Jurnal EurekaMatika 13(1) (2025) 35-46

Jurnal EurekaMatika

Journal homepage: https://ejournal.upi.edu/index.php/JEM

Pemilihan Model Regresi Linear Berganda Terbaik untuk Menentukan Faktor-Faktor Penyebab Kasus Balita Gizi Buruk di Jawa Tengah

Moch. Anjas Aprihartha, Salwa Paramita Azzahro, dan Rahmatul Aziza*[3]

Program Studi PJJ Informatika, Universitas Dian Nuswantoro

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

ABSTRAK

Metode regresi linear berganda adalah metode statistik yang memodelkan hubungan antara dua atau lebih variabel independen dan variabel depeden dengan memasukkan persamaan linier. Dalam menghasilkan model regresi terbaik, perlu dipastikan tidak ada multikolinearitas, yang diatasi melalui pemilihan model terbaik dengan metode regresi seleksi maju, regresi eliminasi mundur, dan metode regresi bertahap. Permasalahan yang dapat diterapkan dengan metode regresi adalah mengidentifikasi faktor penyebab terjadinya kasus gizi buruk pada balita di Jawa Tengah. Tujuan penelitian ini untuk mengumpulkan variabel optimal yang berpengaruh signifikan terhadap masalah gizi buruk balita di Jawa Tengah. Hasil penelitian diperoleh model terbaik adalah model regresi dengan eliminasi mundur. Variabel yang berpengaruh signifikan terhadap jumlah gizi buruk balita yaitu jumlah penduduk miskin, pengelolaan air minum dan makanan, dan jumlah posyandu aktif.

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima 13 Maret 2025
Direvisi 19 Maret 2025
Disetujui 14 April 2025
Tersedia online 2 Mei 2025
Dipublikasikan 2 Mei 2025

Kata Kunci:
Gizi Buruk,
Metode Eliminasi Mundur,
Metode Seleksi Maju,
Regresi Bertahap,
Regresi Linear Berganda.

ABSTRACT

The multiple linear regression method is a statistical method that models relationship between two or more independent variables and dependent variables by entering a linear equation. To obtain the best regression model, it is necessary to ensure that there is no multicollinearity, which is overcome by selecting a model using forward selection regression, backward elimination, and stepwise regression. The problem that can be applied with the regression method is identifying the causal factors of cases of malnutrition in toddlers in Central Java. The purpose of this study was to collect optimal variables that have a significant effect on the problem of malnutrition in toddlers in Central Java. The results of study obtained best model is backward elimination regression model. The variables that have a significant influence on the amount of malnutrition in toddlers are the number of poor people, management of drinking water and food, and number of active integrated health posts.

Keywords:
Backward Elimination Regression,
Forward Selection Regression,
Malnutrition,
Regression Model,
Stepwise Regression Multiple.

1. PENDAHULUAN

Metode regresi linear berganda adalah salah satu metode statistik penting yang berupaya memodelkan hubungan antara dua atau lebih variabel prediktor independen dan variabel depeden dengan memasukkan persamaan linier ke dalam data yang diamati (Al-Dairi, et.al, 2024). Selain untuk tujuan prediksi kejadian yang diamati, model regresi juga digunakan dalam keperluan lain seperti menentukan pengaruh signifikan variabel independen terhadap variabel dependen. Dalam menghasilkan model regresi terbaik maka terdapat beberapa asumsi yang harus terpenuhi seperti normalitas, homokedastisitas, tidak adanya autokorelasi, dan tidak terdapat multikolinearitas pada model (Sholihah, et.al, 2023). Namun pada beberapa kasus, model regresi linear berganda tidak dapat diterapkan karena beberapa asumsi tidak terpenuhi. Hal ini dapat disebabkan masuknya variabel indepeden yang berlebihan ke dalam model regresi. Salah satu cara mengatasi hal tersebut dengan menghapus beberapa variabel indepeden yang tidak memiliki pengaruh cukup besar terhadap variabel dependen.

Teknik pemilihan variabel sekuensial adalah alat yang digunakan untuk mereduksi ruang variabel berdimensi d menjadi subruang fitur berdimensi k dengan $k < d$ (Shafiee, et.al, 2021). Tujuan dari teknik pemilihan variabel agar secara otomatis memilih subkumpulan variabel indepeden yang paling relevan dengan permasalahan yang diteliti. Hal ini dapat meningkatkan efisiensi komputasi atau mengurangi kesalahan generalisasi model dengan menghilangkan variabel atau noise yang tidak relevan. Metode ini dapat bekerja dalam salah satu dari dua cara, yaitu metode seleksi maju atau metode eliminasi mundur.

Salah satu masalah yang dapat diterapkan dengan metode regresi adalah mengidentifikasi faktor penyebab terjadinya peristiwa gizi buruk pada balita di Jawa Tengah. Berdasarkan data Profil Kesehatan Provinsi Jawa Tengah Tahun 2023 yang dihimpun oleh Dinas Kesehatan Jawa Tengah yang dapat diakses pada https://dinkesjatengprov.go.id, jumlah balita yang terkena gizi buruk mengalami peningkatan tiap tahunnya, mulai dari 2018 sampai 2022. Pada tahun 2022, jumlah balita yang menderita gizi buruk mencapai hampir 3000 jiwa. Namun pada tahun 2023, jumlah balita gizi buruk mengalami penurunan hanya berkisar 6,67%. Beberapa penelitian pernah dilakukan terkait peristiwa tersebut diantaranya, penelitian oleh (Maulani & Suherman, 2016) yang menganalisis faktor-faktor penyebab gizi buruk di Jawa Barat. Hasil penelitian diperoleh faktor-faktor yang berpengaruh signifikan terhadap gizi buruk balita adalah berat badan bayi lahir rendah, pemberian vitamin A, sarana kesehatan, pemberian ASI eksklusif, jumlah penduduk miskin, dan usia perkawinan pertama ≤ 15. Penelitian yang dilakukan oleh (Pratnyaningrum, et.al, 2015) tentang pemodelan kasus balita gizi buruk di Jawa Tengah tahun 2012, diperoleh faktor-faktor yang berpengaruh signifikan terhadap kasus tersebut adalah pemberian ASI eksklusif, pemberian imunisasi BCG, pemberian vitamin A, jumlah balita penderita pneumonia, dan jumlah rumah tangga yang memiliki akses air bersih. Penelitian lainnya oleh (Andana, et.al, 2017) yang menggunakan pemodelan regresi pada kasus gizi buruk di Jawa Tengah. Hasil penelitian diperoleh balita dengan gizi buruk dipengaruhi oleh beberapa faktor signifikan yaitu pemberian ASI eksklusif, rumah tangga berprilaku hidup bersih dan sehat, pemberian imunisasi hepatitis B, pemberian imunisasi DPT-HB3, rumah dengan sanitasi yang layak, dan rumah dengan akses air bersih.

Penelitian ini akan menggunakan metode regresi linear berganda dengan teknik seleksi maju, eliminasi mundur dan kombinasi keduanya untuk memilih kumpulan variabel independen terkait masalah gizi buruk pada balita di Jawa Tengah. Kemudian akan dipilih model terbaik yang memiliki pengaruh lebih besar terhadap masalah yang diamati. Model regresi terbaik merupakan model dengan kumpulan variabel optimal yang hanya mencakup berbagai penyebab signifikan terhadap masalah gizi buruk pada balita di Jawa Tengah.

2. METODE

2.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam analisis merupakan data kesehatan Provinsi Jawa Tengah tahun 2023. Data tersebut mencakup faktor-faktor yang diduga berpengaruh terhadap jumlah balita yang menderita gizi buruk tiap Kabupaten/ Kota di Jawa Tengah (y). Faktor-faktor tersebut meliputi berat badan bayi lahir rendah (x_1), pemberian vitamin A (x_2), pemberian ASI eksklusif (x_3), jumlah penduduk miskin (x_4), pengelolaan air minum dan makanan (x_5), imunisasi dasar lengkap (x_6), jumlah puskesmas aktif (x_7), jumlah posyandu aktif (x_8), jumlah rumah sehat (x_9), pengguna air bersih (x_{10}), dan jumlah sanitasi yang layak (x_{11}). Variabel-variabel tersebut digunakan dalam mengidentifikasi jumlah variabel optimal dalam pembentukan model regresi dengan teknik seleksi maju, eliminasi mundur, dan regresi bertahap.

2.2 Uji Asumsi Klasik

Uji asumsi klasik bertujuan untuk menjamin model yang didapatkan memberikan ketetapan dan konsisten, dan tidak bias dalam estimasi. Uji asumsi klasik meliputi uji normalitas, uji homokedastisitas, uji multikolinearitas, dan uji autokorelasi (Sholihah, et.al, 2023).

Uji normalitas bertujuan untuk mengetahui suatu model regresi telah berdistribusi normal atau tidak dengan memeriksa kesalahan acak dari model. Apabila kesalahan acak mengikuti distribusi normal maka model regresi dikatakan berdistribusi normal. Salah satu cara melakukan uji normalitas dengan mengunakan uji Shapiro Wilk (Hanusz & Taransinska, 2014). Hipotesis nol dari uji ini adalah populasi terdistribusi normal. Jadi jika nilai p_{value} kurang dari tingkat toleransi α yang dipilih, maka hipotesis nol ditolak dan ada bukti bahwa data yang diuji tidak berdistribusi normal. Dengan kata lain, data tersebut tidak normal. Sebaliknya, jika nilai p_{value} lebih besar atau sama dengan dari tingkat alfa yang dipilih, maka hipotesis nol bahwa data berasal dari populasi yang berdistribusi normal tidak dapat ditolak.

Commented [1]: Bagaimana keputusan kalau p (value) sama dengan tingkat toleransi alfa yang dipilih?

Homokedastisitas menyiratkan bahwa varians dari kesalahan acak adalah konstan dan sama untuk semua pengamatan (Dalic & Terzic, 2021). Apabila kesalahan acak pada model regresi linier klasik tidak homoskedastis, maka bersifat heteroskedastisitas. Untuk menguji suatu model bersifat homokedastisitas dalam dilakukan uji Breusch Pagan (Anjas, et.al, 2019). Pada prinsipnya uji ini didasarkan estimasi koefisien regresi yang diperoleh dengan metode kuadrat terkecil seharusnya tidak berbeda secara signifikan dari estimasi yang diharapkan jika hipotesis terkait homokedastisitas benar. Hipotesis nol dari uji ini adalah varians dari kesalahan acak bersifat konstan (homokedastisitas). Jadi jika nilai p_{value} lebih kecil dari tingkat toleransi α yang dipilih, maka hipotesis nol ditolak dan ada bukti bahwa varians dari kesalahan acak masih bergantung pada variabel indepeden (heterokedastisitas).

Uji autokorelasi merupakan alat uji untuk mengetahui korelasi antara kesalahan pada suatu periode tertentu dengan kesalahan penggangu pada periode sebelumnya. Autokorelasi disebabkan oleh pengamatan yang semuanya berkaitan satu sama lain. Salah satu cara melakukan uji autokorelasi dengan mengunakan uji Durbin-Watson. Uji Durbin-Watson menggunakan residu untuk menentukan apakah ρ sama dengan nol, dengan ρ adalah koefisien korelasi. Hipotesis nol dari uji ini adalah bahwa tidak ada korelasi antar residual ($\rho = 0$). Jadi jika nilai p_{value} kurang dari tingkat toleransi α yang dipilih, maka hipotesis nol ditolak dan ada bukti bahwa data yang diuji terdapat autokorelasi. Jika autokorelasi signifikan teridentifikasi, dua pendekatan dapat dilakukan. Pertama, menghilangkan satu atau lebih variabel independen utama yang memiliki efek berdasarkan waktu terhadap variabel

dependen atau melakukan analisis dengan variabel yang ditransformasikan (Adrianto, et.al, 2023).

Uji Multikolinearitas pada model regresi bertujuan untuk mendeteksi adanya korelasi antar variabel indepeden. Dalam mengidentifikasi tingkat keparahan multikolinearitas dalam analisis regresi dapat dilakukan dengan menggunakan *Variance Inflation Factor* (VIF) (O'brien, 2007). Selain itu, perhitungan VIF cukup sederhana dan komprehensif. Semakin tinggi nilai VIF maka semakin tinggi kolinearitas antara variabel terkait (Vu & Agalgaonkar, 2015). Nilai VIF yang lebih besar dari sepuluh biasanya menunjukkan adanya multikolinearitas sehingga variabel tersebut disarankan untuk dihapus atau disesuaikan agar tidak menggangu hasil analisis (Choi & Yun, 2025). VIF_j dari salah satu variabel independen x_j dihitung berdasarkan hubungan linear antara variabel indepeden x_j dan variabel independen lainnya $\{x_1, x_2, \cdots, x_{j-1} \cdots, x_{j+1}, x_m\}$ (Alin, 2010).

$$VIF_j = \frac{1}{(1-{R_j}^2)}$$

dengan ${R_j}^2$ adalah koefisien determinasi regresi x_j pada semua variabel independen lainnya dalam kumpulan data $\{x_1, x_2, \cdots, x_{j-1} \cdots, x_{j+1}, x_m\}$.

2.3 Metode Regresi Linear Berganda

Metode regresi linier berganda (MLR) adalah teknik statistik yang menggunakan beberapa variabel independen untuk memprediksi hasil dari variabel depeden. Model regresi linier berganda menganggap variabel dependen sebagai kombinasi linier dari variabel-variabel independen. Nilai koefisien model ditentukan dengan metode kuadrat terkecil dengan meminimalkan galat kuadrat antara nilai variabel dependen eksperimental dan nilai yang diperoleh dari model. Struktur dasar model ditunjukkan sebagai 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 mengacu pada individu ke-i dalam populasi, y_i mewakili nilai-nilai yang diamati, yang sering disebut variabel terikat atau variabel respons, x_i disebut variabel bebas, variabel penjelas atau regressor, β_0 mewakili suku intersep sedangkan β_1, β_2, …, β_p mengacu pada koefisien regresi dan ε_i disebut suku kesalahan yang mencakup semua faktor lain yang mempengaruhi variabel terikat selain dari regressor.

Persamaan (1) pada dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$Y = X * \beta + \varepsilon \quad (2)$$

Berdasarkan persamaan (2) dengan X' menjadi matriks tranpose dari X, estimasi kuadrat terkecil dari β dapat dihitung sebagai $\hat{\beta}$ yang disajikan dalam bentuk sebagai berikut:

$$\hat{\beta} = [(X' * X)^{-1}] * X' * Y$$

2.4 Metode Seleksi Maju *(Forward Selection Method)*

Metode seleksi maju merupakan teknik yang digunakan untuk mengurangi jumlah variabel melalui seleksi bertahap satu per satu variabel independen. Metode ini dimulai dengan memasukkan satu per satu variabel secara berurutan untuk dimodelkan dengan menggunakan metode yang ditentukan (Shafiee, et.al, 2021). Kemudian memilih model terbaik berdasarkan kriteria tertentu seperti R^2 terbesar untuk dipertahankan (Fiola, et.al, 2024). Pada langkah kedua, dari model yang telah ditetapkan sebelumnya, masing-masing variabel independen yang tersisa ditambahkan secara berturut-turut sedemikian sehingga variabel berdasarkan kriteria R^2 terbesar dimasukkan ke dalam model. Proses dilakukan

berulang sedemikian hingga mendapatkan model regresi terbaik dengan jumlah variabel optimum.

2.5 Metode Eliminasi Mundur *(Backward Elimination Method)*

Metode eliminasi mundur merupakan metode yang digunakan untuk mereduksi jumlah variabel melalui eliminasi variabel indepeden secara bertahap satu per satu. Metode ini dimulai dengan membuat model yang mencakup semua variabel independen. Pembatasan jumlah variabel independen pada model akan dicapai melalui tahapan berturut-turut dengan membandingkan efek setiap variabel dalam model berdasarkan ambang batas kriteria eliminasi. Variabel independen yang memiliki nilai kontribusi lebih rendah dari nilai yang ditetapkan dari kriteria eliminasi akan dihapus dari model, dengan kata lain variabel yang memberikan penurunan kinerja paling sedikit berdasarkan kriteria F_{value} atau p_{value} (Fariha & Subekti, 2018). Nilai p_{value} atau nilai probabilitas adalah tingkat signifikasi marjinal dalam uji hipotesis statistik, yang menggambarkan seberapa besar kemungkinan hasil akan terjadi jika hipotesis nol benar (Wasserstein & Lazar, 2016). Pada setiap langkah, variabel yang tidak penting dieliminasi sedemikian hingga menghasilkan model dengan variabel optimum, ketika model memiliki nilai F_{value} lebih rendah atau p_{value} lebih tinggi dibandingkan model sebelumnya maka proses eliminasi terhenti.

2.6 Metode Regresi Bertahap *(Stepwise Regression Method)*

Dalam kasus metode regresi linear berganda, ada kemungkinan untuk menguji tingkat signifikansi statistik dari koefisien variabel prediksi. Disarankan untuk mempertahankan variabel dalam model (dengan toleransi tinggi) yang memiliki koefisien signifikan secara statistik. Semakin tinggi rasio jumlah variabel independen yang signifikan secara statistik terhadap jumlah total variabel independen, maka semakin kuat prediksi model regresi dengan kesalahan model regresi. Model regresi bertahap digunakan jika jumlah variabel independen tinggi (Cozmuta, 2025). Teknik ini telah diterapkan untuk memilih jumlah prediktor optimal yang akan disertakan dalam setiap model regresi berganda (Wang, et al., 2007). Metode bertahap dapat diterapkan dalam dua varian, yaitu metode regresi dengan seleksi maju dan metode regresi dengan eliminasi mundur. Tahapan pembentukan model diawali dari model konstan, algoritma akan menambahkan dan menghapus variabel di dalam model berdasarkan nilai F_{value} atau p_{value} . Proses berhenti ketika nilai p_{value} terendah masih signifikan dan tidak ada lagi variabel indepeden yang dapat ditambahkan atau dihapus.

2.7 Uji Kebaikan Model

Salah satu cara dalam mengukur kebaikan model regresi adalah dengan menghitung nilai koefisien determinasi (R^2) dan simpangan baku residual (RMSE). Koefisien determinasi (R^2) menunjukkan jumlah variasi proporsional dalam variabel depeden y yang dijelaskan oleh variabel independen x dalam model regresi linier (Lewis-Beck & Skalaban, 1990). Nilai R^2 dapat dihitung dengan persamaan berikut:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

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 $\underline{y}$ adalah rata-rata variabel dependen. Sementara itu, RMSE dapat dihitung sebagai ukuran tentang seberapa akurat model dalam memprediksi variabel depeden dengan memperkirakan

standar deviasi dari distribusi kesalahan. Nilai RMSE dapat dihitung dengan persamaan berikut:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

dengan n total observasi.

2.8 Uji Signifikasi Parameter Model

Dalam analisis regresi, uji statistik seperti uji t dan uji F untuk menilai signifikansi model. Uji t digunakan untuk mengevaluasi adanya signifikasi koefisien regresi secara individual (β_i) dengan ambang batas α. Hasil uji t dapat diperoleh menggunakan persamaan berikut:

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

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

Berbeda dengan uji t, uji F digunakan untuk mengevaluasi adanya signifikasi koefisien regresi secara keseluruhan. Hasil uji F dihitung dengan persamaan berikut:

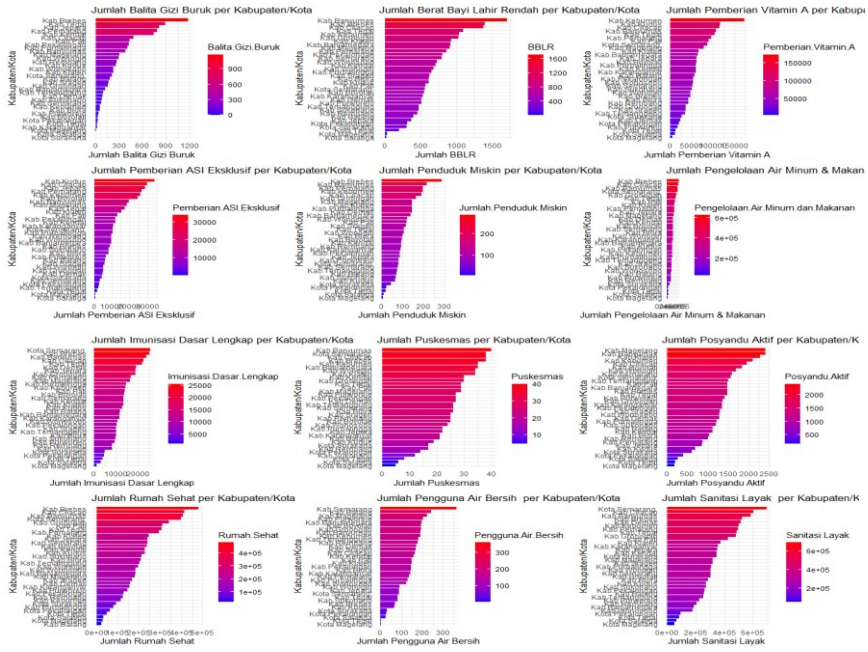
$$F_{hitung} = \frac{\frac{(SS_{tot} - SS_{res})}{k}}{\frac{SS_{res}}{(n-k-1)}}$$

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

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi dan Visualisasi Data

Representasi pada karakteristik variabel-variabel dalam penelitian ini disajikan pada Gambar 1. Jumlah balita yang menderita gizi buruk tertinggi berasal dari Kabupaten Brebes dengan total 1181 balita. Jumlah bayi dengan berat lahir rendah tertinggi berasal dari Kabupaten Banyumas, disusul Kabupaten Brebes dan Kabupaten Cilacap. Kota Magelang memiliki jumlah pemberian vitamin A paling rendah dibandingkan daerah lainnya. Kota Salatiga dan Kota Magelang memiliki jumlah pemberian ASI eksklusif terendah. Jumlah penduduk miskin terbesar dimiliki Kabupaten Brebes. Jumlah pengelolaan air minum dan makanan terendah berasal dari Kota Malang. Kota Semarang memiliki jumlah balita yang telah diimunisasi dasar lengkap tertinggi sedangkan yang terendah adalah Kota Magelang. Kabupaten Banyumas memiliki jumlah puskesmas aktif tertinggi, yaitu 40 buah puskesmas. Jumlah posyandu aktif paling tinggi berasal dari Kabupaten Magelang. Kabupaten Brebes memiliki jumlah rumah sehat tertinggi sedangan yang teredah berasal dari Kabupaten Batang. Jumlah pengguna air bersih paling tinggi berasal dari Kabupaten Semarang. Terakhir, Kota Semarang memiliki jumlah sanitasi layak tertinggi, disusul oleh Kabupaten Cilacap dan Kabupaten Banyumas sedangkan jumlah sanitasi layak terendah berasal dari Kota Magelang.



Gambar 1. Visualisasi Variabel Dependen dan Indepeden

3.2 Model Regresi Linear Berganda

Model regresi linear berganda digunakan untuk membangun hubungan linear antara jumlah balita yang menderita diare dan faktor-faktor yang mempengaruhinya. Model regresi disajikan dalam bentuk persamaan berikut:

$$\hat{y}_i = 0{,}025 + 0{,}080x_{i1} + 0{,}034x_{i2} + 0{,}155x_{i3} + 0{,}351x_{i4} + 0{,}882x_{i5} + 0{,}427x_{i6} + 0{,}167x_{i7} - 0{,}524x_{i8} - 0{,}263x_{i9} - 0{,}229x_{i10} - 0{,}616x_{i11}$$

Kemudian langkah selanjutnya adalah melakukan uji asumsi klasik pada model yang diperoleh.

Tabel 1. Uji Normalitas, Homokedastisitas, dan Autokorelasi Model Awal

Pengujian	p_{value}	α
Normalitas	0,491	0,05
Homokedastisitas	0,156	0,05
Autokorelasi	0,071	0,05

Berdasarkan Tabel 1, hasil uji Shapiro Wilk diperoleh $p_{value} = 0{,}491$ yang lebih besar dari $\alpha = 0{,}05$. Ini menunjukkan bahwa model regresi berdistribusi normal. Selanjutnya hasil uji Breusch Pagan diperoleh $p_{value} = 0{,}156$ lebih besar dari $\alpha = 0{,}05$ yang menunjukkan bahwa model memenuhi homokedastisitas. Hasil uji Durbin Watson diperoleh $p_{value} = 0{,}071$ lebih besar dari $\alpha = 0{,}05$. Hal ini berarti model regresi tidak mengandung autokorelasi.

Tabel 2. Uji Multikolinearitas Model Awal

Variabel	VIF	Variabel	VIF
x_1	7,778	x_7	6,9661
x_2	5,465	x_8	4,658
x_3	3,711	x_9	14,373
x_4	5,062	x_{10}	1,895
x_5	44,269	x_{11}	7,521
x_6	13,155		

Pada Tabel 2, terlihat bahwa nilai VIF dari x_5, x_6, dan x_9 lebih besar dari sepuluh yang menunjukkan adanya korelasi antarvariabel independen. Oleh karena itu, proses selanjutnya dilakukan seleksi, eliminasi, dan kombinasi keduanya pada variabel independen untuk menghasilkan model regresi linear terbaik yang tidak mengandung multikolinearitas.

a. Model Regresi Linear dengan Metode Eleminasi Mundur
   Metode eliminasi mundur dimulai dengan semua variabel independen diregresikan dengan variabel dependen. Salah satu variabel dikeluarkan secara manual untuk menghasilkan subset baru setiap saat, jika setelah dihilangkan memenuhi kondisi dimana p_{value} lebih kecil dari $\alpha = 0{,}05$. Ketika tidak ada varibel independen dalam subset yang dapat memenuhi tersebut, proses eliminasi dihentikan dan subset model regresi optimal terselesaikan. Hasil model regresi dengan metode eliminasi mundur disajikan dalam persamaan berikut:
   $$\hat{y}_i = 0{,}021 + 0{,}207x_{i3} + 0{,}461x_{i4} + 0{,}925x_{i5} - 0{,}555x_{i8} - 0{,}393x_{i11}$$
   Berdasarkan persamaan diatas faktor-faktor yang memengaruhi gizi buruk pada balita antara lain pemberian asi eksklusif (x_3), jumlah penduduk miskin (x_4), pengelolaan air minum dan makanan (x_5), jumlah posyandu aktif (x_8), dan sanitasi yang layak (x_{11}).
b. Model Regresi Linear dengan Metode Seleksi Maju
   Metode seleksi maju diawali dengan model regresi tanpa variabel independen. Kemudian secara bertahap menambahkan variabel independen, jika setelah ditambahkan memenuhi kondisi dimana RMSE lebih kecil atau sama dengan nilai RMSE sebelum ditambahkan atau R^2 lebih besar dari nilai R^2 sebelumnya. Ketika tidak ada varibel independen dalam subset yang dapat memenuhi kedua kondisi secara bersamaan, proses seleksi dihentikan dan model regresi optimal terselesaikan. Hasil model regresi dengan metode maju disajikan dalam persamaan berikut:
   $$\hat{y}_i = 0{,}038 + 0{,}156x_{i3} + 0{,}444x_{i4} + 0{,}528x_{i5} + 0{,}487x_{i6} - 0{,}403x_{i8} - 0{,}201x_{i10} - 0{,}558x_{i11}$$
   Berdasarkan persamaan diatas faktor-faktor yang memengaruhi gizi buruk pada balita antara lain pemberian asi eksklusif (x_3), jumlah penduduk miskin (x_4), pengelolaan air minum dan makanan (x_5), imunisasi dasar lengkap (x_6), jumlah posyandu aktif (x_8), jumlah pengguna air bersih (x_{10}) dan sanitasi yang layak (x_{11}).
c. Model Regresi Bertahap
   Model regresi bertahap dihasilkan dari dua jenis proses, yaitu metode seleksi maju dan metode eliminasi mundur. Teknik ini dimulai dengan menambahkan variabel independen ke model regresi, lalu mengeliminasi variabel lainnya setelah variabel

independen sebelumnya ditambahkan. Hasil model regresi bertahap disajikan dalam persamaan berikut:

$$\hat{y}_i = 0{,}021 + 0{,}207x_{i3} + 0{,}461x_{i4} + 0{,}925x_{i5} - 0{,}555x_{i8} - 0{,}393x_{i11}$$

Pada model diatas diperlihatkan bahwa metode regresi bertahap menghasilkan model yang sama dengan model regresi linear dengan metode eliminasi mundur.

3.3 Pemilihan Model Terbaik

Setelah didapatkan model regresi dengan tiga metode berbeda. Kemudian dilakukan pemilihan model terbaik dengan melakukan uji asumsi klasik untuk mendeteksi kemurnian dari model yang diperoleh. Hasil uji asumsi klasik ditampilkan pada Tabel 3 dan Tabel 4 sebagai berikut:

Tabel 3. Uji Normalitas, Homokedastisitas, dan Autokorelasi Ketiga Model

	Model RL dengan Seleksi Maju	**Model RL dengan Eliminasi Mundur**	**Model Regresi Bertahap**
Pengujian	p_{value}	p_{value}	p_{value}
Normalitas	0,318	0,416	0,416
Homokedastisitas	0,479	0,207	0,207
Autokorelasi	0,105	0,171	0,171

Tabel 4. Uji Multikolinearitas Ketiga Model

Model RL dengan Metode Seleksi Maju		**Model RL dengan Metode Eliminasi Mundur**		**Model Regresi Bertahap**	
Variabel	**VIF**	**Variabel**	**VIF**	**Variabel**	**VIF**
x_3	1,420	x_3	1,324	x_3	1,324
x_4	2,816	x_4	2,801	x_4	2,801
x_5	10,879	x_5	5,993	x_5	5,993
x_6	11,080	x_8	1,995	x_8	1,995
x_8	3,177	x_{11}	4,582	x_{11}	4,582
x_{10}	1,825				
x_{11}	5,548				

Pada Tabel diperoleh bahwa ketiga model regresi memiliki nilai p_{value} lebih besar dari $\alpha = 0{,}05$ yang menunjukkan ketiga model memenuhi normalitas, homokedastisitas, dan tidak adanya autokorelasi. Sementara itu, hasil uji multikolinearitas model regresi linear dengan metode seleksi maju terdapat nilai $VIF > 10$, yaitu x_5 dan x_6. Ini menandakan bahwa model tersebut masih mengandung multikolinearitas. Sedangkan model regresi linear dengan metode eliminasi mundur dan model regresi bertahap bebas dari multikolinearitas, ditunjukkan dengan nilai VIF setiap variabel lebih kecil dari sepuluh.

Berdasarkan hasil uji asumsi klasik pada ketiga model maka terpilih model regresi terbaik, yaitu regresi linear dengan metode eliminasi mundur dan model regresi bertahap untuk menduga jumlah gizi buruk pada balita di Kabupaten/Kota di Jawa Tengah. Dalam penelitian Pasaribu (2024), metode regresi dengan eliminasi mundur lebih efisien dan tidak

membutuhkan waktu komputasi yang lama dibandingan dengan metode regresi bertahap. Ini disebabkan metode regresi bertahap dijalankan dengan dua jenis teknik, yaitu teknik eliminasi mundur dan seleksi maju. Oleh karena itu, pada kasus ini dipilih model regresi dengan eliminasi mundur sebagai model terbaik.

Dalam mengukur seberapa baik model dalam menduga jumlah gizi buruk pada balita terhadap data aktualnya maka digunakan indikator koefisien determinasi (R^2) dan simpangan baku residual (RMSE). Model regresi dengan eliminasi mundur menghasilkan nilai R^2 sebesar 60%, artinya sebesar 60% keragaman dari jumlah gizi buruk pada balita di Kabupaten/ Kota di Jawa Tengah dijelaskan oleh model sedangkan sisanya dijelaskan oleh faktor lain diluar model. Sementara itu, nilai RMSE yang dihasilkan model sebesar 0,191 yang berarti bahwa tingkat kesalahan rata-rata hasil prediksi model terhadap data aktualnya hanya 0,191 (19,1%).

Kemudian menyelidiki ada atau tidaknya pengaruh seluruh variabel independen terhadap jumlah gizi buruk pada balita pada model melalui uji F. Hasil uji F diperoleh nilai $p_{value} = 3{,}95 \times 10^{-5}$ lebih kecil dari $\alpha = 0{,}05$. Hal ini menjelaskan bahwa ada pengaruh yang signifikan secara simultan antara variabel independen dengan jumlah gizi buruk pada balita di Kabupaten/ Kota di Jawa Tengah. Selanjutnya menggali informasi variabel-variabel independen yang berpengaruh pada model terhadap jumlah gizi buruk pada balita pada model dengan uji t. Hasil uji t diperlihatkan pada Tabel 5.

Tabel 5. Uji Signifikasi Parameter Parsial

Variabel	Parameter	p_{value}
x_3	0,207	0,109
x_4	0,461	0,049
x_5	0,925	0,003
x_8	−0,555	0,001
x_{11}	−0,393	0,139

Pada Tabel 5 diperlihatkan bahwa nilai $p_{value} < 0{,}05$ yaitu x_4, x_5, x_8 yang diartikan jumlah penduduk miskin (x_4), pengelolaan air minum dan makanan (x_5), dan jumlah posyandu aktif (x_8) masing-masing memberikan pengaruh signifikan terhadap jumlah gizi buruk pada balita di Kabupaten/ Kota di Jawa Tengah. Sedangkan variabel pemberian ASI eksklusif (x_3) dan sanitasi layak (x_5) tidak memberikan pengaruh yang cukup kuat pada model regresi.

4. KESIMPULAN

Analisis faktor-faktor yang memengaruhi jumlah gizi buruk pada balita di Kabupaten/ Kota di Jawa Tengah dilakukan dengan metode regresi linear berganda. Terdapat sebelas variabel indepeden yang diduga mempengaruhi jumlah gizi buruk pada balita di Kabupaten/ Kota di Jawa Tengah. Hasil Model regresi linear berganda menunjukkan adanya multikolinearitas. Oleh karena itu, dilakukan pemilihan variabel independen terbaik dengan teknik metode regresi dengan seleksi maju, metode regresi dengan eliminasi mundur, dan regresi bertahap. Hasil uji ketiga metode tersebut menghasilkan tiga model regresi. Model regresi dengan seleksi maju masih mengandung multikolinearitas. Sementara itu, tidak ada perbedaan antara model regresi dengan eliminasi mundur dengan model regresi bertahap. Kedua model

tersebut memenuhi uji asumsi klasik, yaitu normalitas, homokedastisitas, tidak adanya autokorelasi, dan tidak mengandung multikolinearitas. Model regresi dengan eliminasi mundur dipilih sebagai model terbaik karena model tersebut lebih efisien dan tidak membutuhkan waktu komputasi yang lama dibandingan dengan metode regresi bertahap. Model regresi dengan eliminasi mundur menghasilkan nilai R^2 sebesar 60% dan RMSE sebesar 19,1%. Variabel yang berpengaruh signifikan terhadap jumlah gizi buruk pada balita di Kabupaten/ Kota di Jawa Tengah yaitu jumlah penduduk miskin, pengelolaan air minum dan makanan, dan jumlah posyandu aktif.

5. DAFTAR PUSTAKA

Adrianto, S., Balqis, I. H. N., Soetanto, C. Z. N., & Ohyver, M. (2023). Cochrane orcutt method to overcome autocorrelation in modeling factors affecting the number of hotel visitors in Indonesia. *Procedia Computer Science*, *216*, 630-638.

Al-Dairi, M., Pathare, P. B., Al-Yahyai, R., Al-Habsi, N., Jayasuriya, H., & Al-Attabi, Z. (2024). Machine vision system combined with multiple regression for damage and quality detection of bananas during storage. *Applied Food Research*, *4*(2), 1-14.

Alin, A. (2010). Multicollinearity. *Wiley interdisciplinary reviews: computational statistics*, *2*(3), 370-374.

Andana, A. P., Safitri, D., & Rusgiyono, A. (2017). Model regresi menggunakan least absolute shrinkage and selection operator (LASSO) pada data banyaknya gizi buruk kabupaten/kota di Jawa Tengah. *Jurnal Gaussian*, *6*(1), 21-30.

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.

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*, 1-8.

Cozmuta, L. M. (2025). The application of multiple linear regression methods to FTIR spectra of fingernails for predicting gender and age of human subjects. *Heliyon*, *11*(4), 1 -13.

Đalić, 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.

Fariha, N. F., & Subekti, R. (2018). Pemilihan model regresi terbaik dalam kasus pengaruh premi, klaim, hasil investasi dan hasil underwriting terhadap laba asuransi jiwa (studi kasus PT. Asuransi Jiwasraya (persero)). *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya*, 674-684.

Fiola, E., Yulius, F., Presilia, P., Risani, D. M., Alvionita, M., & Irawati, F. D. 2024. Metode seleksi variabel dalam pemodelan regresi linear data curah hujan Provinsi Lampung. *Prosiding Seminar Nasional Sains Data*, *4*(1), 351-366.

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

Lewis-Beck, M. S., & Skalaban, A. (1990). The R-squared: Some straight talk. *Political Analysis*, *2*, 153-171.

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.

Maulani, A., Herrhyanto, N., & Suherman, M. (2016). Aplikasi model geographically weighted regression (GWR) untuk menentukan faktor-faktor yang mempengaruhi kasus gizi buruk anak balita di Jawa Barat. *Jurnal EurekaMatika*, *4*(1), 46-63.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*, 673-690.

Pasaribu, S. G. (2024). *Analisis perbandingan model regresi linier metode forward selection, backward elimination, dan stepwise regression (studi kasus pendapatan oli mesin pada UD. Bengkel Mobil Abadi).* (Doctoral dissertation, Universitas Sumatera Utara).

Pratnyaningrum, N., Yasin, H., & Hoyyi, A. (2015). Pemodelan persentase balita gizi buruk di Jawa Tengah dengan Pendekatan geographically weighted regression principal components analysis (GWRPCA). *Jurnal Gaussian*, *4*(2), 171-180.

Shafiee, S., Lied, L. M., Burud, I., Dieseth, J. A., Alsheikh, M., & Lillemo, M. (2021). Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture*, *183*, 106036.

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.

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.

Wang, Q., Koval, J. J., Mills, C. A., & Lee, K. I. D. (2007). Determination of the selection statistics and best significance level in backward stepwise logistic regression. *Communications in Statistics-Simulation and Computation*, *37*(1), 62-72.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, *70*(2), 129-133.