaruhi oleh variabel lain.

Berdasarkan hasil setiap *cluster,* diperoleh variabel yang signifikan yang berbeda pada setiap *cluster* sehingga dapat dilakukan optimalisasi terkait variabel yang bersangkutan untuk memperbaiki angka prevalensi stunting di Jawa Tengah sesuai dengan kondisi setiap *cluster.* Koefisien determinasi dari semua *cluster* menunjukkan angka yang kuat yang menunjukkan bahwa variabel bebas mampu menjelaskan variabel terikat dengan baik.

5. KESIMPULAN

Pengelompokan kabupaten/kota di Provinsi Jawa Tengah dilakukan menggunakan metode *Clusterwise Linear Regression* untuk menganalisis faktor-faktor yang memengaruhi prevalensi stunting. Berdasarkan nilai minimum *Akaike Information Criterion* (AIC), yaitu sebesar 150.5967, diperoleh tiga *cluster* optimal. *Cluster* 1 terdiri dari 13 kabupaten/kota yang meliputi Banyumas, Batang, Blora, Kebumen, Kota Magelang, Kudus, Magelang, Pati, Pekalongan, Rembang, Semarang, Temanggung, Wonosobo. *Cluster* 2 terdiri dari 11 kabupaten/kota yang meliputi Boyolali, Cilacap, Demak, Jepara, Kendal, Kota Semarang, Kota Tegal, Sragen, Sukoharjo, Tegal, Wonogiri. *Cluster* 3 terdiri dari 11 kabupaten/kota yang meliputi Banjarnegara, Brebes, Grobogan, Karanganyar, Klaten, Kota Pekalongan, Kota Salatiga, Kota Surakarta, Pemalang, Purbalingga, Purworejo. Faktor-faktor yang memengaruhi prevalensi stunting di setiap *cluster* menunjukkan bahwa *cluster* 1 dipengaruhi oleh X_2, X_4, X_5, dan X_6 dengan nilai koefisien determinasi R^2 sebesar 99.52%. *Cluster* 2 dipengaruhi oleh X_1, X_2, X_3, X_4, X_5, dan X_6 dengan nilai koefisien determinasi R^2 sebesar

99.76%. *cluster* 3 dipengaruhi oleh X_1, X_3, X_4, X_5, dan X_6 dengan nilai koefisien determinasi R^2 sebesar 98.26%. Berdasarkan hasil setiap *cluster,* dapat dilakukan optimalisasi yang lebih terarah sesuai kondisi dan variabel yang signifikan untuk memperbaiki angka prevalensi stunting di Jawa Tengah.

DAFTAR PUSTAKA

Badan Pusat Statistik Provinsi Jawa Tengah. (21 Februari 2024). *Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2023*. Diakses pada 14 Juni 2025, dari https://jateng.bps.go.id/id/statistics-table/3/VGtGTU5qbDFlQzl1VWxCTVNWZElXbWRhWkUwMFVUMDkjMw==/persentase-rumah-tangga-yang-memiliki-akses-terhadap-sanitasi-layak-menurut-kabupaten-kota-di-provinsi-jawa-tengah--2023.html?year=2023

Badan Pusat Statistik Provinsi Jawa Tengah. (22 Juli 2024). *Jumlah Bayi Lahir, Bayi Berat Badan Lahir Rendah (BBLR), dan Bergizi Kurang Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2023*. Diakses pada 14 Juni 2025, dari https://jateng.bps.go.id/id/statistics-table/2/Mzc4IzI=/jumlah-bayi-lahir-bayi-berat-badan-lahir-rendah-bblr-bblr-dirujuk-dan-bergizi-buruk-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html

Badan Pusat Statistik Provinsi Jawa Tengah. (23 Agustus 2024). *Persentase Penduduk Wanita Berumur 10 Tahun ke Atas yang Pernah Kawin menurut Kabupaten/Kota dan Umur Perkawinan Pertama di Provinsi Jawa Tengah, 2023*. Diakses pada 14 Juni 2025, dari https://jateng.bps.go.id/id/statistics-table/2/MTE0OCMy/persentase-penduduk-wanita-berumur-10-tahun-ke-atas-yang-pernah-kawin-menurut-kabupaten-kota-dan-umur-perkawinan-pertama-di-provinsi-jawa-tengah.html

Badan Pusat Statistik Provinsi Jawa Tengah. (23 Juli 2024). *Persentase Rumah Tangga menurut Kabupaten/Kota, Sumber Air Minum Bersih, dan Akses Air Minum Layak, 2023*. Diakses pada 14 Juni 2025, dari https://jateng.bps.go.id/id/statistics-table/2/MTU2MCMy/persentase-rumah-tangga-menurut-kabupaten-kota--sumber-air-minum-bersih--dan-akses-air-minum-layak--persen-.html

Badan Pusat Statistik Provinsi Jawa Tengah. (26 Agustus 2024). *Persentase Rumah Tangga yang Menerima Kartu Perlindungan Sosial (KPS)/Kartu Keluarga Sejahtera (KKS) menurut Kabupaten/Kota, 2023*. Diakses pada 14 Juni 2025, dari https://jateng.bps.go.id/id/statistics-table/2/MTYyNSMy/persentase-rumah-tangga-yang-menerima-kartu-perlindungan-sosial--kps--kartu-keluarga-sejahtera--kks--menurut-kabupaten-kota--persen-.html

Bagirov, A. M., Mahmood, A., & Barton, A. (2017). Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach. *Atmospheric research*, *188*, 20-29.

Chin, W. W. (1998) The partial least squares approach for structural equation modeling. *Modern Methods for Business Research*.

Corrêa, E. M., Gallo, C. D. O., Antunes, J. L. F., & Jaime, P. C. (2023). The tendency of stunting among children under five in the Northern Region of Brazil, according to the Food and Nutrition Surveillance System, 2008-2017. *Jornal de Pediatria*, *99*, 120-126.

Erni, L. (2024, July 4). *Jumlah Bayi Mendapat ASI Eksklusif*. Retrieved from Portal Data Jawa Tengah: https://data.jatengprov.go.id/dataset/d37-sdjt-jumlah-bayi-mendapat-asi-eksklusif

Erni, L. (2024, July 4). *Prevalensi Stunting*. Retrieved from Portal Data Jawa Tengah: https://data.jatengprov.go.id/dataset/42f-sdjt-prevalensi-stunting

Ghozali, I. (2018). Aplikasi Analisis Multivariate Dengan Pogram IBM SPSS Edisi Sembilan. Semarang: Badan Penerbit Universitas Diponegoro.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Education.

Hardinata, R., Oktaviana, L., Husain, F. F., Putri, S., & Kartiasih, F. (2023, October). Analisis faktor-faktor yang memengaruhi stunting di Indonesia tahun 2021. In *Seminar Nasional Official Statistics* (Vol. 2023, No. 1, pp. 817-826).

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T. C. (1991). *The theory and practice of econometrics*. John wiley & sons.

Kementerian Kesehatan Republik Indonesia. (2016). Situasi Balita Pendek. *ACM SIGAPL APL Quote Quad*, *29*(2), 63–76. https://doi.org/10.1145/379277.312726

Kementerian Kesehatan Republik Indonesia. (2018). Situasi Balita Pendek (Stunting) Di Indonesia. *Jakarta: Pusat Data dan Informasi Kementerian Kesehatan RI*.

Meylisah, E., Rini, D. S., Fransiska, H., Agwil. W. & Sartono. B. (2023). Modeling Clusterwise Linear Regression on Poverty Rate in Indonesia. BAREKENG: Journal of Mathematics and Its Applications.

Rahmawati, S. D., Sudarno, & Wilandari, Y. (2025). Cluster Optimization Using Cluster wise Linear Regression Method with Akaike Information Criterion to Identify Factors Influencing the Number of DHF Cases. *International Journal of Research Publication and Reviews,*.

Wulandari, S., & Kurniawan, R. (2019). Pengelompokan Kabupaten/Kota Di Jawa Timur Berdasarkan Kasus Stunting Balita Menggunakan Algoritme Fuzzy Particle Swarm Optimization-Fuzzy C-Means. *Jurnal Statistika Universitas Muhammadiyah Semarang*, *7*(1).