Jurnal Matematika Integratif. p-ISSN:1412-6184, e-ISSN:2549-903
Vol. 21, No. 1 (2025), pp. 89–102. doi:10.24198/jmi.v21.n1.62937.89-102

Regresi Logistik Multinomial Bayesian dengan Algoritma *Gibbs Sampling* untuk Menentukan Faktor-Faktor Tingkat Kemiskinan di Indonesia

Hani Syifana, Nurul Gusriani, Kankan Parmikanti

Departemen Matematika, Fakultas MIPA, Universitas Padjadjaran
Jl. Ir. Soekarno Km 21 Jatinangor Sumedang 45363
Email: hani21003@mail.unpad.ac.id

Abstrak
Kemiskinan adalah keadaan serba kekurangan yang dialami oleh individu atau kelompok dengan pengeluaran per kapita bulanan yang tidak mencukupi untuk memenuhi kebutuhan dasar. Berdasarkan profil kemiskinan Indonesia yang dirilis oleh Badan Pusat Statistik (BPS) pada Maret 2024, tercatat bahwa 9,03% penduduk Indonesia dinyatakan sebagai penduduk miskin, dimana angka ini masih terhitung jauh dari target penurunan kemiskinan ke angka 6,5% sampai dengan 7,5% yang ditargetkan dalam Rencana Pembangunan Jangka Menengah Nasional Tahun 2020-2024. Salah satu usaha yang dapat dilakukan untuk mengakhiri kemiskinan di Indonesia adalah dengan menganalisis faktor apa saja yang memengaruhi tingkat kemiskinan. Metode yang digunakan dalam penelitian ini adalah regresi logistik multinomial Bayesian dengan menggunakan algoritma *Markov Chain* Monte Carlo (MCMC) *Gibbs Sampling*. Variabel respon yang digunakan sebagai pengukur tingkat kemiskinan merupakan garis kemiskinan yang merupakan indikator resmi bersumber dari BPS. Hasil penelitian menunjukkan bahwa setelah 20.000 iterasi, rantai Markov mencapai keadaan stasioner. Hasil uji interval kredibel yang didukung oleh hasil uji *deviance* menyatakan bahwa faktor-faktor yang berpengaruh signifikan terhadap tingkat kemiskinan di Indonesia tahun 2024 adalah Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan dan rata-rata lama sekolah.
Kata kunci: Garis Kemiskinan, Regresi Logistik Multinomial, Bayesian, *Markov Chain* Monte Carlo, *Gibbs Sampling*.

2000 Mathematics Subject Classification:
Received: 27-04-2025, accepted: 17-05-2025.

Abstract
Poverty is a state of deprivation experienced by individuals or groups with monthly per capita expenditure that is insufficient to meet basic needs. Based on the Indonesian poverty profile released by the Statistics Indonesia (BPS) in March 2024, it was recorded that 9.03% of the Indonesian population was declared poor, which is still far from the poverty reduction target of 6.5% to 7.5% targeted in the National Medium-Term Development Plan 2020-2024. One of the efforts that can be made to end poverty in Indonesia is to analyze what factors affect the poverty rate. The method used in this research is Bayesian multinomial logistic regression by using the algorithm Markov Chain Monte Carlo (MCMC) algorithm. The response variable used as a measure of poverty level is the poverty line, which is an official indicator sourced from BPS. The results showed that after 20,000 iterations, the Markov chain reached a stationary state. The results of the credible interval test supported by the results of the deviance test state that the factors that significantly affect the poverty rate in Indonesia in 2024 are Gross Regional Domestic Product (GRDP) at constant prices and average years of schooling.

Keywords: *Poverty Line, Multinomial Logistic Regression; Bayesian; Markov Chain Monte Carlo; Gibbs Sampling*

1. Pendahuluan

Badan Pusat Statistik Indonesia mendefinisikan kemiskinan sebagai keadaan serba kekurangan yang dialami oleh individu atau kelompok dengan pengeluaran per kapita bulanan yang tidak mencukupi untuk memenuhi kebutuhan dasar. Untuk mengukur tingkat kemiskinan, Indonesia menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*) yang lebih dikenal dengan istilah garis kemiskinan. BPS menyatakan bahwa garis kemiskinan dapat digunakan sebagai indikator tingkat kemiskinan dengan standar untuk penentuan status miskin penduduk adalah sebesar Rp582.932 per kapita per bulan.

Urgensi untuk memberantas kemiskinan merupakan tujuan utama di antara tujuh belas tujuan dan sasaran yang menjadi komitmen global dan nasional dalam *Sustainable Development Goals* (SDGs). Hal ini ditegaskan pula dalam Rencana Pembangunan Jangka Menengah Nasional Tahun 2020-2024, dengan target penurunan kemiskinan di Indonesia adalah sebesar 6,5% sampai dengan 7,5%. Namun, pada kenyataannya, berdasarkan profil kemiskinan Indonesia yang dirilis oleh BPS pada Maret 2024, tercatat bahwa 9,03% atau sekitar 25,22 juta penduduk Indonesia dinyatakan sebagai penduduk miskin. Ini menunjukkan bahwa Indonesia belum bisa mencapai target penurunan kemiskinan tersebut. Salah satu upaya yang bisa dilakukan untuk mencapai target penurunan kemiskinan tersebut adalah dengan mengidentifikasi faktor apa saja yang memengaruhinya. Untuk mengidentifikasi faktor-faktor yang memengaruhi tingkat kemiskinan, dapat digunakan analisis regresi. Namun, faktor-faktor signifikan yang memengaruhi tingkat kemiskinan yang tinggi dan tingkat kemiskinan yang rendah dapat berbeda. Hal ini dapat diatasi dengan mengategorikan garis kemiskinan ke dalam beberapa kategori. Dengan demikian, salah satu jenis analisis regresi yang dapat diterapkan untuk masalah ini adalah regresi logistik multinomial yang cocok digunakan untuk menangani masalah dengan variabel respon berupa kategori.

Pada data sosial-ekomomi seperti tingkat kemiskinan, sangat rentan terjadi multikolinearitas. Salah satu metode yang dapat diterapkan untuk mengatasi masalah multikolinearitas ini adalah digunakannya pendekatan Bayesian [12]. Penelitian yang menggunakan regresi logistik multinomial Bayesian pun telah banyak dilakukan. Beberapa di antaranya adalah yang dilakukan oleh Apsari, Yasin, dan Sugito [3] serta Cahyani, Goejantoro, dan Siringoringo [6]. Apsari, Yasin, dan Sugito [3] menggunakan metode Bayesian dengan algoritma Markov *Chain* Monte Carlo (MCMC) *Metropolis-Hastings* untuk mengestimasi parameter regresinya. Uji konvergensi rantai Markov-nya dilakukan menggunakan *trace plot*, *ergodic mean plot*, dan *autocorrelation plot*. Sedangkan Cahyani, Goejantoro, dan Siringoringo [6] menggunakan metode Bayesian dengan algoritma MCMC Gibbs Sampling, sedangkan metode pengujian konvergensi rantai Markov-nya dilakukan menggunakan *trace plot, autocorrelation plot*, dan *running quantiles plot*. Dua penelitian tersebut hanya menggunakan analisis *plot* untuk pengujian konvergensi estimasi parameternya, sedangkan *plot* hanya memberikan visualisasi kasar terkait stabilitas rantai Markov yang dihasilkan. Tidak hanya itu, setelah dilakukan uji signifikansi, variabel yang tidak signifikan tidak dikeluarkan dari model. Hal ini dapat menyebabkan model menjadi kurang baik dalam merepresentasikan data.

Pada penelitian ini, tingkat kemiskinan di Indonesia dimodelkan menggunakan regresi logistik multinomial Bayesian menggunakan algoritma MCMC *Gibbs Sampling*. Dipilihnya algoritma MCMC *Gibbs Sampling* adalah karena algoritma ini dinilai lebih efisien daripada algoritma MCMC *Metropolis-Hastings* [10]. Pengujian konvergensi rantai Markov dilakukan menggunakan *trace plot, autocorrelation plot*, dan dilengkapi dengan metode statistik yakni uji Geweke. Setelah itu, dilakukan uji interval kredibel untuk menganalisis variabel prediktor yang memberikan pengaruh signifikan terhadap model. Berbeda dari yang dilakukan oleh [6] dan [3], variabel yang tidak signifikan dikeluarkan dari model sehingga diperoleh model yang hanya berisi variabel yang signifikan saja. Kemudian, sebagai pengujian model terbaik, dilakukan uji *deviance* untuk menguji dua model antara model yang berisi semua variabel prediktor dan model yang hanya berisi variabel yang signifikan saja untuk menentukan model yang lebih akurat dalam merepresentasikan data.

2. Metode Penelitian

2.1. **Objek dan Variabel Penelitian.** Objek yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Indonesia (BPS) mengenai tingkat kemiskinan di Indonesia berdasarkan provinsi pada tahun 2024. Tingkat kemiskinan diukur menggunakan garis kemiskinan dengan enam variabel prediktor, meliputi, laju pertumbuhan penduduk (X_1), tingkat pengangguran terbuka (X_2), laju pertumbuhan ekonomi yang diukur dengan Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan (X_3), rasio gini (X_4), rata-rata lama sekolah (X_5), dan upah minimum provinsi (X_6).

Pada penelitian ini, tingkat kemiskinan dibedakan menjadi tiga kategori, yaitu kategori normal yang diberi kode 0, di bawah standar garis kemiskinan yang diberi kode 1, dan di atas standar garis kemiskinan yang diberi kode 2. Berdasarkan pernyataan Badan Pusat Statistik Indonesia, suatu individu dinyatakan miskin apabila pengeluarannya di bawah Rp582.932 per bulan. Namun, tidak ada sumber kredibel yang mengategorikan garis kemiskinan ke dalam tiga kategori, sehingga diperlukan penggolongan data ke dalam tiga kategori. Tabel 1 menunjukkan kategorisasi data garis kemiskinan mengacu pada [4] berbasis statistik dengan tetap mempertimbangkan standar garis kemiskinan yang ditetapkan oleh BPS:

Tabel 1. Kategorisasi variabel respon

Kode	Keterangan	
0	Normal	Rp582.932 $\leq y \leq \hat{\mu} + \hat{\sigma}$
1	Di bawah standar garis kemiskinan	$y <$ Rp582.932
2	Di atas standar garis kemiskinan	$\hat{\mu} + \hat{\sigma} \leq y$

dengan $\hat{\mu}$ dan $\hat{\sigma}$ berturut-turut menyatakan rata-rata dan standar deviasi garis kemiskinan di Indonesia.

2.2. **Metode Penelitian.** Metode yang digunakan pada penelitian ini adalah regresi logistik multinomial dengan metode pengestimasian parameter menggunakan pendekatan Bayes.

2.2.1. *Regresi Logistik Multinomial.* Regresi logistik multinomial adalah model regresi logistik untuk memprediksi nilai variabel respon yang lebih dari dua kategori atau bersifat polikotomus [11]. Untuk variabel respon dengan kategori sebanyak J, dibutuhkan sebanyak $J-1$ fungsi logit dengan satu kategori lainnya yang digunakan sebagai pembanding atau referensi. Misalkan, jika variabel respon Y mempunyai kategori (dinotasikan dengan j) sebanyak J yang diberikan kode $0, 1, 2, \ldots, J-2$ dan $J-1$, dan akan digunakan $Y = 0$ sebagai pembanding (kategori referensi), maka [11] mendefinisikan dua fungsi logit seperti pada persamaan (1)

$$g_j(\mathbf{x}) = \ln\left(\frac{P(Y=j)}{P(Y=0)}\right) = \beta_{j0} + \sum_{k=1}^{K} \beta_{jk} X_k \tag{1}$$

dengan $k = 1, 2, 3, \ldots, K$ menunjukkan variabel prediktor, dan peluang kejadian dari tiap kategori ditunjukkan oleh persamaan (2).

$$\pi_j(\mathbf{x}) = P(Y=j) = \frac{e^{(g_j(\mathbf{x}))}}{1 + \sum_{j=1}^{J-1} e^{(g_j(\mathbf{x}))}} \tag{2}$$

2.2.2. *Pendekatan Bayes.* Metode Bayes merupakan metode estimasi parameter yang menggunakan penggabungan data sampel dengan data prior (informasi masa lalu) lalu hasilnya dinyatakan dengan distribusi posterior [3]. Distribusi *prior* adalah informasi awal dari data yang bersifat subjektif, sedangkan distribusi *posterior* merupakan keyakinan yang sudah diperbarui setelah dilakukan pengolahan data menggunakan keyakinan awal (distribusi *prior*) dan data yang diobservasi (fungsi *likelihood*). Bentuk analitik yang rumit dari distribusi *posterior* dalam konteks estimasi parameter regresi mengharuskan dilakukannya metode simulasi dan salah satunya yang dapat diterapkan adalah simulasi Markov *Chain* Monte Carlo (MCMC) [8].

2.2.3. *Algoritma Gibbs Sampling.* MCMC adalah metode simulasi untuk mendapatkan data sampel suatu variabel acak dengan teknik sampling berdasarkan sifat rantai Markov yang berguna untuk mengintegrasikan fungsi yang kompleks atau untuk menangani peluang distribusi yang kompleks [10]. Metode ini dilakukan dengan cara membangun rantai Markov yang konvergen terhadap distribusi target, yaitu distribusi *posterior* dari parameter yang ditaksir. Rantai Markov yang dihasikan memuat ruang keadaan kontinu berupa nilai *posterior* dengan ruang parameter diskrit. Algoritma ini dilakukan secara iteratif hingga diperoleh rantai Markov yang stasioner.

Ada beberapa metode simulasi MCMC yang dapat digunakan, salah satu yang paling populer di antaranya adalah algoritma *Gibbs Sampling.* Menurut Casella dan George [7], *Gibbs Sampling* adalah teknik untuk membangkitkan variabel acak dari suatu distribusi secara tidak langsung tanpa perlu menghitung fungsi densitasnya.

Hanada dan Matsuura [10] meringkas algoritma *Gibbs Sampling* dalam langkah-langkah sebagai berikut:

(1) Diberikan nilai awal $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_K^{(0)})$
(2) Untuk iterasi $t = 1, 2, 3, \ldots, T^*$, parameter $\theta^{(t+1)}$ dibangkitkan dari fungsi densitas bersyarat θ, yaitu:
$$f(\theta_k \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, , \theta_{k-1}^{(t+1)}, \theta_{k+1}^{(t)}, , \theta_K^{(t)})$$
sehingga diperoleh parameter yang sudah diperbarui menjadi:

$$\boldsymbol{\theta}^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, \ldots, \theta_K^{(t+1)})$$

Fungsi densitas bersyarat $f(\theta_k \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, , \theta_{k-1}^{(t+1)}, \theta_{k+1}^{(t)}, , \theta_K^{(t)})$ merupakan fungsi densitas dari distribusi kondisional masing-masing parameter yang diperoleh dengan menurunkan distribusi *posterior*-nya. Pada kasus pengestimasian parameter dengan *prior* konjugat, bentuk distribusi kondisional akan mudah diidentifikasi karena distribusi *posterior* akan mengikuti distribusi *prior*. Namun, apabila *prior* dan *likelihood* tidak konjugasi, bentuk distribusi kondisional akan sulit diidentifikasi karena tidak memiliki bentuk analitis, sehingga diperlukan pendekatan lain untuk *sampling* parameter. Salah satu metode yang dapat dilakukan untuk mengatasi masalah ini adalah dengan menggunakan metode *Metropolis-Hastings* dalam *Gibbs Sampling* atau lebih dikenal dengan algoritma *Metropolis-within-Gibbs*.

2.2.4. *Algoritma Metropolis-within-Gibbs.* Algoritma *Metropolis-within-Gibbs* merupakan algoritma yang mengintegrasikan proses *Metropolis-Hastings* ke dalam Gibbs Sampling. Metode ini mengganti langkah *sampling* langsung dari distribusi kondisional pada *Gibbs Sampling* dengan proses *Metropolis-Hastings*. Tahapan iterasi *Metropolis-within-Gibbs* adalah sebagai berikut [14]:

(1) Inisialisasi nilai awal $\boldsymbol{\beta}^{(0)}$ dan distribusi proporsal, yaitu distribusi probabilitas yang digunakan untuk usulan nilai baru pada setiap iterasi.
(2) Untuk iterasi $t = 1, 2, 3, ..., T^*$, lakukan:
  (a) Definisikan $\hat{\boldsymbol{\beta}}^{(t)} = \hat{\boldsymbol{\beta}}^{(t-1)}$
  (b) Hitung usulan nilai baru untuk $\hat{\beta_{jk}}^{(t)}$, yaitu:

$$\hat{\beta_{jk}}^{(t)} = \hat{\beta_{jk}}^{(t-1)} + \epsilon \tag{3}$$

  dengan ϵ merupakan hasil pembangkitan nilai dari distribusi proporsal.
  (c) Definisikan $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}^{(t)}$ dan memperbarui elemen $\hat{\beta_{jk}}$ pada $\hat{\boldsymbol{\beta}}^*$ dengan hasil perhitungan $\hat{\beta_{jk}}^{(t)}$ pada persamaan (3).
  (d) Hitung rasio penerimaan $a = \min(1, A)$ dengan A ditunjukkan oleh persamaan (4)

$$A = \log \frac{L(\hat{\boldsymbol{\beta}}^*) f(\hat{\boldsymbol{\beta}}^*)}{L(\hat{\boldsymbol{\beta}}^{(t)}) f(\hat{\boldsymbol{\beta}}^{(t)})} \tag{4}$$

  dengan:
  $L(\hat{\boldsymbol{\beta}}^*)$: Fungsi *likelihood* untuk $\hat{\boldsymbol{\beta}}^*$
  $f(\hat{\boldsymbol{\beta}}^*)$: Fungsi densitas distribusi *prior* untuk $\hat{\boldsymbol{\beta}}^*$
  $L(\hat{\boldsymbol{\beta}}^{(t)})$: Fungsi *likelihood* untuk $\hat{\boldsymbol{\beta}}^{(t)}$
  $f(\hat{\boldsymbol{\beta}}^{(t)})$: Fungsi densitas distribusi *prior* untuk $\hat{\boldsymbol{\beta}}^{(t)}$
  (e) Bangkitkan nilai $u \sim U(0, 1)$.
  (f) Definisikan $U = \log(u)$.
  (g) Jika $U \leq a$, maka usulan nilai baru untuk $\hat{\beta_{jk}}^{(t)}$ pada persamaan (5) diterima dan perbarui vektor parameter menjadi $\hat{\boldsymbol{\beta}}^{(t)} = \hat{\boldsymbol{\beta}}^*$. Apabila sebaliknya, maka tetapkan $\hat{\beta_{jk}}^{(t)} = \hat{\beta_{jk}}^{(t-1)}$.
  (h) Ulangi tahap (b) sampai (g) untuk semua parameter $\hat{\beta_{jk}}$, dengan $j = 1, 2, 3, ..., J-1$ dan $k = 1, 2, 3, ..., K$

2.2.5. *Pengujian Konvergensi Rantai Markov.* Untuk melihat konvergensi dari rantai Markov yang telah diperoleh, dapat dilakukan:

(1) *Trace Plot*
Trace Plot adalah gambaran sebuah *plot* dari nilai estimasi parameter terhadap iterasinya. Trend naik turun nilai parameter pada *trace plot* menunjukkan bahwa

burn-in period belum selesai dan hal ini mengindikasikan bahwa konvergensi belum tercapai. Rantai Markov dikatakan stasioner apabila sebaran *plot*-nya sudah stabil dengan fluktuasi yang kecil dan membentuk pola pada satu titik [13].

(2) *Autocorrelation plot*
Nilai autokorelasi merupakan nilai yang menyatakan besarnya korelasi antara nilai pada saat ini dengan nilai masa lalunya. Pada *autocorrelation plot* atau grafik autokorelasi, nilai autokorelasi pada lag 1 menyatakan korelasi antara nilai pada saat ini dengan nilai pada satu periode sebelumnya (misalkan, $\theta^{(t)}$ dengan $\theta^{(t-1)}$), nilai autokorelasi pada lag 2 menyatakan korelasi antara nilai pada saat ini dengan nilai pada dua periode sebelumnya (misalkan, $\theta^{(t)}$ dengan $\theta^{(t-2)}$), dan seterusnya. Jika grafik autokorelasi pada lag pertama mendekati 1 dan pada lag selanjutnya terus berkurang menuju 0, maka dapat dikatakan bahwa konvergensi sudah tercapai [13].

(3) Uji Geweke
Uji Geweke dilakukan dengan membandingkan dua nilai pada bagian awal rantai Markov dengan bagian akhir rantai Markov untuk mendeteksi kegagalan konvergensi. Dua bagian dari rantai Markov (setelah menghapus *burn-in period*) diambil, misalkan $\boldsymbol{\theta}_1^{(t)} : t = 1, \dots, T_1$ dan $\boldsymbol{\theta}_2^{(t)} : t = t_a, \dots, T^*$, dengan $1 < T_1 < t_a < T^*$, T_1 merupakan banyaknya elemen pada rantai Markov yang pertama, t_a merupakan indeks iterasi elemen pertama dari rantai Markov yang kedua, dan T^* merupakan banyaknya iterasi keseluruhan rantai Markov. Hipotesis yang digunakan pada uji Geweke adalah sebagai berikut:
H_0 : $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ (Rantai Markov sudah stasioner dan estimasi nilai *posterior* belum konvergen)
H_1 : $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ (Rantai Markov belum stasioner dan estimasi nilai *posterior* sudah konvergen)
dengan statistik uji ditunjukkan oleh persamaan (5)

$$Z = \frac{\bar{\boldsymbol{\theta}_1} - \bar{\boldsymbol{\theta}_2}}{\sqrt{Var(\boldsymbol{\theta}_1) + Var(\boldsymbol{\theta}_2)}} \tag{5}$$

dan kriteria uji:
Tolak H_0 jika $Z > \mathcal{Z}_{(\frac{\alpha}{2})}$.
Untuk pengujjiannya, Geweke menyarankan untuk mengambil 10% bagian rantai Markov paling awal, dan 50% bagian rantai Markov paling akhir [9].

2.2.6. *Uji Interval Kredibel.* Uji interval kredibel adalah pengujian statistik yang digunakan untuk mengukur ketidakpastian yang terkait dengan estimasi parameter. Dengan tingkat signifikansi α, pengujian ini menggambarkan adanya kepercayaan sebesar $(1-\alpha)\times 100\%$ bahwa parameter model berada di dalam interval kredibel. Hipotesis yang digunakan pada uji interval kredibel adalah sebagai berikut:
H_0 : $\beta_k = 0$ (Variabel ke-k tidak berpengaruh signifikan terhadap model)
H_1 : $\beta_k \neq 0$ (Variabel ke-k berpengaruh signifikan terhadap model).
Apabila interval kredibel memuat nilai nol, ada kemungkinan bahwa parameter model adalah nol sehingga parameter tidak signifikan dan tidak memberi pengaruh terhadap model. Dengan demikian, H_0 ditolak apabila interval kredibelnya tidak memuat nilai nol [1].

2.2.7. *Uji Deviance.* Uji *Deviance* merupakan salah satu metode pemilihan model terbaik yang dilakukan dengan melihat nilai *deviance*. Metode ini dilakukan untuk melihat apakah penambahan atau pengurangan variabel prediktor dapat meningkatkan kemampuan model dalam menjelaskan data. Nilai deviance dinyatakan pada persamaan (6) berikut:

$$D = -2 \times (l_A - l_B) \tag{6}$$

dengan l_A merupakan fungsi log-*likelihood* dari model yang hanya memuat variabel signifikan saja dan l_B merupakan fungsi log-*likelihood* dari model yang berisi semua variabel prediktor. Hipotesis yang digunakan pada uji *deviance* adalah sebagai berikut:

H_0 : Model penuh lebih baik dalam merepresentasikan data
H_1 : Model yang hanya memuat variabel signifikan saja lebih baik dalam merepresentasikan data.
dengan kriteria uji:
H_0 ditolak apabila nilai $|D| \leq \mathcal{X}^2_{(\alpha,df)}$ dimana *df* merupakan selisih jumlah parameter antara kedua model.

3. Hasil dan Pembahasan

3.1. **Kategorisasi Tingkat Kemiskinan.** Karena variabel respon memiliki tiga kategori, maka data garis kemiskinan akan diubah dan dikategorikan ke dalam tiga kategori. Kategorisasi garis kemiskinan ini didasarkan pada Tabel 1, mengacu pada [4] dengan tetap mempertimbangkan standar garis kemiskinan di Indonesia pada tahun 2024, yakni Rp582.932 (BPS, 2024). Dari data, diperoleh bahwa rata-rata ($\hat{\mu}$) garis kemiskinan adalah Rp643.734,895 dengan standar deviasi ($\hat{\sigma}$) adalah Rp134.076,187. Dengan demikian, nilai pada variabel garis kemiskinan diubah menjadi:

(1) Dikategorikan normal dan diberi kode 0 apabila garis kemiskinan lebih dari atau sama dengan Rp582.932 dan apabila garis kemiskinan kurang dari Rp777.811,082.
(2) Dikategorikan di bawah standar garis kemiskinan dan diberi kode 1 apabila garis kemiskinan kurang dari Rp582.932.
(3) Dikategorikan di atas standar garis kemiskinan dan diberi kode 2 apabila garis kemiskinan lebih dari Rp777.811,082.

Setelah dilakukan kategorisasi, data variabel respon disajikan menjadi seperti pada Tabel 2.

Tabel 2. Hasil kategorisasi variabel respon

Nomor	Provinsi	Garis kemiskinan	Kategori
1	Aceh	Rp661.227	0
2	Bali	Rp568.510	1
3	Banten	Rp654.213	0
⋮	⋮	⋮	⋮
37	Sumatera Selatan	Rp554.197	1
38	Sumatera Utara	Rp642.243	0

3.2. **Hasil Uji Multikolinearitas.** Pengujian multikolinearitas yang dilakukan terhadap data variabel prediktor dilakukan dengan VIF. Perhitungan nilai VIF tersebut dibantu menggunakan program Python dengan hasil yang diperoleh ditunjukkan oleh Tabel 3.

Tabel 3. Hasil uji multikolinearitas

Variabel	Nilai VIF
X_1	18,36
X_2	17,59
X_3	3,46
X_4	44,61
X_5	70,45
X_6	29,55

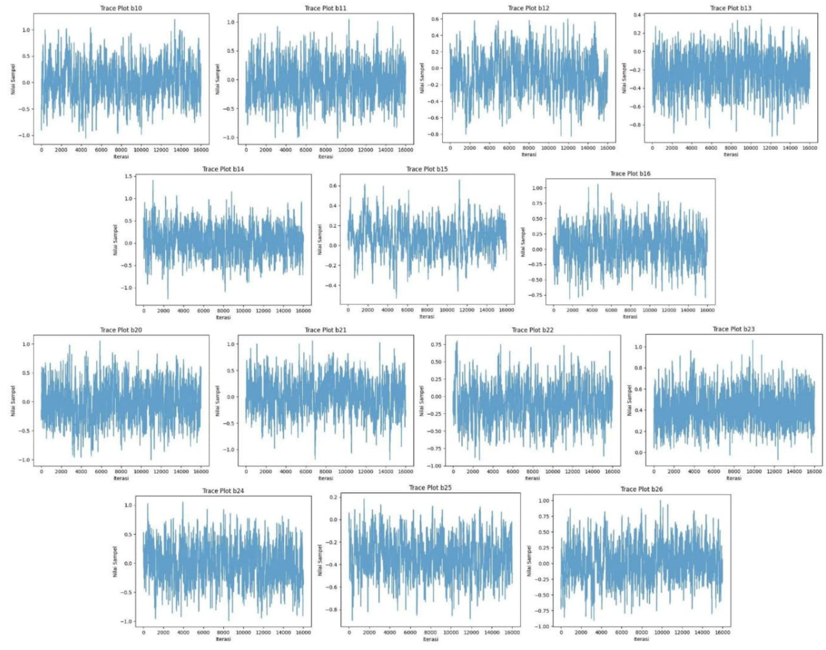
Data mengalami multikolinearitas apabila terdapat satu atau lebih variabel prediktornya yang memiliki nilai VIF lebih dari 10. Pada Tabel 3, nilai VIF semua variabel prediktor kecuali X_3 adalah lebih dari 10, yang artinya data mengalami multikolinearitas. Untuk mengatasi masalah multikolinearitas tersebut, akan digunakan metode Bayesian untuk mengestimasi parameter model regresi logistik multinomial.

3.3. **Estimasi Parameter Model Regresi Logistik Multinomal menggunakan Metode *Maximum Likelihood Estimation* (MLE) dengan Iterasi *Newton-Raphson*.** Langkah ini dilakukan untuk memperoleh estimasi parameter model regresi logistik multinomial menggunakan metode MLE sebagai nilai awal pada algoritma MCMC. Dengan nilai toleransi sebesar 10^{-5}, proses iterasi *Newton-Raphson* mencapai kondisi konvergen setelah 7 iterasi, dengan hasil estimasi parameter ditunjukkan oleh Tabel 4.

Tabel 4. Hasil estimasi parameter model menggunakan MLE

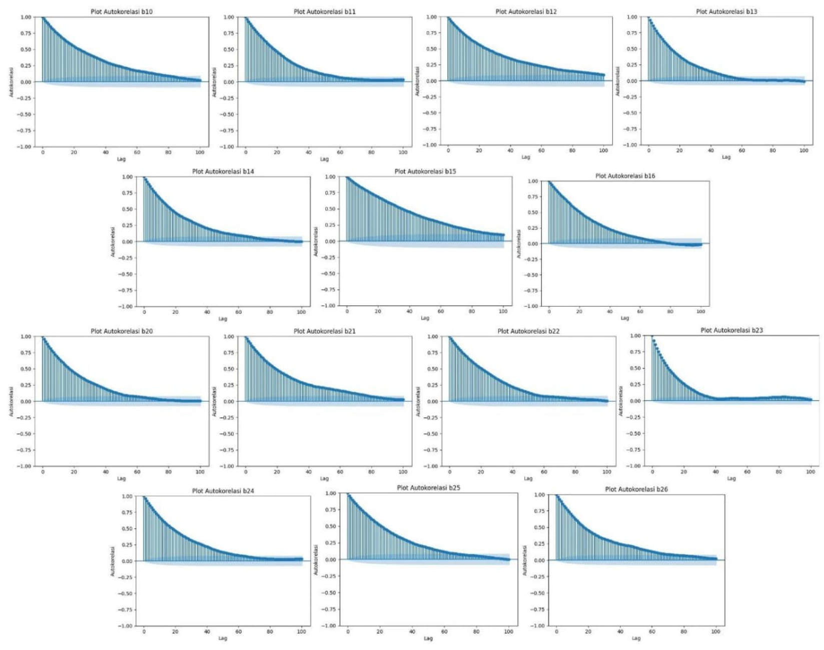
Parameter	Estimasi Parameter
$\hat{\beta}_{10}$	-4,59
$\hat{\beta}_{11}$	-0,82
$\hat{\beta}_{12}$	-0,18
$\hat{\beta}_{13}$	-0,34
$\hat{\beta}_{14}$	29,63
$\hat{\beta}_{15}$	-0,16
$\hat{\beta}_{16}$	-0,46
$\hat{\beta}_{20}$	1,46
$\hat{\beta}_{21}$	1,75
$\hat{\beta}_{22}$	0,31
$\hat{\beta}_{23}$	0,61
$\hat{\beta}_{24}$	-12,61
$\hat{\beta}_{25}$	-1,02
$\hat{\beta}_{26}$	1,14

3.4. **Estimasi Parameter Model Regresi Logistik Multinomial menggunakan Algoritma MCMC *Gibbs Sampling*.** Pada penelitian ini, digunakan *prior* non-konjugat berdistribusi $\mathcal{N}(0, 10^{-1})$. Karena *prior* yang digunakan tidak konjugat dengan fungsi *likelihood* yang berdistribusi multinomial, maka distribusi*posterior* tidak memiliki bentuk analitis dan sulit diidentifikasi bentuk standarnya. Dengan demikian, dilakukan pendekatan lain, yakni dengan melakukan algoritma MCMC *Metropolis-within-Gibbs*. Dengan distribusi proporsal yang berdistribusi $\mathcal{N}(0, 10^{-2})$, rantai Markov mencapai kondisi stasioner setelah dilakukan sebanyak 20.000 iterasi dengan *burn-in period* sebanyak 4000 setelah dilakukan pengujian konvergensi.



Gambar 1. *Trace plot* rantai Markov 20.000 iterasi

Pada Gambar 1, *trace plot* rantai Markov setiap parameter memiliki fluktuasi yang kecil dengan sebaran *plot* yang sudah stabil. Dengan demikian, rantai Markov dikatakan stasioner dengan analisis *trace plot*.



Gambar 2. *Autocorrelation plot* rantai Markov 20.000 iterasi

Pada Gambar 2, *autocorrelation plot* untuk semua parameter menunjukkan grafik yang menurun terus setelah lag ke-1 menuju 0, sehingga berdasarkan analisis *autocorrelation plot*, rantai Markov dinyatakan stasioner.

Selanjutnya, dilakukan pengujian konvergensi menggunakan uji Geweke. Uji Geweke dilakukan pada tiap rantai Markov dengan mengambil 10% bagian awal dan 50% bagian akhir rantai Markov, lalu menghitung nilai Z berdasarkan persamaan (7). Dengan tingkat sigifikansi $\alpha = 0,05$, diperoleh $\mathcal{Z}_{\frac{\alpha}{2}}) = 1,96$.

Tabel 5. Hasil uji Geweke

Parameter	Z	Keputusan
$\hat{\beta_{10}}$	-0,07	Stasioner
$\hat{\beta_{11}}$	-0,03	Stasioner
$\hat{\beta_{12}}$	0,13	Stasioner
$\hat{\beta_{13}}$	0,02	Stasioner
$\hat{\beta_{14}}$	-0,02	Stasioner
$\hat{\beta_{15}}$	-0,12	Stasioner
$\hat{\beta_{16}}$	0,11	Stasioner
$\hat{\beta_{20}}$	0,01	Stasioner
$\hat{\beta_{21}}$	0,18	Stasioner
$\hat{\beta_{22}}$	-0,03	Stasioner
$\hat{\beta_{23}}$	0,01	Stasioner
$\hat{\beta_{24}}$	0,06	Stasioner
$\hat{\beta_{25}}$	0,02	Stasioner
$\hat{\beta_{26}}$	-0,09	Stasioner

Berdasarkan Tabel 5, diperoleh bahwa rantai Markov tiap parameter stasioner dengan hasil estimasi nilai posterior yang konvergen. Dengan demikian, karena pengujian konvergensi menggunakan *trace plot, autocorrelation plot*, dan uji Geweke memberikan hasil bahwa rantai Markov sudah stasioner, maka keadaan stasioner dan konvergen tercapai setelah melakukan iterasi sebanyak 20.000 kali

3.5. **Hasil Uji Signifikansi Parameter.** Untuk menentukan signifikansi masing-masing parameter yang sudah diestimasi sebelumnya, dapat digunakan uji interval kredibel. Suatu parameter dikatakan signifikan terhadap model apabila interval kredibelnya tidak memuat nilai nol. Menggunakan bantuan program Python, hasil uji signifikansi parameter beserta estimasi parameter dengan *mean posterior* disajikan pada Tabel 6.

Tabel 6. Hasil uji interval kredibel

Parameter	*Mean posterior*	Nilai Persentil ke-2,5	9,75	Keputusan
$\hat{\beta_{10}}$	0,04	-0,57	0,68	Tidak Signifikan
$\hat{\beta}_{11}$	-0,03	-0,60	0,54	Tidak Signifikan
$\hat{\beta_{12}}$	-0,09	-0,49	0,36	Tidak Signifikan
$\hat{\beta_{13}}$	-0,21	-0,61	0,15	Tidak Signifikan
$\hat{\beta_{14}}$	0,01	-0,59	0,62	Tidak Signifikan
$\hat{\beta_{15}}$	0,07	-0,21	0,33	Tidak Signifikan
$\hat{\beta_{16}}$	0,06	-0,39	0,56	Tidak Signifikan
$\hat{\beta_{20}}$	-0,01	-0,55	0,52	Tidak Signifikan
$\hat{\beta_{21}}$	0,03	-0,64	0,58	Tidak Signifikan
$\hat{\beta_{22}}$	-0,10	-0,59	0,38	Tidak Signifikan
$\hat{\beta}_{23}$	-0,41	0,17	0,71	Signifikan
$\hat{\beta_{24}}$	-0,02	-0,63	0,59	Tidak Signifikan
$\hat{\beta_{25}}$	-0,33	-0,65	-0,03	Signifikan
$\hat{\beta_{26}}$	-0,04	-0,44	0,57	Tidak Signifikan

Dari Tabel 6 diperoleh bahwa semua parameter tidak signifikan kecuali $\hat{\beta_{23}}$ dan $\hat{\beta_{25}}$, sehingga perlu dilakukan penghapusan variabel X_1, X_2, X_4, dan X_6. Kemudian, lakukan kembali pengestimasian parameter model regresi logistik multinomial menggunakan algoritma MCMC yang dilakukan sebelumnya dengan hanya melibatkan variabel X_3 dan X_5. Proses pengestimasian parameter ulang dengan hanya melibatkan variabel X_3 dan X_5 tersebut berhasil mencapai keadaan konvergen pada iterasi ke-20.000, sehingga menghasilkan 6 buah rantai Markov (untuk parameter $\hat{\beta_{10}}$,$\hat{\beta_{13}}$, $\hat{\beta_{15}}$, $\hat{\beta_{20}}$, $\hat{\beta_{23}}$, dan $\hat{\beta_{25}}$ dengan masing-masing rantai Markov sepanjang 16.000. Uji signifikansi yang dilakukan terhadap model baru tersebut serta estimasi parameternya yang diperoleh dari rantai Markov menggunakan nilai *mean posterior* disajikan dalam Tabel 7.

Tabel 7. Hasil uji interval kredibel untuk model dengan X_3 dan X_5

Parameter	*Mean posterior*	Nilai Persentil ke-2,5	9,75	Keputusan
$\hat{\beta_{10}}$	0,04	-0,57	0,65	Tidak Signifikan
$\hat{\beta_{13}}$	-0,24	-0,62	0,09	Tidak Signifikan
$\hat{\beta_{15}}$	0,06	-0,01	0,24	Tidak Signifikan
$\hat{\beta_{20}}$	-0,03	-0,66	0,64	Tidak Signifikan
$\hat{\beta}_{23}$	0,4	0,14	0,70	Signifikan
$\hat{\beta_{25}}$	-0,35	-0,60	-0,15	Signifikan

Dari Tabel 7 diperoleh hasil yang sama seperti pengetimasian parameter dengan model yang menggunakan semua variabel prediktor, bahwa semua parameter tidak signifikan kecuali $\hat{\beta}_{23}$ dan $\hat{\beta}_{25}$. Pada tahap ini, tidak ada variabel yang dapat dihapus lagi, sehingga disimpulkan bahwa variabel X_3 dan X_5 hanya memengaruhi kategori 2

3.6. **Hasil Uji *Deviance*.** Karena terdapat 2 model, yaitu model penuh yang berisi variabel X_k dengan $k = 1, 2, 3, \ldots, 6$ dan model baru yang hanya berisi variabel signifikan saja, yakni X_3 dan X_5, perlu dilakukan pemilihan model yang lebih baik dalam merepresentasikan data. Pada penelitian ini, digunakan uji deviance yang menggunakan nilai fungsi log-*likelihood* dari kedua model pada nilai deviance diperoleh dari persamaan (8).

Tabel 8. Hasil uji *deviance*

Nilai D	Nilai $\mathcal{X}^2_{(0,05;8)}$
-0,57	2,73

Dari tabel 8 diperoleh bahwa $|D| < \mathcal{X}^2_{(0,05;8)}$, maka H_0 ditolak, dan model yang hanya berisi variabel X_3dan X_5 lebih baik dalam merepresentasikan data. Dengan demikian, model menghasilkan dua fungsi logit, yaitu:

$$g_1(\mathbf{x}) = 0,04 - 0,24X_3 + 0,06X_5 \tag{7}$$

$$g_2(\mathbf{x}) = -0,03 + 0,40X_3 - 0,35X_5 \tag{8}$$

Berdasarkan model di atas, dapat diartikan bahwa apabila nilai PDRB atas dasar harga konstan dan rata-rata lama sekolah bernilai 0, maka peluang suatu provinsi untuk berstatus di bawah garis kemiskinan adalah $e^{\hat{\beta_{10}}} = e^{0,04} = 1,04$ atau 4% lebih besar dibandingkan berstatus normal dan peluang berstatus di atas garis kemiskinan adalah $e^{\hat{\beta_{20}}} = e^{-0,03} = 0,97$ atau 3% lebih kecil dibandingkan berstatus normal. Kemudian, untuk setiap peningkatan PDRB sebesar Rp10 juta, peluang suatu provinsi untuk berstatus di bawah garis kemiskinan adalah $e^{\hat{\beta_{13}}} = e^{-0,24} = 0,79$ atau 21% lebih kecil dibandingkan berstatus normal dan peluang berstatus di atas garis kemiskinan adalah $e^{\hat{\beta_{23}}} = e^{0,39} = 1,48$ atau 48% lebih besar dibandingkan berstatus normal. Ini menunjukkan bahwa pertumbuhan ekonomi berperan dalam menurunkan tingkat kemiskinan. Selain itu, ketika nilai rata-rata lama sekolah meningkat 1 tahun, peluang suatu provinsi untuk berada di bawah standar garis kemiskinan adalah $e^{\hat{\beta_{15}}} = e^{0,06} = 1,06$ atau 6% lebih besar dibandingkan berstatus normal dan peluang berstatus di atas garis kemiskinan adalah sebesar $e^{\hat{\beta_{25}}} = e^{-0,35} = 0,70$ atau 30% lebih kecil dibandingkan berstatus normal.

4. Simpulan

Pada penelitian ini, likelihood berdistribusi multinomial sedangkan prior yang digunakan berdistribusi $\mathcal{N}(0, 10^{-1})$. Karena *prior* dan *likelihood* tidak konjugasi, distribusi kondisional tidak memiliki bentuk analitis dan proses sampling parameter tidak dapat dilakukan dengan algoritma *Gibbs Sampling*. Dengan demikian, digunakan algoritma *Metropolis-within-Gibbs* yang mengintegrasikan proses *Metropolis-Hastings* ke dalam *Gibbs Sampling*. Menggunakan algoritma tersebut, rantai Markov mencapai kondisi stasioner setelah dilakukan iterasi. Selain itu, dengan uji *deviance* diperoleh bahwa model yang lebih baik dalam merepresentasikan data adalah model yang hanya memuat variabel X_3 dan X_5 saja. Model regresi logistik multinomial yang diperoleh ditunjukkan dengan dua fungsi logit pada persamaan (7) dan (8). Selain itu, diperoleh pula faktor-faktor yang berpengaruh secara signifikan terhadap tingkat kemiskinan di Indonesia pada 2024 adalah variabel PDRB atas dasar harga konstan dan rata-rata lama sekolah.

Daftar Pustaka

[1] Agresti, A., *An introduction to categorical data analysis.*,John Wiley & Sons Inc., New Jersey.
[2] Ainul, A.M., Junaidi and Utami, I.T., 2018, Penerapan model analisis regresi linier berganda dengan pendekatan bayesian pada data aset bank di Indonesia, *Jurnal Keteknikan dan Sains (JUTEKS)*, Volume 1, Issue 1, June 2018, Pages 41-47.
[3] Apsari, W., Yasin, H. and Sugito., 2013, estimasi parameter regresi logistik multinomial dengan metode bayes, Jurnal Gaussian, Volume 2, Issue 1, January 2013, Pages 79-88.
[4] Azwar, S., 2012, *Penyusunan skala psikologi*, Pustaka Pelajar, Yogyakarta.
[5] Box, G.E.P. and Tiao, G.C., 1973, *Bayesian inference in statistical analysis (Wiley Classic Library)*, Wiley-Interscience, Massachusetts.

[6] Cahyani, E.T., Goejantoro, R. and Siringoringo, M., 2022, Analisis regresi logistik multinomial bayes untuk pemodelan minat peserta didik MAN 2 Samarinda tahun Ajaran 2018/2019, *Jurnal EKSPONENSIAL*, Volume 13, Issue 1, Mei 2022, Pages 1-8.

[7] Casella, G and George, E.I., 1992, Explaining the Gibbs sampler, *The American Statistician*, Volume 46, Issue 3, August 1992, Pages 167-174.

[8] Dewi, S.S., Resmawan and Nashar, L.O., 2023, analisis regresi logistik multinomial dengan metode bayes untuk identifikasi faktor-faktor terjadinya diabetes melitus, *Journal of Mathematics: Theory and Applications*, Volume 5,Issue 2, October2023, Pages 51-60.

[9] Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation posterior moments, *Bayesian Statistics* Volume 4., Noveber 1992, Pages. 169-193.

[10] Hanada, M. and Matsuura, S., 2022, *MCMC from scratch: a practical introduction to Markov Chain Monte Carlo*, Springer, Singapore.

[11] Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X., 2013, *Applied logistic regression*, John Wiley & Sons Inc, New Jersey.

[12] Jaya, I.G.N.M., Tantular, B., and Andriyana, Y., 2019, A Bayesian approach on multicollinearity problem with an informative prior', *Journal of Physics*, Volume 1265, Issue 1), Pages 1-10.

[13] Johnson, A.A., Ott, M.Q. and Dogucu, M., 2022, *Bayes rules! An introduction to applied Bayesian modeling*, Chapman and Hall/CRC, Boca Raton.

[14] Liu, Q. and Tong, X.T., 20202 'Accelerating Metropolis-within-Gibbs sampler with localizedcomputations of differential equations', *Statistics and Computing*, Volume 30, Fabruary 2020, Pages 1037-1056.

[15] Montgomery, D.C., Peck, E.A. and Vining G.G., 2021, *Introduction to linear regression analysis*,John Wiley & Sons Inc, New Jersey.