Jurnal EurekaMatika

Journal homepage: https://ejournal.upi.edu/index.php/JEM
doi: https://doi.org/10.17509/jem.v13i2.90359

Implementasi Metode Regresi dengan Analisis Komponen Utama pada Kasus Diare Balita di Jawa Tengah

*Moch. Anjas Aprihartha *, Mohamad Rijal Arfani , dan Indah Putianik*

PJJ Informatika, Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia

*Correspondence: anjas.aprihartha@dsn.dinus.ac.id

A B S T R A K

Multikolinearitas merupakan kondisi yang terjadi pada model regresi ketika adanya korelasi tinggi antara variabel independen. Untuk mengatasi masalah multikolinearitas maka digunakan teknik regresi komponen utama. Melalui pendekatan tersebut, faktor-faktor risiko kesehatan dapat diidentifikasi secara lebih akurat. Provinsi Jawa Tengah menduduki urutan ke 11 dari 34 provinsi dengan proporsi kasus diare pada balita ialah 11,1%. Angka proporsi diare balita yang cukup tinggi mempertegas urgensi pemodelan faktor risikonya, sehingga tujuan penelitian ini adalah menerapkan metode regresi komponen utama pada kasus diare balita di Jawa Tengah. Model regresi menghasilkan nilai $R^2 = 96{,}2\%$ yang dapat diartikan sebagai model sangat baik dalam menggambarkan hubungan penyebab diare pada balita terhadap jumlah balita menderita diare di kabupaten/ kota di Jawa Tengah. Adapun faktor-faktor yang berpengaruh signifikan diantaranya pengguna sanitasi layak, balita yang kekurangan gizi, posyandu aktif, dan balita yang diimunisasi dasar lengkap.

© 2025 Kantor Jurnal dan Publikasi UPI

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima 18 September 2025
Direvisi 21 Oktober 2025
Disetujui 1 November 2025
Tersedia online 2 November 2025
Dipublikasikan 2 November 2025

Kata Kunci:
Analisis komponen utama,
Balita,
Diare,
Multikolinearitas,
Regresi.

A B S T R A C T

Multicollinearity is a condition that occurs in a regression model when there is a high correlation between independent variables. To address the problem of multicollinearity, the principal component regression technique is used. Through this approach, the causes of health problems can be identified more accurately. Central Java Province ranks 11th out of 34 provinces with a proportion of diarrhea cases in toddlers of 11.1%. The relatively high proportion of diarrhea in toddlers emphasizes the urgency of modeling its risk factors. Therefore, the purpose of this study was to apply the principal component regression method to cases of diarrhea in toddlers in Central Java. The regression model produced an R^2 value of 96.2%, which is interpreted as a very good model in identifying the relationship between the causes of diarrhea in toddlers and the number of toddlers suffering from diarrhea in districts/cities in Central Java. The factors that have a significant influence include the use of proper sanitation, malnourished toddlers, active integrated health posts, and fully immunized toddlers.

© 2025 Kantor Jurnal dan Publikasi UPI

Keywords:
Analysis principal component,
Toddlers,
Diarrhea,
Multicollinearity,
Regression.

1. PENDAHULUAN

Secara umum, analisis regresi dapat dianggap sebagai proses penyesuaian model dengan data eksperimen. Menurut teori analisis regresi, terdapat jenis model regresi linier yang merupakan bentuk regresi spesifik, dan fungsi prediktor linier dapat digunakan untuk memodelkan data, dan parameter keluaran diestimasi dari data tersebut (Li *et al.*, 2024). Dalam perkembangannya, model regresi linier memiliki berbagai turunan dan metode khusus yang dapat digunakan untuk mengatasi permasalahan tertentu pada data. Salah satu diantaranya adalah analisis regresi komponen utama. Teknik regresi komponen utama merupakan turunan dari teknik regresi linear berganda yang diintegrasikan dengan Analisis Komponen Utama (AKU), dengan AKU ditetapkan dalam tahapan analisis. AKU merupakan teknik mereduksi dimensi data dengan melibatkan transformasi orthogonal untuk menciptakan komponon utama yang menggantikan variabel aslinya (Kristjanpoller & Minutolo, 2018). Metode ini dikembangkan sebagai cara mengatasi masalah multikolinearitas dalam analisis regresi linear berganda (Pendi, 2021).

Multikolinearitas terjadi ketika adanya korelasi tinggi antara variabel independen pada model regresi linear. Banyaknya variabel independen menjadi faktor memicu adanya multikolinearitas pada model regresi. Apabila tidak diatasi maka variabel independen yang seharusnya signifikan dapat terlihat tidak signifikan secara statistik, dengan tanda koefisien regresi yang menyimpang dari kondisi nyata. Ketidakstabilan estimasi koefisien ini menyulitkan prediksi variabel dependen dan menurunkan akurasi peramalan (Marcus *et al*., 2012).

Multikolinearitas dapat muncul apabila variabel-variabel yang digunakan untuk menduga penyebab diare pada balita memiliki tingkat keterkaitan yang tinggi satu sama lain. Pada penelitian Jannah & Susanti (2023), penyebab terjadinya diare pada balita dipengaruhi oleh pemberian ASI eksklusif, pemberian imunisasi dasar lengkap, akses air minum layak, dan akses sanitasi layak. Penelitian Widyaningrum *et al*. (2021), jumlah kasus diare di Jawa Timur disebabkan oleh balita kurang gizi, jumlah penduduk miskin, jumlah posyandu aktif, dan pemberian vitamin A. Penelitian oleh Putri *et al.* (2023), kasus diare pada balita di Jawa Timur diduga oleh faktor-faktor yang diantranya pemberian ASI eksklusif, jumlah rumah tangga dengan sanitasi layak, pemberian vitamin A, pemberian imunisasi dasar lengkap, sarana air minum yang memenuhi syarat, dan kepadatan penduduk. Penelitian dari Khoirunnisa *et al*. (2019), faktor-faktor yang diduga memengaruhi diare pada balita di Bandung di antaranya jumlah rumah sehat, jumlah jamban sehat, akses air bersih, akses air layak minum, perilaku hidup bersih dan sehat, pemberian vitamin A, status gizi buruk pada balita, dan kepadatan penduduk.

Diare menjadi polemik utama pada kematian balita di berbagai provinsi Widyaningrum *et al*. (2021). Provinsi Jawa Tengah menduduki urutan ke 11 dari 34 provinsi dengan proporsi kasus diare 7,2%, sedangkan proporsi kasus diare pada balita ialah 11,1% (Kemenses RI, 2019). Angka kejadian penyakit diare mayoritas dialami oleh balita, hal ini terjadi karena pada fase balita masih memiliki sistem imun yang belum cukup kuat dalam melawan gangguan mikroorganisme dari luar yang menjadi penyebab balita rentan terhadap serangan penyakit. Hal ini dipaparkan pada buku berjudul *Metadata penelitian Badan Litbangkes Tahun 2019* yang diterbitkan oleh Badan Penelitian Dan Pengembangan Kesehatan.

Penelitian ini bertujuan untuk memodelkan dan mengamati faktor-faktor yang mempengaruhi kasus diare balita di Jawa Tengah. Perbedaan penelitian ini dengan penelitian-penelitian sebelumnya terletak pada pemilihan variabel yang digunakan. Dalam penelitian ini, variabel-variabel yang digunakan merupakan hasil sintesis dari beberapa penelitian yang telah diidentifikasi sebelumnya, yang menjadi penyebab terhadap kejadian diare. Namun

demikian, terdapat beberapa variabel yang tidak dimasukkan dalam penelitian ini karena keterbatasan ketersediaan data, meskipun telah dilakukan upaya pencarian dari berbagai sumber. Adapun faktor-faktor yang diduga mempengaruhi kejadian diare balita di Jawa Tengah di antaranya jumlah balita yang diberikan ASI eksklusif , jumlah balita yang diberikan vitamin A, kepadatan penduduk, jumlah pengguna sanitasi layak, jumlah pengelolaan air minum dan makanan, jumlah balita yang kekurangan gizi, jumlah penduduk miskin, jumlah posyandu aktif, jumlah balita yang diimunisasi dasar lengkap (Widyaningrum *et al*., 2021; Putri *et al.*, 2023; Khoirunnisa *et al*., 2019). Dalam menghindari asumsi multikolinearitas tidak terpenuhi yang disebabkan jumlah dimensi yang besar maka digunakan teknik analisis regresi linear yang diintegrasikan dengan analisis komponen utama (AKU).

2. METODE

2.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam analisis ini merupakan data sekunder dari data Dinas Kesehatan Provinsi Jawa Tengah Tahun 2023 pada buku berjudul *Profil Kesehatan Jawa Tengah tahun 2023*. Variabel-variabel yang diduga berpengaruh terhadap jumlah balita yang terkena diare disetiap kabupaten/kota di Jawa Tengah (y) meliputi, jumlah balita yang diberikan ASI eksklusif (x_1), jumlah balita yang diberikan vitamin A (x_2), kepadatan penduduk (x_3), jumlah pengguna sanitasi layak (x_4), jumlah pengelolaan air minum dan makanan (x_5), jumlah balita yang kekurangan gizi (x_6), jumlah penduduk miskin (x_7), jumlah posyandu aktif (x_8), jumlah balita yang diimunisasi dasar lengkap (x_9).

2.2 Uji Asumsi Klasik

Uji asumsi klasik menjadi syarat sebelum proses analisis regresi linier berganda. Tujuannya dilakukan uji ini untuk memastikan model yang dihasilkan dapat memberikan ketepatan prediksi dan tidak bias dalam estimasi. Uji asumsi klasik terdiri dari uji normalitas, uji homokedastisitas, uji autokorelasi, dan uji multikolinearitas (Sholihah *et al*., 2023).

Uji normalitas bertujuan untuk memeriksa kesalahan (ε) dari model regresi dalam mendeteksi suatu model telah berdistribusi normal atau tidak. Uji Shapiro Wilk merupakan salah satu teknik dalam menguji normalitas suatu model (Hanusz & TarasiŃska, 2014). Suatu model dikatakan berdistribusi normal apabila nilai p_{value} lebih besar atau sama dengan tingkat alfa yang dipilih (α).

Homokedastisitas dapat diartikan sebagai varians dari kesalahan (ε) yang bernilai konstan dan sama untuk semua pengamatan (Dalić & Terzić, 2021). Uji Breusch Pagan merupakan salah satu teknik dalam mengindentikan suatu model memenuhi asumsi homokedastisitas atau tidak (Anjas *et al.*, 2019). Suatu model dikatakan memenuhi asumsi homokedastisitas apabila nilai p_{value} lebih besar atau sama dengan dari tingkat toleransi α yang dipilih.

Uji autokorelasi merupakan uji untuk mendeteksi adanya korelasi pada kesalahan periode tertentu (ε_t) dengan kesalahan periode sebelumnya (ε_{t-1}). Cara melakukan uji autokorelasi dengan mengunakan uji Durbin-Watson (Aprihartha, 2025). Suatu model dikatakan mengandung autokorelasi apabila nilai p_{value} kurang dari tingkat toleransi α yang dipilih.

Uji Multikolinearitas bertujuan untuk mengidentifikasi adanya korelasi antar variabel indepeden pada model regresi. Dalam mendeteksi skala pengukuran multikolinearitas dalam analisis regresi dapat dilakukan dengan menggunakan *Variance Inflation Factor* (VIF). Nilai VIF yang lebih besar dari sepuluh biasanya menunjukkan adanya multikolinearitas yang sangat tinggi sehingga variabel tersebut disarankan untuk dihapus atau disesuaikan agar tidak menggangu hasil analisis (Choi & Yun, 2025).

2.3 Metode Regresi Linear Berganda

Metode regresi linier berganda adalah teknik statistik klasik yang digunakan untuk meramalkan hasil variabel dependen berdasarkan nilai dua atau lebih variabel independen (Aflah *et al.*, 2025). Model regresi linear menganggap variabel dependen merupakan kombinasi linier dari variabel-variabel independen yang dapat ditunjukkan pada persamaan berikut (Lu *et al*., 2025):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \varepsilon_i, i = 1, 2, \cdots, n \quad (1)$$

dengan i menyatakan pengamatan ke-i dalam populasi, y_i menyatakan variabel terikat atau variabel respons atau variabel dependen atau variable yang dipengaruhi, $x_i$disebut variabel bebas atau variabel penjelas atau variabel independent atau variabel yang memengaruhi, β_0, β_1, β_2, …, β_p mewakili parameter regresi dan ε_i disebut suku kesalahan yang terdiri dari variabel lain diluar model regresi.

Dalam menentukan parameter regresi dapat melalui pendekatan estimasi kuadrat terkecil. Sehingga diperoleh penduga parameter ($\widehat{\boldsymbol{\beta}}$) yang dinyatakan sebagai berikut (Vu *et al*., 2015):

$$\widehat{\boldsymbol{\beta}} = [(\boldsymbol{X}' * \boldsymbol{X})^{-1}] * \boldsymbol{X}' * \boldsymbol{Y} \quad (2)$$

dengan $\boldsymbol{X}$ merupakan matriks variabel indepeden, $\boldsymbol{Y}$ merupakan matriks variabel dependen, dan $\boldsymbol{X}'$ menjadi matriks tranpose dari $\boldsymbol{X}$.

2.4 Analisis Komponen Utama

Analisis komponen utama (AKU) merupakan teknik analisis dalam mengatasi multikolinearitas dengan menghilangkan korelasi antara variabel asli dan mengurangi jumlah dimensi variabel indepeden (Chen *et al.,* 2025). AKU menggunakan transformasi ortogonal untuk menemukan beberapa variabel komprehensif (komponen utama) dari banyak variabel asli untuk menggantikan variabel asli (Kristjanpoller & Minutolo, 2018). Komponen utama ini tidak hanya tidak berkorelasi satu sama lain, tetapi juga mempertahankan informasi utama dari variabel asli (Charitaki *et al.*, 2021).

Misalkan terdapat n data historis dalam himpunan data asli dan setiap data memiliki p faktor yang memengaruhi, yaitu $\boldsymbol{X} = (x_1, x_2, \cdots, x_p)$ dan $\boldsymbol{x_j} = \left[x_{1j}, x_{2j}, \cdots, x_{nj}\right]^T$. Untuk menghilangkan kemungkinan pengaruh yang disebabkan oleh dimensi yang berbeda, himpunan data tersebut distandarisasi dengan persamaan berikut (Chen *et al.,* 2025):

$$x_{ij}{}^* = \frac{x_{ij} - M_j}{S_j} \quad (3)$$

dengan $x_{ij}{}^*$ merupakan data terstandarisasi, x_{ij} merupakan data mentah, $M_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$, dan $S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - M_j)^2}$

Matriks kovariansi dari kumpulan data terstandar dinyatakan sebagai (Chen *et al.,* 2025):

$$\boldsymbol{R} = \frac{1}{n-1}\boldsymbol{X^T X} \quad (4)$$

dengan $\boldsymbol{X}$ merupakan matriks ternormalisasi dan $\boldsymbol{X^T}$ merupakan matriks transpos dari $\boldsymbol{X}$.

Berdasarkan algoritma dekomposisi nilai singular dapat dilakukan perhitungan untuk nilai eigen dari matriks kovariansi (λ_i) dan vektor eigen satuan ortogonal (a_i). Komponen utama ke-i (PC_i) dari variabel asli dapat dinyatakan sebagai:

$$PC_i = a_i \times \boldsymbol{X} \quad (5)$$

dengan $a_i = \frac{\lambda_i}{\sum_{i=1}^{m}\lambda_i}$ dan m adalah jumlah komponen utama.

Tingkat kontribusi dan tingkat kontribusi kumulatif dapat dinyatakan sebagai berikut:

$$\eta_i = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \quad (6)$$

$$\eta_\Sigma = \sum_{i=1}^{m} \eta_i \quad (7)$$

Tingkat kontribusi varians kumulatif mencerminkan proporsi informasi yang dicadangkan oleh komponen utama. Hubungan antara tingkat kontribusi kumulatif dan pengurangan dimensionalitas data mencerminkan pertukaran antara hilangnya informasi dan efisiensi komputasi. Jumlah komponen utama bergantung pada tingkat kontribusi kumulatif dari komponen utama. Semakin besar tingkat kontribusi kumulatif, semakin kecil hilangnya informasi data. Ketika tingkat kontribusi kumulatif mencapai 95%, pada m komponen utama pertama yang sesuai berisi informasi terbanyak dalam set data asli.

Matriks *loading factor* dari komponen utama dinyatakan sebagai (Chen *et al.,* 2025):

$$\boldsymbol{l} = \sqrt{\lambda_i}\boldsymbol{a_{ij}} \quad (8)$$

Skor komponen utama dapat dihitung dengan memasukkan data asli ke dalam ekspresi komponen utama.

$$PC_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip}x_p \quad (9)$$

2.5 Uji Kebaikan Model

Salah satu cara dalam mengukur kebaikan model regresi adalah dengan menghitung nilai koefisien determinasi (R^2). Koefisien determinasi (R^2) menunjukkan jumlah variasi proporsional dalam variabel depeden y yang dijelaskan oleh variabel independen x dalam model regresi linier. Nilai R^2 dapat dihitung dengan persamaan berikut (Aprihartha, 2025):

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (10)$$

dengan $SS_{res} = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2$ merupakan jumlah kuadrat residual dan $SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ merupakan jumlah kuadrat total, $\hat{y}_i$ merupakan hasil prediksi model dan $\bar{y}$ adalah rata-rata variabel dependen.

2.6 Uji Signifikasi Parameter

Dalam analisis regresi, uji t diterapkan untuk mengukur signifikasi koefisien regresi secara masing-masing (β_i) dengan batas toleransi α. Nilai uji t dapat dihasilkan dengan persamaan berikut (Aprihartha, 2025):

$$t_{hitung} = \frac{\beta_i}{SE(\beta_i)} \quad (11)$$

dengan β_i adalah estimasi koefisien regresi untuk variabel independen x_i dan $SE(\beta_i)$ adalah standar *error* dari β_i. Variabel x_i dikatakan berpengaruh signifikan terhadap y apabila nilai $p_{value} < \alpha$ atau $t_{hi} \quad > t_{tabel}$ dengan $df = n - 1$, df merupakan derajat kebebasan.

Sedangkan uji F digunakan untuk mengukur signifikasi koefisien regresi secara keseluruhan. Nilai uji F dihitung dengan persamaan berikut:

$$F_{hitung} = \frac{(SS_{tot} - SS_{res})/_k}{SS_{res}/_{(n-k-1)}} \quad (12)$$

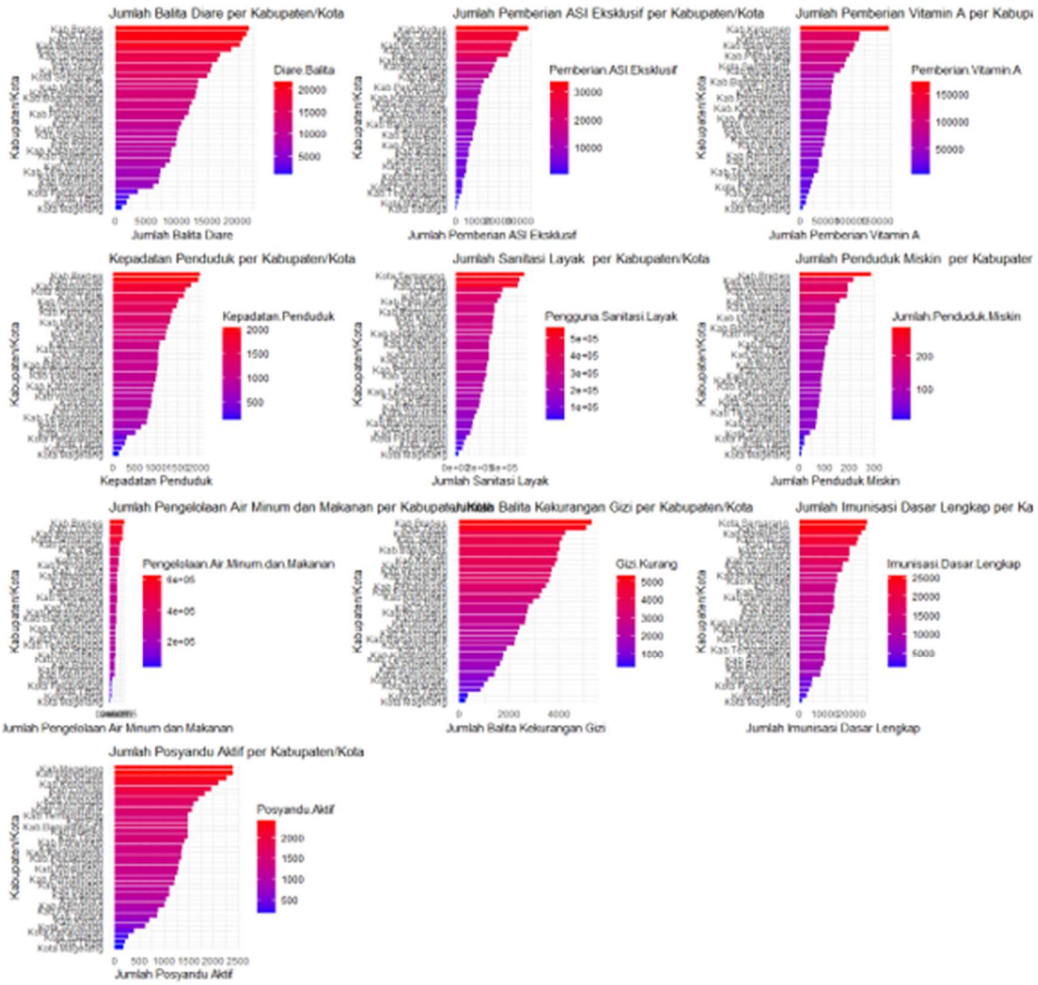
dengan k menyatakan jumlah variabel indepeden. Seluruh variabel independen dikatakan berpengaruh signifikan terhadap y apabila $p_{value} < \alpha$ atau $F_{hitun} \quad > F_{tabel}$ dengan $df_1 = k$ dan $df_2 = n - k - 1$.

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi dan Visualisasi Data

Total kasus diare pada balita dan faktor-faktor yang mempengaruhinya di Jawa Tengah bervariasi berdasarkan kabupaten/kota berdasarkan laporan pada buku berjudul *Profil*

kesehatan Jawa Tengah tahun 2023 yang diterbitkan oleh Dinas Kesehatan Provinsi Jawa Tengah tercermin pada Gambar 1.



Gambar 1. Visualisasi Variabel Independen dan Dependen

Berdasarkan Gambar 1, wilayah dengan jumlah kasus balita yang terkena diare tertinggi digambarkan dengan warna merah, sedangkan jumlah kasus balita yang terkena diare terendah digambarkan dengan warna biru. Kabupaten Brebes memiliki jumlah kasus diare pada balita tertinggi, disusul oleh Kabupaten Tegal dan Kabupaten Cilacap. Sedangkan, Kota Magelang dan Kota Salatiga menunjukkan jumlah kasus diare balita terendah. Terdapat perbedaan jumlah kasus diare pada balita di setiap wilayah yang disebabkan kurangnya penanganan kasus secara menyeluruh dan fasilitas kesehatan yang belum memadai disetiap wilayah di Provinsi Jawa Tengah.

Lokasi penyebaran faktor-faktor yang mempengaruhi kasus diare pada balita di Jawa Tengah ditunjukkan pada Gambar 1. Jumlah balita yang diberi ASI eksklusif tertinggi berasal dari Kabupaten Kudus dan terendah adalah Kota Salatiga. Jumlah balita yang diberi vitamin A tertinggi berasal dari Kabupaten Kebumen dan terendah berasal dari Kota Magelang. Kepadatan penduduk terbesar berasal dari Kabupaten Brebes dan terendah dari Kota Magelang. Kota Semarang memiliki jumlah pengguna sanitasi layak tertinggi sedangkan Kota Magelang memiliki jumlah pengguna sanitasi layak terendah. Jumlah penduduk miskin terbesar berasal dari Kabupaten Brebes dan terendah berasal dari Kota Magelang. Jumlah pengelolaan air minum dan makanan tertinggi berasal dari Kabupaten Brebes dan terendah

berasal dari Kota Magelang. Kabupaten Brebes memiliki jumlah balita kurang gizi tertinggi dan Kota Magelang dengan jumlah balita kurang gizi terendah. Jumlah balita yang diimunisasi dasar lengkap tertinggi berasal dari Kota Semarang dan terendah dari Kota Magelang. Jumlah posyandu aktif tertinggi berasal dari Kabupaten Magelang dan terendah dari Kota Magelang.

3.2 Uji Asumsi Klasik

Uji asumsi klasik digunakan untuk menentukan kelayakan suatu model regresi linear. Adapun uji asumsi klasik terdiri dari uji normalitas, uji homokedastisitas, uji autokorelasi, dan uji multikolinearitas. Dalam setiap pengujian tersebut, hasil kelayakan model dinilai berdasarkan nilai probabilitas (p_{value}) yang diperoleh. p_{value} merupakan probabilitas untuk memperoleh hasil yang setidaknya sama ekstremnya dengan hasil yang diamati, dengan asumsi bahwa hipotesis nol (*null hypothesis*) adalah benar (Biau *et al*., 2010). Hasil analisis uji asumsi klasik pada studi kasus ini disajikan pada Tabel 1 dan Tabel 2.

Tabel 1. Uji Normalitas, Homokedastisitas, dan Autokorelasi

Pengujian	$\boldsymbol{p_{value}}$	$\boldsymbol{\alpha}$
Normalitas	0,772	0,05
Homokedastisitas	0,418	0,05
Autokorelasi	0,902	0,05

Tabel 2. Uji Multikolinearitas

Variabel	**VIF**	**Variabel**	**VIF**
x_1	3,295	x_6	3,615
x_2	5,203	x_7	5,404
x_3	64,401	x_8	2,572
x_4	4,484	x_9	12,800
x_5	23,954		

Pada Tabel 1, diperlihatkan hasil uji normalitas dengan uji Shapiro-Wilk diperoleh $p_{value}\ (0{,}772) > \alpha(0{,}05)$ yang berarti data pada studi kasus ini berasal dari populasi berdistribusi normal. Hasil uji homokedastisitas dengan uji Breusch-Pagan memperlihatkan nilai $p_{value}\ (0{,}418) > \alpha(0{,}05)$ yang berarti homokedastisitas terpenuhi, dengan kata lain tidak ada perbedaan keragamaan sisaan antar observasi. Hasil uji autokorelasi dengan uji Durbin-Watson menunjukkan nilai $p_{value}\ (0{,}902) > \alpha(0{,}05)$ yang berarti tidak ada korelasi sisaan antar observasi.

Sementara itu, pada uji multikolineritas ditetapkan batas toleransi $VIF \leq 10$ agar memenuhi asumsi tidak ada multikolinearitas antar variabel independen. Tabel 2 memperlihatkan variabel x_3, x_5, dan x_9 memiliki nilai $VIF > 10$ yang berarti adanya korelasi antar ketiga variabel tersebut. Oleh karena asumsi multikolinearitas tidak terpenuhi maka digunakan metode analisis komponen utama untuk mereduksi variabel x_3, x_5, dan x_9 dalam mengatasi multikolinearitas.

Model regresi linear berganda dinyatakan dalam persamaan (13) sebagai berikut:

$$\hat{y} = \hat{y} = 9{,}29 \times 10^{-17} + 0{,}096x_1 - 0{,}017x_2 + 1{,}119x_3 - 0{,}233x_4 - 0{,}137x_5 + 0{,}158x_6 - 0{,}00258x_7 - 0{,}151x_8 + 0{,}365x_9 \quad (13)$$

3.3 Analisis Komponen Utama

Dalam analisis komponen utama, setiap variabel distandarisasi untuk meminimalisir kesalahan dalam analisis data dan menyelaraskan jangkauan pada setiap variabel. Tahap pertama menguji kelayakan data dari variabel x_3, x_5, dan x_9 untuk dianalisis dengan metode

komponen utama menggunakan uji Kaiser Meyer Olkin (KMO) dan uji *Measure of Sampling Adequacy* (MSA). KMO dan MSA masing-masing digunakan untuk menentukan kebutuhan akan kecukupan data dalam analisis faktor (Ari, 2014). KMO dan MSA dinyatakan dalam persamaan berikut (Ari, 2014):

$$KMO = \frac{\sum {r_{ij}}^2}{\sum {r_{ij}}^2 + \sum {p_{ij}}^2} \quad (14)$$

$$MSA = \frac{\sum_{i \neq j} {r_{ij}}^2}{\sum_{i \neq j} {r_{ij}}^2 + \sum_{i \neq j} {p_{ij}}^2} \quad (15)$$

dengan r_{ij} adalah koefisien korelasi antara variabel ke-i dan variabel ke-j, dan p_{ij} adalah koefisien korelasi parsial antara variabel ke-*i* dan variabel ke-*j*. Hasil uji ditampilkan pada Tabel 3.

Tabel 3. Uji KMO dan MSA

	x_3	x_5	x_9
MSA	0,730	0,760	0,860
KMO	0,779		

Berdasarkan Tabel 3 diperoleh nilai MSA untuk variabel x_3, x_5, dan x_9 masing-masing berada pada batas nilai lebih dari 0,7 yang mengindikasikan bahwa ketiga variabel layak untuk dianalisis lebih lanjut. Hal ini juga didukung oleh nilai KMO yang lebih dari 0,7 yang berarti secara keseluruhan variabel x_3, x_5, dan x_9 dapat dianalisis dengan metode komponen utama.

Selanjutnya dalam analisis komponen utama adalah menghitung kovarians dari variabel yang mengandung multikolinearitas, yaitu variabel x_3, x_5, dan x_9. Hasil perhitungan tersaji dalam matriks kovarians sebagai berikut:

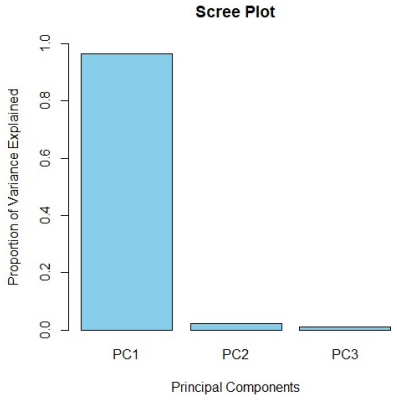
$$\begin{bmatrix} 1 & 0{,}964 & 0{,}946 \\ 0{,}964 & 1 & 0{,}938 \\ 0{,}946 & 0{,}938 & 1 \end{bmatrix}$$

Nilai eigen dari matriks kovarians diperoleh sebagai berikut: $e_1 = 2{,}899$; $e_2 = 0{,}066$; $e_3 = 0{,}034$. Setelah diperoleh nilai eigen maka vektor eigen (skor komponen utama) disajikan Tabel 4 berikut:

Tabel 4. Skor Komponen Utama

Variabel	$\boldsymbol{PC_1}$	$\boldsymbol{PC_2}$	$\boldsymbol{PC_3}$
$\boldsymbol{x_3}$	0,580	0,300	0,758
$\boldsymbol{x_5}$	0,578	0,504	-0,642
$\boldsymbol{x_9}$	0,574	-0,810	-0,119

Skor komponen utama pada Tabel 4 terdiri dari PC_1, PC_2 dan PC_3 yang berkonstribusi dalam menggantikan variabel aslinya x_3, x_5, dan x_9. Pada Gambar ditampilkan proporsi keragaman oleh tiap-tiap komponen utama.



Gambar 2. Proporsi Variansi Skor Komponen Utama

Berdasarkan Gambar 2 dijelaskan bahwa PC_1 memiliki proporsi keragaman sebesar 96,64% yang berarti variabel PC_1 memiliki keragaman yang sangat tinggi dalam mewakili variabel x_3, x_5, dan x_9. Sedangkan PC_2 dan PC_3 masing-masing memiliki proporsi keragaman sangat rendah, sebesar 2,20% dan 1,15%. Ini menunjukkan keduanya menyumbang konstribusi kecil dalam menjelaskan keragaman. Oleh karena itu, dipilih PC_1 sebagai variabel baru yang akan dianalisis lebih lanjut karena mengandung informasi yang sangat besar dalam mewakili variabel x_3, x_5, dan x_9 dibandingkan PC_2 dan PC_3. Persamaan PC_1 yang terbentuk sebagai berikut:

$$PC_1 = 0{,}580x_3 + 0{,}587x_5 + 0{,}574x_9 \quad (16)$$

3.4 Analisis Regresi Komponen Utama

Proses analisis regresi komponen utama dilakukan dengan metode regresi linear dengan memasukkan variabel independen baru (PC_1) dan variabel independen lainnya yang memiliki nilai $VIF \leq 10$ ke dalam model. Model regresi yang terbentuk sebagai berikut:

$$\hat{y} = 9{,}29 \times 10^{17} + 0{,}513PC_1 + 0{,}096x_1 - 0{,}017x_2 - 0{,}233x_4 + 0{,}158x_6 - 0{,}00258x_7 - 0{,}151x_8 \quad (17)$$

Pada variabel PC_1 persamaan (16) disubstitusikan ke dalam model (17) maka diperoleh model regresi sebagai berikut:

$$\hat{y} = 9{,}29 \times 10^{-17} + 0{,}096x_1 - 0{,}017x_2 + 0{,}297x_3 - 0{,}233x_4 + 0{,}296x_5 + 0{,}158x_6 - 0{,}00258x_7 - 0{,}151x_8 + 0{,}294x_9 \quad (18)$$

Berdasarkan model (18), dilakukan perhitungan R^2 dengan parsamaan (10) sehingga diperoleh hasil koefisien determinasi, $R^2 = 96{,}2\%$. Hasil ini dapat dimaknai sebagai model sangat baik dalam menggambarkan hubungan faktor-faktor penyebab diare pada balita terhadap jumlah balita menderita diare disetiap kabupaten/ kota di Jawa Tengah. Sisanya 3,8% disebabkan oleh faktor-faktor lain di luar model.

3.5 Uji Signifikasi Parameter

Berdasarkan parameter dari model regresi komponen utama, faktor-faktor signifikan penyebab balita terkena diare ditampilkan dalam hasil uji parameter parsial pada Tabel 5.

Tabel 5. Hasil Uji Signifikasi Parsial

Variabel	Parameter	p_{value}
x_1	0,062	0,378
x_2	0,158	0,164
x_3	0,317	0,352
x_4	-0,267	0,008
x_5	0,316	0,137
x_6	0,217	0,008
x_7	0,138	0,217
x_8	-0,145	0,031
x_9	0,314	0,039

Berdasarkan Tabel 5, faktor-faktor yang berpengaruh signifikan terhadap jumlah balita terkena diare di setiap kabupaten/kota di Jawa Tengah diantaranya adalah jumlah pengguna sanitasi layak (x_4), jumlah balita yang kekurangan gizi (x_6), jumlah posyandu aktif (x_8), dan jumlah balita yang diimunisasi dasar lengkap (x_9). Hal ini ditunjukkan dengan nilai p_{value} pada variabel tersebut lebih rendah dari batas toleransi $\alpha = 0{,}05$. Sementara itu hasil uji

parameter secara simultan diperoleh nilai $p_{value} = 1{,}680 \times 10^{-15}$. Pada tingkat kepercayaan 95% atau batas toleransi $\alpha = 0{,}05$ maka nilai $p_{value} < \alpha\ (0{,}05)$ yang mengindikasikan bahwa setiap variabel penyebab balita terkena diare berpengaruh signifikan secara bersama-sama terhadap jumlah balita terkena diare.

Hasil penelitian ini senada dengan hasil penelitian oleh Jannah & Susanti (2023), kejadian diare pada balita dipengaruhi oleh imunisasi dasar lengkap dan akses sanitasi layak. Berdasarkan Gambar 1, lima daerah dengan sanitasi terendah, diantaranya Kota Magelang, Kota Salatiga, Kota Tegal, Kabupaten Pekalongan dan Kabupaten Grobogan. Sedangkan lima daerah dengan jumlah imunisasi dasar lengkap terendah adalah Kota Magelang, Kota Salatiga, Kota Tegal, Kota Pekalongan, dan Kota Surakarta. Penelitian lain oleh Widyaningrum *et al*. (2021), variabel yang berpengaruh signifikan terhadap jumlah diare pada balita diantaranya jumlah balita yang kurang gizi dan jumlah posyandu aktif. Kabupaten/ kota yang dengan jumlah posyandu aktif terendah adalah Kota Magelang, Kota Tegal, Kota Surakarta, Kota Salatiga, dan Kota Pekalongan. Sementara itu, untuk kasus jumlah balita kekurangan gizi terbesar diantaranya Kabupaten Brebes, Kabupaten Tegal, Kabupaten Cilacap, Kabupaten Jepara, Kabupaten Klaten. Dalam penelitian Sitanggang *et al*. (2022), faktor imunisasi dasar lengkap memiliki risiko yang berpengaruh dalam kejadian diare pada balita. Penelitian oleh Azhar *et al*. (2015) memberikan hasil bahwa akses sanitasi layak dapat menjadi faktor penyebab dalam kejadian diare pada balita.

Dari sisi penerapan, hasil ini memiliki beberapa makna penting bagi pemerintah daerah dan lembaga kesehatan masyarakat, yaitu Pemerintah daerah perlu memperhatikan daerah dengan sanitasi layak yang rendah. Melalui peningkatan akses sanitasi layak diharapkan mampu menanggulangi resiko penularan diare pada balita. Bertambahnya jumlah posyandu aktif dapat menjadi kunci dalam pencegahan penyakit karena posyandu berperan dalam edukasi sanitasi, gizi, dan deteksi dini penyakit. Daerah dengan jumlah imunisasi dasar lengkap yang rendah perlu menjadi fokus dalam pemerataan layanan imunisasi dan edukasi Kesehatan. Terakhir, wilayah dengan angka balita kurang gizi tinggi (seperti Brebes dan Tegal) perlu menjadi prioritas program intervensi gizi dan peningkatan kesejahteraan masyarakat.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan diperoleh hasil uji asumsi klasik menunjukkan terpenuhinya uji normalitas, homokedastisitas, dan tidak adanya autokorelasi dalam model regresi. Namun, pada uji multikolinearitas terdapat tiga variabel independen yang saling berkorelasi diantaranya jumlah balita yang diberikan kepadatan penduduk (x_3), jumlah pengelolaan air minum dan makanan (x_5), dan jumlah balita yang diberikan imunisasi dasar lengkap (x_9). Oleh karena itu, dalam mengatasi masalah multikolinearitas maka digunakan analisis komponen utama untuk mereduksi variabel yang saling berkorelasi.

Hasil uji analisis komponen utama menghasilkan tiga variabel baru, yaitu PC_1, PC_2, dan PC_3. Pada ketiga skor komponen utama terpilih variabel PC_1 untuk analisis lebih lanjut karena memiliki proporsi keragaman sebesar 96,64%, yang berarti variabel PC_1 memiliki keragaman yang sangat tinggi dalam mewakili variabel x_3, x_5, dan x_9 dibandingkan PC_2 dan PC_3. Selanjutnya melakukan analisis regresi komponen utama melibatkan variabel PC_1dan variabel independen lainnya yang tidak saling berkorelasi. Model regresi menghasilkan nilai $R^2 = 96{,}2\%$ yang dapat diartikan sebagai model sangat baik dalam menggambarkan hubungan faktor-faktor penyebab diare pada balita terhadap jumlah balita menderita diare disetiap kabupaten/ kota di Jawa Tengah. Adapun faktor-faktor yang memberikan pengaruh signifikan terhadap kasus jumlah balita menderita diare di kabupaten/kota di Jawa Tengah diantaranya

jumlah pengguna sanitasi layak (x_4), jumlah balita yang kekurangan gizi (x_6), jumlah posyandu aktif (x_8), dan jumlah balita yang diimunisasi dasar lengkap (x_9).

Penelitian selanjutnya disarankan untuk menambahkan variabel lain yang mungkin berpengaruh terhadap kejadian diare pada balita, seperti curah hujan atau kondisi iklim. Selain itu, penggunaan model analisis spasial dapat memberikan gambaran yang lebih komprehensif mengenai distribusi risiko diare di berbagai wilayah.

5. DAFTAR PUSTAKA

Aflah, F. R., Risnawati, R., & Hamdani, M. F. (2025). Penerapan regresi linier berganda dalam menilai hubungan antar variabel dalam penelitian kuantitatif. *Innovative: Journal Of Social Science Research*, *5*(3), 4195-4211.

Anjas, M., Sukarsa, I. K. G., & Kencana, I. P. E. N. (2019). Penerapan metode Geographically Weighted Regression (GWR) pada kasus penyakit pneumonia di provinsi Jawa Timur. *E-Jurnal Matematika*, *8*(1), 27-34.

Aprihartha, M. A. (2025). Pemilihan model regresi linear berganda terbaik untuk menentukan faktor-faktor penyebab kasus balita gizi buruk di Jawa Tengah. *Jurnal EurekaMatika*, *13*(1), 35-46.

Ari, M. (2014). Identifikasi perilaku calon mahasiswa baru dalam memilih program studi menggunakan analisis faktor. *Paradigma*, *16*(1), 34-44.

Azhar, K., Dwi Sisca, K., & Hapsari, D. (2015). Diare balita di Provinsi DKI Jakarta ditinjau dari aspek air minum, sanitasi dan PHBS (Analisis Data Riskesdas 2013). *Jurnal Ekologi Kesehatan*, 14(1), 29-40.

Biau, D. J., Jolles, B. M., & Porcher, R. (2010). P value and the theory of hypothesis testing: an explanation for new researchers. *Clinical Orthopaedics and Related Research®*, *468*(3), 885-892.

Charitaki, G., Soulis, S. G., & Tyropoli, R. (2021). Academic self-regulation in autism spectrum disorder: A principal components analysis. *International Journal of Disability, Development and Education*, *68*(1), 26-45.

Choi, J., & Yun, J. I. (2025). Optimization of water chemistry to mitigate corrosion products in nuclear power plants using big data and multiple linear regression in machine learning. *Progress in Nuclear Energy*, *183*, 105685.

Chen, K., Liu, K., Zhou, Y., Li, Y., Wu, G., Gao, G., ... & Djerdir, A. (2025). State of health prognosis for polymer electrolyte membrane fuel cell based on principal component analysis and Gaussian process regression. *International Journal of Hydrogen Energy*, *98*, 933-943.

Dalić, I., & Terzić, S. (2021). Violation of the assumption of homoscedasticity and detection of heteroscedasticity. *Decision Making: Applications in Management and Engineering*, *4*(1), 1-18.

Hanusz, Z., & TarasiŃska, J. (2014). Simulation study on improved Shapiro–Wilk tests for normality. *Communications in Statistics-Simulation and Computation*, *43*(9), 2093-2105.

Jannah, M., & Susanti, R. (2023). Faktor yang memengaruhi kejadian diare pada balita di Indonesia dengan menggunakan regresi data panel. *Prosiding Seminar Nasional Matematika dan Statistika,* 3(1), 168-182.

Khoirunnisa, A., Nyoman, I. G., & Jaya, M. (2019). Modelling and mapping the average of incidence rate diarrhea among toddlers in Bandung city in 2013-2018 using Geographically Weighted Regression. *International Journal of Innovative Science, Engineering & Technology*, *6*(11), 115-19.

Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications*, *109*, 1-11.

Li, B., Lu, Y., Sun, X., Chen, X., Gong, W., & Miao, F. (2024). Radial artery pulse wave age-related assessment for diabetic patients based on multiple linear regression time domain analysis method. *Extreme Mechanics Letters*, *70*, 102185.

Lu, X., Teh, S. Y., Tay, C. J., Kassim, N. F. A., Fam, P. S., & Soewono, E. (2025). Application of multiple linear regression model and long short-term memory with compartmental model to forecast dengue cases in Selangor, Malaysia based on climate variables. *Infectious Disease Modelling*, *10*(1), 240-256.

Marcus, G. L., Wattimanela, H. J., & Lesnussa, Y. A. (2012). Analisis regresi komponen utama untuk mengatasi masalah multikolinieritas dalam analisis regresi linier berganda. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *6*(1), 31-40.

Pendi, P. (2021). Analisis regresi dengan metode komponen utama dalam mengatasi masalah multikolinearitas. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, *10*(1).

Putri, F. C., Suciptawati, N. L. P., & Susilawati, M. (2023). Implementasi metode Geographically Weighted Regression (GWR) pada kasus diare balita di Provinsi Jawa Timur. *E-Jurnal Matematika*, *12*(2), 92.

Sholihah, S. M. A., Aditiya, N. Y., Evani, E. S., & Maghfiroh, S. (2023). Konsep uji asumsi klasik pada regresi linier berganda. *Jurnal Riset Akuntansi Soedirman*, *2*(2), 102-110.

Sitanggang, H. D., Ramadhanti, I. T., & Halim, R. D. (2022). Faktor risiko kejadian diare pada anak balita (12-59 bulan) di Puskesmas Kota Jambi. *Riset Informasi Kesehatan*, *11*(1), 54-63.

Widyaningrum, A. R., Susanti, Y., & Slamet, I. (2021). Pemodelan penyakit diare balita di Jawa Timur menggunakan regresi robust. In *SINASIS (Seminar Nasional Sains)* (Vol. 2, No. 1).

Vu, D. H., Muttaqi, K. M., & Agalgaonkar, A. P. (2015). A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, *140*, 385-394.