

ANALISIS FAKTOR-FAKTOR PENENTU KEMISKINAN DI KALIMANTAN TIMUR:

Perbandingan Model *Weighted Logistic Regression*, *Naive Bayes*, dan
Regresi Binomial

(*Determinants of Poverty in East Kalimantan: A Comparison of Weighted Logistic
Regression, Naive Bayes, and Binomial Regression Models*)

Arifatus Solikhah

BPS Provinsi Kalimantan Timur

E-mail: arifatus.solikhah@bps.go.id

ABSTRAK

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi status kemiskinan rumah tangga di Kalimantan Timur dengan membandingkan tiga model: *Weighted Logistic Regression* (WLR), *Naive Bayes*, dan Regresi Binomial dengan fungsi tautan logit tergeneralisasi (glogit). Data yang digunakan diperoleh dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dilaksanakan pada Maret 2023. Estimasi parameter dilakukan dengan pendekatan Bayesian menggunakan algoritma *Hamiltonian Monte Carlo* (HMC) melalui program RStan. Hasil analisis menunjukkan bahwa jumlah anggota rumah tangga, lokasi tempat tinggal, tingkat pendidikan kepala rumah tangga, status pekerjaan kepala rumah tangga, usia kepala rumah tangga, dan rasio ketergantungan merupakan variabel yang signifikan mempengaruhi status kemiskinan rumah tangga di Kalimantan Timur. Perbandingan kinerja ketiga model menunjukkan bahwa model WLR dan *Naive Bayes* lebih baik dalam mendeteksi rumah tangga miskin dibandingkan model regresi Binomial dengan fungsi tautan glogit, meskipun model regresi Binomial menunjukkan akurasi keseluruhan yang lebih tinggi. Temuan ini memberikan wawasan penting tentang determinan kemiskinan dan efektivitas berbagai model dalam menangani data biner yang tidak seimbang.

Kata kunci: Kemiskinan, *Weighted Logistic Regression*, *Naive Bayes*, Regresi Binomial, Fungsi Tautan Logit Tergeneralisasi, Pendekatan Bayesian, Kalimantan Timur

ABSTRACT

This study aims to analyze the factors influencing household poverty status in East Kalimantan by comparing three models: Weighted Logistic Regression (WLR), Naive Bayes, and Binomial Regression with a generalized logit link function (glogit). The data used were obtained from the National Socioeconomic Survey (SUSENAS) conducted in March 2023. Parameter estimation was performed using a Bayesian approach with the Hamiltonian Monte Carlo (HMC) algorithm through the RStan program. The analysis results indicate that the number of household members, place of residence, education level of household head, employment status of household head, age of household head, and dependency ratio are significant variables affecting household poverty status in East Kalimantan. The comparison of the three models' performance shows that the WLR and Naive Bayes models are better at detecting poor households compared to the Binomial Regression with a generalized logit link function model, despite the Binomial Regression model showing higher overall accuracy. These findings provide important insights into the determinants of poverty and the effectiveness of various models in handling unbalanced binary data.

Keywords: Poverty, *Weighted Logistic Regression*, *Naive Bayes*, *Binomial Regression*, *Generalized Logit Link Function*, *Bayesian Approach*, *East Kalimantan*

PENDAHULUAN

Mengatasi kemiskinan merupakan agenda utama yang harus menjadi prioritas bagi pemerintah pusat maupun daerah. Hal ini tidak hanya penting untuk memenuhi hak asasi manusia seperti akses terhadap pangan, perumahan, pendidikan, dan kesehatan bagi setiap individu, tetapi juga mampu mencegah jutaan kematian yang sebenarnya bisa dihindari, terutama di kalangan anak-anak. Pengentasan kemiskinan juga berperan penting dalam mendorong pertumbuhan ekonomi yang

inklusif dengan meningkatkan daya beli masyarakat miskin. Lebih jauh, upaya ini dapat menciptakan masyarakat yang lebih stabil, damai, dan kohesif dengan mengurangi ketimpangan ekonomi yang tinggi yang seringkali memicu ketidakstabilan sosial dan konflik.

Kemiskinan juga menjadi masalah serius yang perlu penanganan yang komprehensif di Kalimantan Timur. Persentase penduduk di Kalimantan Timur yang hidup di bawah garis kemiskinan menunjukkan penurunan dari 6,23% pada Maret 2015 menjadi 5,91% pada September 2019. Namun, pandemi COVID-19 pada tahun 2020 menyebabkan lonjakan angka kemiskinan hingga 6,64% pada September 2020. Setelah pandemi, angka kemiskinan berfluktuasi, turun ke 6,27% pada September 2021 dan naik lagi ke 6,44% pada September 2022. Pada Maret 2023, angka kemiskinan kembali turun menjadi 6,11%.

Faktor-faktor yang mempengaruhi kemiskinan sangat kompleks dan melibatkan berbagai dimensi. Penelitian menunjukkan bahwa usia, lokasi tempat tinggal di perkotaan atau perdesaan, pendidikan, status pernikahan, disabilitas, remitansi, dan wilayah geografis memainkan peran penting dalam menentukan tingkat kemiskinan (Iqbal dkk., 2020). Selain itu, pertumbuhan ekonomi yang rendah, pengangguran, produktivitas yang rendah, kebijakan makro ekonomi yang tidak tepat, degradasi ekonomi, kejahatan, dan kekerasan juga diidentifikasi sebagai faktor lain yang mendorong kemiskinan (Aderounmu dkk., 2021). Faktor lain seperti rasio ketergantungan (*dependency ratio*), jumlah anggota rumah tangga, tingkat pendidikan kepala rumah tangga, lokasi geografis tempat tinggal, ras kepala rumah tangga, jenis kelamin kepala rumah tangga, status pernikahan kepala rumah tangga, serta status pekerjaan kepala rumah tangga juga turut berpengaruh terhadap kemiskinan (Biyase dan Zwane, 2018).

Model regresi logistik sering digunakan dalam analisis data biner. Tetapi model tersebut memiliki beberapa kelemahan signifikan ketika diterapkan pada data dengan kejadian langka. Seperti halnya kemiskinan di Kalimantan Timur, jumlah rumah tangga miskin jauh lebih sedikit dibandingkan dengan rumah tangga tidak miskin. Menurut King dan Zeng (2001), model regresi logistik cenderung mengabaikan probabilitas kejadian langka, yang dapat menyebabkan bias yang signifikan dalam estimasi probabilitas tersebut. Oleh karena itu, beberapa model yang lebih sesuai untuk menangani data dengan ketidakseimbangan kelas telah dikembangkan, seperti model *Weighted Logistic Regression* (WLR) yang diperkenalkan oleh King dan Zeng (2001), model Regresi Binomial dengan fungsi tautan logit tergeneralisasi yang diperkenalkan oleh Prasetyo, dkk. (2020), dan model *Naive Bayes* yang diperkenalkan oleh Murty dan Devi (2011).

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi kemiskinan di Kalimantan Timur pada Maret 2023 menggunakan model WLR, model *Naive Bayes*, dan model Regresi Binomial dengan fungsi tautan logit tergeneralisasi yang telah disempurnakan. Estimasi parameter dari ketiga model tersebut dilakukan menggunakan metode Bayesian dengan algoritma *Hamiltonian Monte Carlo* (HMC) seperti yang dijelaskan oleh Gelman, dkk. (2014) dan diimplementasikan menggunakan program RStan. Tujuan dari analisis ini adalah untuk memahami lebih dalam tentang faktor-faktor yang berkontribusi terhadap kemiskinan di daerah tersebut dan untuk menguji keefektifan model-model tersebut dalam menangani ketidakseimbangan kelas dalam data.

Penelitian ini juga membandingkan kinerja ketiga model tersebut dalam mengidentifikasi faktor-faktor yang mempengaruhi kemiskinan di Kalimantan Timur. Kinerja model dievaluasi menggunakan ukuran *recall* atau sensitivitas, seperti yang dijelaskan oleh He dan Ma (2013), yang mengukur seberapa baik model dapat mendeteksi observasi positif yang sebenarnya. *Recall* atau sensitivitas yang tinggi menunjukkan bahwa model mampu memprediksi sebagian besar rumah tangga miskin dengan baik, sehingga mengurangi kesalahan prediksi. Variabel bebas yang digunakan dalam penelitian ini meliputi rasio ketergantungan (*dependency ratio*)¹, usia kepala rumah tangga, jumlah

¹ Rasio ketergantungan (*dependency ratio*) adalah ukuran yang menunjukkan jumlah individu yang dianggap tidak produktif secara ekonomi (anak-anak dan orang tua) dibandingkan dengan jumlah individu yang produktif secara ekonomi dalam suatu populasi, dalam hal ini adalah rumah tangga. Rasio ini dihitung dengan membagi jumlah anggota rumah tangga usia non-produktif (mereka yang berusia di bawah 15 tahun dan di atas 64 tahun) dengan jumlah anggota rumah tangga usia produktif (mereka yang berusia 15-64 tahun), lalu dikalikan dengan 100.

anggota rumah tangga, lokasi tempat tinggal, tingkat pendidikan kepala rumah tangga, dan status pekerjaan kepala rumah tangga.

METODOLOGI PENELITIAN

1. Data dan Variabel

Penelitian ini memanfaatkan data mentah yang diperoleh dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dilaksanakan di Kalimantan Timur pada Maret 2023. Variabel dependen yang dianalisis dalam penelitian ini adalah status kemiskinan rumah tangga, sedangkan variabel independennya atau fiturnya mencakup berbagai aspek kondisi sosial, karakteristik demografi, serta lokasi tempat tinggal rumah tangga, dengan rincian sebagai berikut:

Tabel 1. Variabel yang digunakan pada penelitian.

Variabel	Tipe	Keterangan
Y	<i>Dummy</i>	Status kemiskinan rumah tangga; 1 = miskin, 0 = tidak miskin.
X_1	Kontinu	<i>Dependency ratio</i> .
X_2	Kontinu	Usia kepala rumah tangga.
X_3	<i>Dummy</i>	Jumlah anggota rumah tangga; 1 = 5 orang atau lebih, 0 = kurang dari 5 orang.
X_4	<i>Dummy</i>	Lokasi tempat tinggal; 1 = perdesaan, 0 = perkotaan.
X_5	<i>Dummy</i>	Tingkat pendidikan kepala rumah tangga; 1 = SLTP kebawah, 0 = SLTA keatas.
X_6	<i>Dummy</i>	Status pekerjaan kepala rumah tangga; 1 = tidak bekerja atau bekerja di sektor Pertanian, 0 = bekerja di sektor lainnya.

2. Model *Weighted Logistic Regression*

Salah satu masalah utama dalam klasifikasi data adalah ketidakseimbangan data. Dalam klasifikasi dua kelas, masalah ketidakseimbangan terjadi ketika salah satu kelas secara signifikan lebih banyak daripada yang lain. Situasi ini menyebabkan model klasifikasi yang standar seperti model regresi logistik cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas. Model *Weighted Logistic Regression* (WLR) merupakan pengembangan dari regresi logistik yang memperhitungkan bobot untuk mengatasi masalah ketidakseimbangan data. Fungsi *likelihood* dan fungsi *log-likelihood* dari model WLR masing-masing dapat dinyatakan dalam Persamaan (1) dan Persamaan (2) berikut (King dan Zeng, 2001; Sulasih dkk., 2015)

$$L_w = \prod_{i=1}^n \left(\frac{1}{1+e^{-X_i' \beta}} \right)^{w_1 y_i} \left(1 - \left(\frac{1}{1+e^{-X_i' \beta}} \right) \right)^{w_0 (1-y_i)} ; \quad (1)$$

$$\ln L_w = w_1 \sum_{Y_i=1} \ln \left(\frac{1}{1+e^{-X_i' \beta}} \right) + w_0 \sum_{Y_i=0} \ln \left(1 - \left(\frac{1}{1+e^{-X_i' \beta}} \right) \right), \quad (2)$$

dengan n menyatakan total rumah tangga dalam data, w_1 menyatakan bobot untuk rumah tangga miskin, w_0 menyatakan bobot untuk rumah tangga tidak miskin, β merupakan vektor koefisien regresi yang harus diestimasi, $X_i = (1, X_{i1}, X_{i2}, \dots, X_{i6})'$, yang mana 1 merupakan vektor satuan dan X_{ij} menyatakan variabel independen ke- j pada rumah tangga ke- i . Terdapat berbagai metode yang dapat digunakan untuk menentukan bobot dalam model WLR (He dan Garcia, 2009; Kuhn dan Johnson, 2013), salah satunya adalah modifikasi dari bobot yang diusulkan oleh Ting (2002) dalam persamaan berikut:

$$w_0 = \frac{n}{n_0} ; w_1 = \frac{n}{n_1},$$

dengan n_1 menyatakan banyaknya rumah tangga miskin dalam data dan n_0 menyatakan banyaknya rumah tangga tidak miskin.

3. Model *Naive Bayes*

Naive Bayes adalah model klasifikasi probabilistik yang didasarkan pada teorema Bayes dengan asumsi bahwa setiap fitur atau variabel independen dalam model bersifat saling bebas satu sama lain (Murty dan Devi, 2011). Teorema Bayes dapat dinyatakan sebagai berikut:

$$P(Y_i = c | \mathbf{X}_i) = \frac{P(Y_i = c)P(\mathbf{X}_i | Y_i = c)}{P(\mathbf{X}_i)}; i = 1, 2, \dots, n; c = 1, 2,$$

yang mana n menyatakan banyaknya rumah tangga dalam data, $P(Y_i = c | \mathbf{X}_i)$ merupakan probabilitas posterior rumah tangga i berada di kelas c berdasarkan fitur $\mathbf{X}$ dengan $\mathbf{X} = (X_1, X_2, \dots, X_6)$. Simbol $P(\mathbf{X}_i | Y_i = c)$ menyatakan *likelihood*, yaitu probabilitas kondisional fitur $X_j; j = 1, 2, \dots, 6$ dari rumah tangga i berada pada kelas c , dan $P(Y_i = c)$ menyatakan probabilitas *prior* kelas c , serta $P(\mathbf{X}_i)$ merupakan probabilitas fitur $\mathbf{X}$. Dalam model *Naive Bayes*, semua fitur diasumsikan saling bebas, sehingga fungsi *log-likelihood* dari model ini dapat dinyatakan dalam persamaan berikut (Collins, 2012):

$$\ln L_{NB} = \ln P(\mathbf{X} | Y_i = c) = \sum_{i=1}^n \sum_{j=1}^6 \ln P(X_{ij} = x_{ij} | Y_i = c). \quad (3)$$

Dalam penelitian ini, fitur X_1 atau *dependency ratio* dan X_2 atau usia kepala rumah tangga yang bersifat kontinu diasumsikan mengikuti distribusi normal seperti pada Persamaan (4). Sedangkan fitur lainnya yang bersifat biner diasumsikan mengikuti distribusi Bernoulli seperti pada Persamaan (5).

$$P(X_{ij} = x_{ij} | Y_i = c) = \frac{1}{\sqrt{2\pi\sigma_{jc}^2}} e^{-\frac{(x_{ij} - \mu_{jc})^2}{2\sigma_{jc}^2}}; c = 1, 2 \quad (4)$$

dengan μ_{jc} dan σ_{jc}^2 masing-masing menyatakan *mean* dan variansi dari fitur $X_j; j = 1, 2$ pada kelas c , dan

$$P(X_{ij} = x_{ij} | Y_i = c) = (p_{jc})^{x_{ij}} (1 - p_{jc})^{1-x_{ij}}; 0 < p_{jc} < 1, \quad (5)$$

yang mana p_{jc} menyatakan probabilitas kelas c dari fitur X_j . Berdasarkan Persamaan (4) dan Persamaan (5), fungsi *log-likelihood* model *Naive Bayes* yang terdapat pada Persamaan (3) dapat dinyatakan dengan persamaan berikut:

$$\ln L_{NB} = \sum_{i=1}^n \left(- \sum_{j=1}^2 \left(\frac{(x_{ij} - \mu_{jc})^2}{2\sigma_{jc}^2} + \log \sqrt{2\pi\sigma_{jc}^2} \right) + \sum_{j=3}^6 (x_{ij} \log(p_{jc}) + (1 - x_{ij}) \log(1 - p_{jc})) \right). \quad (6)$$

Jika dimisalkan $p_{jc} = \frac{1}{1 + e^{-\gamma_{jc}}}$; $-\infty < \gamma_{jc} < \infty; j = 3, 4, 5, 6$, maka fungsi *log-likelihood* model *Naive Bayes* juga dapat dinyatakan dengan persamaan berikut:

$$\ln L_{NB} = \sum_{i=1}^n \left(- \sum_{j=1}^2 \left(\frac{(x_{ij} - \mu_{jc})^2}{2\sigma_{jc}^2} + \log \sqrt{2\pi\sigma_{jc}^2} \right) + \sum_{j=3}^6 \left(x_{ij} \log \left(\frac{1}{1 + e^{-\gamma_{jc}}} \right) + (1 - x_{ij}) \log \left(1 - \frac{1}{1 + e^{-\gamma_{jc}}} \right) \right) \right). \quad (7)$$

4. Model Regresi Binomial Dengan Fungsi Tautan Logit Tergeneralisasi

Dalam model regresi Binomial, fungsi tautan digunakan untuk menghubungkan variabel independen dengan variabel dependen. Salah satu fungsi tautan yang dapat digunakan dalam regresi Binomial adalah fungsi tautan logit tergeneralisasi (glogit) yang diperkenalkan oleh Prasetyo, dkk. (2020). Fungsi tautan ini memungkinkan untuk menangani data dengan distribusi simetris maupun asimetris, sehingga mampu mengatasi data yang tidak seimbang. Selain itu, fungsi glogit dapat menangani kasus data dengan ekor distribusi yang lebih ringan atau lebih berat dibandingkan distribusi logistik standar.

Prasetyo, dkk. (2020) mengembangkan fungsi tautan glogit dari fungsi kuantil (*quantile function*) distribusi *Exponentiated-Exponential Logistic* (EEL). Namun, fungsi tautan tersebut masih

perlu disempurnakan karena fungsi tersebut berbeda dengan usulan Ghosh dan Alzaatreh (2018). Fungsi kuantil yang didefinisikan oleh Prasetyo, dkk. (2020) adalah sebagai berikut:

$$F_{gl}^{-1}(p) = \mu + s \log \left(1 - \left(1 - p^{\frac{1}{\alpha}} \right)^{\frac{-1}{\lambda}} \right); \quad 1 > p > 0, s > 0, \alpha > 0, \lambda > 0, \infty > \mu > -\infty, \quad (8)$$

yang mana α, λ merupakan parameter bentuk, dan s merupakan parameter skala dari distribusi EEL. Sedangkan Ghosh dan Alzaatreh (2018) mendefinisikan fungsi kuantil distribusi EEL sebagai berikut:

$$F_{\square}^{-1}(p) = s \log \left(\left(1 - p^{\frac{1}{\alpha}} \right)^{\frac{-1}{\lambda}} - 1 \right); \quad 1 > p > 0, s > 0, \alpha > 0, \lambda > 0. \quad (9)$$

Perbedaan fungsi kuantil tersebut menghasilkan fungsi tautan glogit dan model regresi yang berbeda. Prasetyo, dkk. (2020) mengusulkan model regresi binomial dengan fungsi tautan glogit yang dinyatakan sebagai berikut:

$$\text{glogit}(p_i) = \log \left(1 - \left(1 - p_i^{\frac{1}{\alpha}} \right)^{\frac{-1}{\lambda}} \right) = \mathbf{X}'_i \boldsymbol{\beta}; \quad 1 > p_i > 0, \alpha > 0, \lambda > 0, \quad (10)$$

dengan p_i adalah probabilitas rumah tangga i berada di bawah garis kemiskinan, $\mathbf{X}_i = (1, X_{i1}, X_{i2}, \dots, X_{i6})'$ dan $\boldsymbol{\beta}$ adalah vektor koefisien regresi yang harus diestimasi. Jika fungsi kuantil dari distribusi EEL didefinisikan seperti pada Persamaan (9), maka diperoleh model regresi binomial dengan fungsi tautan glogit yang telah disempurnakan sebagai berikut:

$$\text{glogit}(p_i) = \log \left(\left(1 - p_i^{\frac{1}{\alpha}} \right)^{\frac{-1}{\lambda}} - 1 \right) = \mathbf{X}'_i \boldsymbol{\beta}; \quad 1 > p_i > 0, \alpha > 0, \lambda > 0. \quad (11)$$

Berdasarkan Persamaan (11) diperoleh persamaan inversi-glogit sebagai berikut:

$$p_i = F(p_i) = \left(1 - \left(\frac{1}{1 + e^{\mathbf{X}'_i \boldsymbol{\beta}}} \right)^{\lambda} \right)^{\alpha} = \left(1 - \left(\frac{e^{-\mathbf{X}'_i \boldsymbol{\beta}}}{1 + e^{-\mathbf{X}'_i \boldsymbol{\beta}}} \right)^{\lambda} \right)^{\alpha} = \left(1 - \left(1 - \frac{1}{1 + e^{-\mathbf{X}'_i \boldsymbol{\beta}}} \right)^{\lambda} \right)^{\alpha}. \quad (12)$$

Sedangkan fungsi *log-likelihood* dari model regresi Binomial dengan fungsi glogit yang telah disempurnakan ini dapat dinyatakan dengan persamaan berikut:

$$\ln L_{gl} = \sum_{i=1}^n \left\{ (y_i) \alpha \log \left(1 - \left(1 - \frac{1}{1 + e^{-\mathbf{X}'_i \boldsymbol{\beta}}} \right)^{\lambda} \right) + (1 - y_i) \log \left[1 - \left(1 - \left(1 - \frac{1}{1 + e^{-\mathbf{X}'_i \boldsymbol{\beta}}} \right)^{\lambda} \right)^{\alpha} \right] \right\}. \quad (13)$$

Fungsi tautan logit yang digunakan dalam model regresi logistik merupakan bentuk spesifik dari fungsi tautan glogit dengan $\alpha = \lambda = 1$. Jika parameter α dan λ pada Persamaan (11) diganti dengan nilai 1 maka diperoleh persamaan model regresi logistik sebagai berikut:

$$\text{logit}(p_{li}) = \log \left(\frac{p_{li}}{1 - p_{li}} \right) = \mathbf{X}'_i \boldsymbol{\beta}; \quad 1 > p_{li} > 0, \quad (14)$$

yang mana p_{li} adalah probabilitas rumah tangga i berada di bawah garis kemiskinan dengan menggunakan model regresi logistik. Berdasarkan Persamaan (14) diperoleh persamaan inversi-logit dari model regresi logistik sebagai berikut:

$$p_{li} = F(p_{li}) = \frac{e^{\mathbf{X}'_i \boldsymbol{\beta}}}{1 + e^{\mathbf{X}'_i \boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{X}'_i \boldsymbol{\beta}}}. \quad (15)$$

Memperhatikan fungsi inversi-glogit pada Persamaan (12) dan fungsi inversi-logit pada Persamaan (15), maka persamaan untuk inversi-glogit juga dapat dinyatakan dengan persamaan berikut:

$$p_i = F(p_i) = \left(1 - \left(1 - \frac{1}{1 + e^{-\mathbf{X}'_i \boldsymbol{\beta}}} \right)^{\lambda} \right)^{\alpha} = (1 - (1 - p_{li})^{\lambda})^{\alpha}, \quad (16)$$

yang mana p_{li} menyatakan probabilitas rumah tangga i berada di bawah garis kemiskinan dengan menggunakan model regresi logistik.

5. Estimasi Parameter Model dengan Pendekatan Bayesian

Estimasi parameter model dalam penelitian ini dilakukan dengan metode Bayesian menggunakan algoritma *Hamiltonian Monte Carlo* (HMC) (Gelman dkk., 2014) dan program RStan. Misal, δ merupakan vektor parameter yang tidak diketahui dari suatu model dan y merupakan vektor variabel dependen pada data. Dengan pendekatan Bayesian, kesimpulan tentang parameter δ dinyatakan dalam bentuk probabilitas yang bergantung pada nilai y yang diamati. Probabilitas ini disebut sebagai distribusi *joint posterior* dari parameter δ dan dinotasikan dengan $f(\delta|y)$. Misalkan, informasi awal tentang parameter tersebut dinyatakan dengan distribusi *joint prior*, yang dilambangkan dengan $f(\delta)$. Fungsi *likelihood* suatu model berdasarkan pada data yang ada dinyatakan dengan $L_w(y|\delta)$ untuk model *Weighted Logistic Regression* (WLR), $L_{NB}(y|\delta)$ untuk model *Naive Bayes*, dan $L_{gl}(y|\delta)$ untuk model regresi Binomial dengan fungsi logit. Maka, distribusi *joint posterior* $f(\delta|y)$ dari masing-masing model secara proporsionalnya dapat dinyatakan sebagai berikut:

$$f_w(\delta|y) \propto L_w(y|\delta)f_w(\delta), \quad (17)$$

$$f_{NB}(\delta|y) \propto L_{NB}(y|\delta)f_{NB}(\delta), \quad (18)$$

$$f_{gl}(\delta|y) \propto L_{gl}(y|\delta)f_{gl}(\delta). \quad (19)$$

Fungsi *log-likelihood* untuk model WLR dinyatakan dalam Persaman (2), fungsi *log-likelihood* model *Naive Bayes* dinyatakan dalam Persamaan (7), dan fungsi *log-likelihood* untuk model regresi Binomial dengan fungsi logit dinyatakan dalam Persamaan (13). Selanjutnya, jika diasumsikan parameter masing-masing model saling bebas maka distribusi *joint prior* dari model WLR memenuhi kondisi berikut:

$$f_w(\delta) = \prod_{j=1}^7 f(\beta_j). \quad (20)$$

Distribusi *joint prior* dari model *Naive Bayes* memenuhi kondisi berikut:

$$f_{NB}(\delta) = \prod_{c=1}^2 \left(\prod_{j=1}^2 f(\mu_{jc})f(\sigma_{jc}) \prod_{j=3}^6 f(\gamma_{jc}) \right), \quad (21)$$

dan distribusi *joint prior* dari model regresi Binomial dengan fungsi logit memenuhi kondisi berikut:

$$f_{gl}(\delta) = f(\alpha)f(\lambda) \prod_{j=1}^7 f(\beta_j), \quad (22)$$

yang mana simbol $f(\alpha)f(\lambda), f(\beta_j), f(\mu_{jc}), f(\sigma_{jc}),$ dan $f(\gamma_{jc})$ masing-masing merupakan distribusi *prior* dari parameter $\alpha, \lambda, \beta_j, \mu_{jc}, \sigma_{jc}$ dan γ_{jc} pada ketiga model tersebut, serta indeks j menyatakan variabel independen atau fitur yang ke- $j; j = 0, 1, 2, \dots, 6$, dan c menyatakan kelas yang ke- $c; c = 1, 2$. Distribusi *prior* dari masing-masing parameter tersebut ditetapkan sebagai berikut:

1. Parameter α, λ , dan σ_{jc} diasumsikan berdistribusi *truncated* Student-t yaitu $\alpha \sim t_3(0, 1)I(0, \infty)$, $\lambda \sim t_3(0, 1)I(0, \infty)$, $\sigma_{jc} \sim t_3(0, 1)I(0, \infty)$.
2. Parameter β_j dan γ_{jc} diasumsikan berdistribusi Normal yaitu $\beta_j \sim N(0, 1)$ dan $\gamma_{jc} \sim N(0, 1)$.
3. Parameter μ_{jc} diasumsikan berdistribusi *Uniform* yaitu $\mu_{1c} \sim \text{Uniform}(0, 400)$ dan $\mu_{2c} \sim \text{Uniform}(18, 92)$.

Distribusi *posterior* berisi semua informasi terkini tentang parameter δ . Idealnya, seluruh distribusi *posterior* ditampilkan secara grafis. Namun, untuk tujuan praktis, perlu diberikan berbagai ringkasan numerik dari distribusi tersebut. Ringkasan numerik tersebut biasanya berupa ringkasan lokasi atau titik, seperti *mean*, median, dan modus. Selain itu, juga bisa diberikan ringkasan interval dengan *Credible Interval* (CI) atau *Highest Posterior Density Interval* (HPDI). Selain sebagai ringkasan numerik, selang interval juga bisa digunakan untuk melihat signifikansi perbedaan suatu parameter dengan angka nol (Gelman dan Tuerlinckx, 2000). Jika CI tidak mencakup angka nol,

dapat dikatakan bahwa secara statistik parameter tersebut berbeda nyata dengan angka nol. Namun, jika sebaliknya, maka parameter tersebut tidak berbeda nyata dengan angka nol.

HASIL DAN PEMBAHASAN

1. Perbandingan Keباikan Model

Kinerja ketiga model pada kasus kemiskinan di Kalimantan Timur dibandingkan berdasarkan kemampuan memprediksi status rumah tangga miskin dengan baik, melalui tingkat akurasi dan sensitivitas/*recall*. Model Regresi Binomial dengan fungsi glogit memiliki akurasi tertinggi, tetapi gagal mendeteksi rumah tangga miskin, sehingga sensitivitasnya nol. Sebaliknya, Model *Naïve Bayes* dan *Weighted Logistic Regression* memiliki akurasi yang lebih rendah tetapi mampu mendeteksi rumah tangga miskin dengan lebih baik. Tabel 2 menunjukkan kemampuan masing-masing model dalam memprediksi status rumah tangga tersebut, termasuk informasi tentang akurasi dan *recall* atau sensitivitas.

Model yang memiliki sensitivitas tertinggi adalah model *Naïve Bayes*. Model tersebut menunjukkan hasil prediksi yang memadai dengan 241 rumah tangga yang benar diklasifikasikan sebagai miskin dan 3.953 rumah tangga yang benar diklasifikasikan sebagai tidak miskin. Namun, terdapat kesalahan klasifikasi pada 102 rumah tangga miskin yang salah dikategorikan sebagai tidak miskin dan 1.724 rumah tangga tidak miskin yang salah dikategorikan sebagai miskin. Akurasi model ini sebesar 69,67%, yang berarti sekitar 69,67% prediksinya benar. Sensitivitasnya adalah 70,26%, menunjukkan bahwa model ini mampu mengidentifikasi 70,26% rumah tangga miskin dengan benar.

Tabel 2. Data aktual, kemampuan prediksi kategori miskin dan tidak miskin masing-masing model, akurasi dan *Recall*.

Model	Kemampuan Prediksi			Akurasi	<i>Recall</i> / Sensitivitas
	Kategori	Aktual Miskin	Aktual Tidak Miskin		
<i>Naïve Bayes</i>	Miskin	241	1.724	0,6967	0,7026
	Tidak Miskin	102	3.953		
<i>Weighted Logistic Regression</i>	Miskin	239	1.704	0,6997	0,6968
	Tidak Miskin	104	3.973		
Regresi Binomial dengan Fungsi glogit	Miskin	0	0	0,9430	0,0000
	Tidak Miskin	343	5.677		

Model *Weighted Logistic Regression* menghasilkan prediksi yang hampir serupa dengan Model *Naïve Bayes*. Model ini benar dalam mengklasifikasikan 239 rumah tangga sebagai miskin dan 3.973 rumah tangga sebagai tidak miskin. Kesalahan klasifikasi terjadi pada 104 rumah tangga miskin yang salah dikategorikan sebagai tidak miskin dan 1.704 rumah tangga tidak miskin yang salah dikategorikan sebagai miskin. Akurasi model ini adalah 69,97%, sedikit lebih tinggi dari Model *Naïve Bayes*, sedangkan sensitivitasnya adalah 69,67%, sedikit lebih rendah dari Model *Naïve Bayes*. Ini menunjukkan bahwa Model *Weighted Logistic Regression* sedikit lebih akurat secara keseluruhan, tetapi sedikit kurang sensitif dalam mendeteksi rumah tangga miskin dibandingkan dengan Model *Naïve Bayes*.

Model Regresi Binomial dengan fungsi glogit menunjukkan hasil yang sangat berbeda. Dalam *Confusion Matrix*, model ini benar dalam mengklasifikasikan 5.677 rumah tangga tidak miskin, tetapi 343 rumah tangga miskin telah salah diklasifikasikan sebagai rumah tangga tidak miskin. Akurasi model ini tercatat sebesar 94,30%, jauh lebih tinggi dibandingkan dua model sebelumnya. Namun, sensitivitas model ini adalah nol, yang berarti model ini gagal mengidentifikasi rumah tangga miskin. Meskipun model ini sangat akurat secara keseluruhan, ia cenderung bias terhadap kelas mayoritas (rumah tangga tidak miskin) dan mengabaikan kelas minoritas (rumah tangga miskin).

2. Hasil dan Pembahasan Model *Weighted Logistic Regression*

Hasil *sampling* yang terdapat pada Tabel 3 untuk model *Weighted Logistic Regression* (WLR) menunjukkan konvergensi yang sangat baik dan *sampling* yang efisien. Model ini dijalankan dengan tiga rantai MCMC, masing-masing terdiri dari 2.000 iterasi dengan 1.000 iterasi *warmup*. Nilai *R-hat* yang konsisten sebesar 1 untuk semua parameter mengindikasikan konvergensi yang optimal dari rantai *Markov Chain Monte Carlo* (MCMC). Jumlah sampel efektif (*n_eff*) yang tinggi, berkisar antara 1.786 hingga 2.867, menandakan bahwa *sampling* telah dilakukan dengan sangat efisien, memberikan estimasi parameter yang dapat diandalkan.

Tabel 3. *Posterior mean, interval kredibel (CI) 95%, jumlah sampel efektif (*n_eff*), dan nilai R-hat Model Weighted Logistic Regression.*

Parameter	Mean	CI 95%		<i>n_eff</i>	<i>R-hat</i>
		Batas Bawah	Batas Atas		
β_1	-0,74	-0,92	-0,56	1.786	1
β_2	3,78e-03	3,26e-03	4,30e-03	2.867	1
β_3	-0,02	-0,02	-0,02	2.179	1
β_4	1,54	1,46	1,62	2.485	1
β_5	0,48	0,40	0,56	2.248	1
β_6	0,43	0,35	0,51	2.498	1
β_7	0,79	0,70	0,89	2.276	1

Intercept model (β_1) memiliki *posterior mean* -0,74 dengan interval kredibel (CI) 95% antara -0,92 dan -0,56. Ini mengindikasikan bahwa ketika semua prediktor bernilai nol, terdapat kecenderungan dasar yang lebih rendah untuk sebuah rumah tangga menjadi miskin. Namun, interpretasi ini harus dipertimbangkan dalam konteks variabel prediktor lainnya, mengingat beberapa variabel adalah kategorikal dan tidak memiliki nilai nol yang bermakna.

Di antara variabel prediktor, jumlah anggota rumah tangga (β_4) muncul sebagai faktor yang paling berpengaruh terhadap kemiskinan. Dengan *posterior mean* 1,54 dan interval kredibel 95% antara 1,46 dan 1,62, hasil ini menunjukkan bahwa rumah tangga dengan 5 orang atau lebih memiliki probabilitas yang jauh lebih tinggi untuk menjadi miskin dibandingkan rumah tangga yang lebih kecil. Temuan ini menyoroti pentingnya ukuran keluarga dalam dinamika kemiskinan dan dapat memiliki implikasi signifikan untuk kebijakan kependudukan dan kesejahteraan sosial.

Status pekerjaan kepala rumah tangga (β_7) juga memiliki pengaruh substansial, dengan *posterior mean* 0,79 dan interval kredibel 95% antara 0,70 dan 0,89. Ini menunjukkan bahwa kepala rumah tangga yang tidak bekerja atau bekerja di sektor pertanian memiliki probabilitas yang jauh lebih tinggi untuk menjadi miskin dibandingkan mereka yang bekerja di sektor lainnya. Temuan ini menekankan pentingnya diversifikasi ekonomi dan pengembangan kesempatan kerja di luar sektor pertanian sebagai strategi pengentasan kemiskinan.

Lokasi tempat tinggal (β_5) dan tingkat pendidikan kepala rumah tangga (β_6) juga memiliki efek yang signifikan terhadap kemiskinan. Rumah tangga di perdesaan dan rumah tangga dengan kepala keluarga berpendidikan SLTP ke bawah memiliki probabilitas yang lebih tinggi untuk menjadi miskin. Ini menunjukkan adanya kesenjangan geografis dan pendidikan yang perlu ditangani dalam upaya pengentasan kemiskinan.

Menariknya, usia kepala rumah tangga (β_3) memiliki efek negatif yang sangat kecil terhadap kemiskinan, sementara *dependency ratio* (β_2) hampir tidak terdeteksi efeknya terhadap kemiskinan. Ini menunjukkan bahwa faktor-faktor struktural seperti ukuran keluarga, pendidikan, dan pekerjaan memiliki pengaruh yang jauh lebih besar terhadap status kemiskinan dibandingkan dengan karakteristik demografis seperti usia kepala rumah tangga atau rasio ketergantungan.

3. Hasil dan Pembahasan Model *Naïve Bayes*

Metodologi yang diimplementasikan dalam studi ini menerapkan inferensi Bayesian untuk mengestimasi parameter model *Naïve Bayes*, dengan menggunakan algoritma HMC melalui *platform* RStan. Proses *sampling* dilakukan dengan tiga rantai independen, masing-masing terdiri dari 2.000 iterasi, dengan 1.000 iterasi pertama ditetapkan sebagai periode *warm-up*. Hasil *sampling* yang

terdapat pada Tabel 4 memberikan estimasi *posterior* untuk berbagai parameter model, termasuk rata-rata (μ), standar deviasi (σ), probabilitas (p), dan transformasi logit dari probabilitas (γ) untuk setiap fitur dan kelas. Semua parameter menunjukkan nilai *R-hat* sebesar 1, yang menandakan konvergensi yang baik dari rantai MCMC.

Secara keseluruhan, model ini memberikan gambaran yang detail mengenai faktor-faktor yang berkorelasi dengan kemiskinan rumah tangga di Kalimantan Timur. *Dependency ratio* yang tinggi, usia kepala rumah tangga yang lebih muda, jumlah anggota keluarga yang besar, tinggal di perdesaan, tingkat pendidikan yang rendah, dan bekerja di sektor pertanian atau tidak bekerja semuanya berkaitan dengan probabilitas yang lebih tinggi untuk menjadi rumah tangga miskin. Model ini dapat menjadi alat yang berguna untuk mengidentifikasi rumah tangga yang berisiko tinggi mengalami kemiskinan dan untuk merancang intervensi kebijakan yang tepat sasaran. Namun, perlu diingat bahwa korelasi tidak selalu berarti kausalitas, dan terdapat faktor-faktor lain yang tidak termasuk dalam model ini yang mungkin juga mempengaruhi status kemiskinan rumah tangga.

Rata-rata *dependency ratio* (X_1) untuk rumah tangga miskin adalah 89,78 dengan interval kredibel 95% antara 82,26 hingga 97,30. Sementara itu, untuk rumah tangga tidak miskin, rata-rata *dependency ratio* lebih rendah, yaitu 60,93 dengan interval kredibel 95% antara 58,78 hingga 63,05. Ini berarti bahwa, secara rata-rata, setiap 100 orang produktif secara ekonomi dalam rumah tangga miskin harus menanggung hampir 97 orang yang tidak produktif, seperti anak-anak di bawah 15 tahun dan orang tua di atas 64 tahun. Sebaliknya, rumah tangga tidak miskin hanya menanggung sekitar 61 orang yang tidak produktif untuk setiap 100 orang produktif. Perbedaan yang signifikan ini menunjukkan bahwa rumah tangga miskin menghadapi beban ekonomi yang lebih besar. Pendapatan yang dihasilkan oleh anggota rumah tangganya yang produktif harus dibagi untuk mendukung lebih banyak anggota yang tidak produktif.

Rata-rata usia kepala rumah tangga (X_2) kategori miskin adalah 46,37 tahun dengan interval kredibel 95% antara 45,21 hingga 47,52 tahun. Sebaliknya, untuk rumah tangga tidak miskin, rata-rata usia kepala rumah tangga sedikit lebih tinggi, yaitu 48,18 tahun dengan interval kredibel 95% antara 47,88 hingga 48,48 tahun. Meskipun perbedaan usia ini relatif kecil, namun signifikan secara statistik karena interval kredibel yang tidak tumpang tindih. Hal ini menunjukkan bahwa kepala rumah tangga miskin cenderung memiliki usia yang lebih muda.

Jumlah anggota rumah tangga (X_3) menunjukkan perbedaan yang cukup jelas. Probabilitas rumah tangga miskin memiliki 5 anggota atau lebih adalah 0,61 dengan interval kredibel 95% antara 0,56 hingga 0,66. Sementara itu, probabilitas untuk rumah tangga tidak miskin adalah 0,24 dengan interval kredibel 95% antara 0,23 hingga 0,25. Ini menunjukkan bahwa rumah tangga miskin cenderung memiliki lebih banyak anggota.

Lokasi tempat tinggal (X_4) juga menunjukkan perbedaan yang signifikan. Probabilitas rumah tangga miskin tinggal di perdesaan adalah 0,57 dengan interval kredibel 95% antara 0,52 hingga 0,62. Sedangkan probabilitas rumah tangga tidak miskin tinggal di perdesaan adalah 0,37 dengan interval kredibel 95% antara 0,36 hingga 0,39. Hal ini mengindikasikan bahwa kemiskinan lebih sering terjadi di daerah perdesaan.

Tingkat pendidikan kepala rumah tangga (X_5) menunjukkan bahwa probabilitas kepala rumah tangga miskin memiliki pendidikan maksimal SLTP adalah 0,56 dengan interval kredibel 95% antara 0,50 hingga 0,61. Sebaliknya, untuk rumah tangga tidak miskin, probabilitasnya lebih rendah, yaitu 0,43 dengan interval kredibel 95% antara 0,41 hingga 0,44. Ini menunjukkan bahwa tingkat pendidikan yang lebih rendah berkorelasi dengan kemiskinan.

Terakhir, status pekerjaan kepala rumah tangga (X_6) menunjukkan bahwa probabilitas kepala rumah tangga miskin tidak bekerja atau bekerja di sektor pertanian adalah 0,57 dengan interval kredibel 95% antara 0,52 hingga 0,62. Sebaliknya, untuk rumah tangga tidak miskin, probabilitasnya lebih rendah, yaitu 0,37 dengan interval kredibel 95% antara 0,36 hingga 0,39. Ini mengindikasikan bahwa bekerja di sektor pertanian atau tidak bekerja berkorelasi dengan kemiskinan.

Parameter γ (gamma) yang ditampilkan di bagian bawah Tabel 4 merepresentasikan transformasi logit dari probabilitas untuk variabel biner. Nilai positif menunjukkan kecenderungan lebih tinggi terhadap kelas 1 (miskin), sementara nilai negatif menunjukkan kecenderungan terhadap kelas 2 (tidak miskin).

Tabel 4. *Posterior mean, interval kredibel (CI) 95%, jumlah sampel efektif (n_eff), dan nilai R-hat Model Naive Bayes.*

Parameter	Mean	CI 95%		n_eff	R-hat
		Batas Bawah	Batas Atas		
μ_{11}	89,78	82,26	97,30	4.084	1
μ_{12}	60,93	58,78	63,05	4.509	1
μ_{21}	46,37	45,21	47,52	3.359	1
μ_{22}	48,18	47,88	48,48	5.518	1
σ_{11}	71,22	66,05	76,75	3.928	1
σ_{12}	80,09	78,63	81,55	4.173	1
σ_{21}	11,23	10,42	12,13	4.260	1
σ_{22}	12,08	11,85	12,31	4.739	1
p_{11}	0,61	0,56	0,66	4.213	1
p_{12}	0,24	0,23	0,25	4.087	1
p_{21}	0,58	0,53	0,63	3.531	1
p_{22}	0,39	0,38	0,40	3.892	1
p_{31}	0,56	0,50	0,61	4.184	1
p_{32}	0,43	0,41	0,44	3.731	1
p_{41}	0,57	0,52	0,62	3.425	1
p_{42}	0,37	0,36	0,39	3.682	1
γ_{11}	0,44	0,23	0,66	4.214	1
γ_{12}	-1,14	-1,20	-1,08	4.085	1
γ_{21}	0,32	0,12	0,53	3.521	1
γ_{22}	-0,45	-0,50	-0,39	3.898	1
γ_{31}	0,23	0,01	0,44	4.182	1
γ_{32}	-0,3	-0,35	-0,25	3.725	1
γ_{41}	0,27	0,07	0,49	3.426	1
γ_{42}	-0,52	-0,57	-0,46	3.679	1

4. Hasil dan Pembahasan Model Regresi Binomial Dengan Fungsi Tautan Glogit

Estimasi parameter model ini juga dilakukan menggunakan pendekatan Bayesian dengan paket RStan. *Sampling* dilakukan dengan 3 *chains*, masing-masing dengan 2.500 iterasi, 1.000 iterasi *warmup*, dan 1.500 *post-warmup* per *chain*, menghasilkan total 4.500 *post-warmup*. Pada penelitian ini, pendekatan Bayesian dengan RStan memberikan estimasi parameter yang stabil dengan nilai *n_eff* yang tinggi dan *R-hat* yang mendekati 1 sebagaimana yang terdapat pada Tabel 5, menunjukkan konvergensi yang baik dari *sampling*.

Model regresi binomial dengan fungsi tautan glogit yang digunakan dalam penelitian ini menunjukkan hasil yang signifikan untuk beberapa variabel kovariat. Variabel usia kepala rumah tangga, *dependency ratio*, jumlah anggota rumah tangga, lokasi tempat tinggal, tingkat pendidikan kepala rumah tangga, dan status pekerjaan kepala rumah tangga memiliki pengaruh yang signifikan terhadap probabilitas kemiskinan rumah tangga.

Intercept (Parameter β_1) dari model ini memiliki nilai *posterior mean* -0,24 dengan interval kepercayaan 95% antara -1,10 dan 0,63. Hal ini menunjukkan bahwa *intercept* tidak signifikan secara statistik karena interval kepercayaan tersebut mencakup nilai nol. Dalam konteks model, *intercept* mewakili probabilitas dasar kemiskinan rumah tangga ketika semua variabel kovariat lainnya adalah nol. Mengingat interval kepercayaan mencakup nol, maka tidak dapat disimpulkan adanya efek yang signifikan dari *intercept* terhadap probabilitas kemiskinan.

Tabel 5. *Posterior Mean*, interval kredibel (CI) 95%, jumlah sampel efektif (n_{eff}), dan nilai *R-hat* Model Regresi Binomial dengan fungsi tautan glogit.

Parameter	<i>Mean</i>	CI 95%		n_{eff}	<i>R-hat</i>
		Batas Bawah	Batas Atas		
β_1	-0,24	-1,10	0,63	2.093	1
β_2	2,11e-03	1,14e-03	3,34e-03	1.447	1
β_3	-0,01	-0,02	-0,01	1.502	1
β_4	0,90	0,60	1,28	1.002	1
β_5	0,27	0,11	0,47	1.646	1
β_6	0,24	0,09	0,41	1.610	1
β_7	0,45	0,27	0,70	1.321	1
α	3,32	1,92	5,33	1.127	1
λ	0,76	0,34	1,55	894	1

Parameter *dependency ratio* (β_2) memiliki nilai *posterior mean* sebesar 0,00211 dengan interval kepercayaan 95% antara 0,00114 dan 0,00334. Meskipun nilai *posterior mean* sangat kecil, interval kepercayaan yang tidak mencakup nol menunjukkan bahwa *dependency ratio* memiliki pengaruh signifikan terhadap status kemiskinan rumah tangga. Hal ini berarti bahwa setiap kenaikan dalam *dependency ratio* cenderung meningkatkan probabilitas rumah tangga tersebut jatuh ke dalam kemiskinan. Dengan kata lain, rumah tangga dengan lebih banyak anggota non-produktif (seperti anak-anak dan orang tua) dibandingkan dengan anggota produktif menghadapi risiko kemiskinan yang lebih tinggi.

Usia kepala rumah tangga (Parameter β_3) memiliki nilai *posterior mean* -0,01 dengan interval kepercayaan 95% antara -0,02 dan -0,01. Meskipun nilai *posterior mean* mendekati nol, interval kepercayaan yang sangat sempit dan tidak mencakup nol menunjukkan bahwa usia kepala rumah tangga memiliki pengaruh yang signifikan terhadap status kemiskinan, meskipun sangat kecil. Pengaruh negatif ini menunjukkan bahwa setiap peningkatan usia kepala rumah tangga akan mengurangi probabilitas kemiskinan.

Jumlah anggota rumah tangga (Parameter β_4) memiliki nilai *posterior mean* 0,90 dengan interval kepercayaan 95% antara 0,60 dan 1,28. Koefisien positif yang cukup besar ini menunjukkan bahwa rumah tangga dengan 5 orang atau lebih memiliki probabilitas kemiskinan yang lebih tinggi dibandingkan dengan rumah tangga yang memiliki kurang dari 5 orang. Ini konsisten dengan anggapan bahwa rumah tangga yang lebih besar akan menghadapi tekanan ekonomi yang lebih besar.

Lokasi tempat tinggal (Parameter β_5) memiliki nilai *posterior mean* 0,27 dengan interval kepercayaan 95% antara 0,11 dan 0,47. Koefisien positif ini menunjukkan bahwa rumah tangga yang tinggal di perdesaan memiliki probabilitas kemiskinan yang lebih tinggi dibandingkan dengan yang tinggal di perkotaan. Interval kepercayaan yang tidak mencakup nol mengonfirmasi bahwa pengaruh ini signifikan secara statistik.

Tingkat pendidikan kepala rumah (Parameter β_6) tangga memiliki nilai *posterior mean* 0,24 dengan interval kepercayaan 95% antara 0,09 dan 0,41. Koefisien positif ini menunjukkan bahwa kepala rumah tangga dengan pendidikan SLTP ke bawah memiliki probabilitas kemiskinan yang lebih tinggi dibandingkan dengan yang berpendidikan SLTA ke atas. Interval kepercayaan yang tidak mencakup nol menunjukkan bahwa efek ini signifikan.

Status pekerjaan kepala rumah tangga (Parameter β_7) memiliki nilai *posterior mean* 0,45 dengan interval kepercayaan 95% antara 0,27 dan 0,70. Koefisien positif ini menunjukkan bahwa kepala rumah tangga yang tidak bekerja atau bekerja di sektor pertanian memiliki probabilitas kemiskinan yang lebih tinggi dibandingkan dengan yang bekerja di sektor lainnya. Interval kepercayaan yang tidak mencakup nol menunjukkan bahwa efek ini juga signifikan.

Parameter bentuk α memiliki nilai *posterior mean* 3,32 dengan interval kepercayaan 95% antara 1,92 dan 5,33, sedangkan parameter bentuk λ memiliki nilai *posterior mean* 0,76 dengan interval kepercayaan 95% antara 0,34 dan 1,55. Kedua parameter ini signifikan, seperti yang ditunjukkan oleh interval kepercayaan yang tidak mencakup nol. Nilai α yang lebih besar dari 1 dan nilai λ yang

lebih rendah dari 1 menunjukkan bahwa distribusi data memiliki ekor yang lebih tebal dibandingkan distribusi logistik standar dan miring ke kiri.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis faktor-faktor penentu kemiskinan di Kalimantan Timur dengan menggunakan tiga model statistik yang berbeda, yaitu *Weighted Logistic Regression* (WLR), *Naive Bayes*, dan Regresi Binomial dengan fungsi tautan logit tergeneralisasi (glogit). Data yang digunakan dalam penelitian ini berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan pada Maret 2023. Model-model tersebut dievaluasi menggunakan pendekatan Bayesian dengan algoritma *Hamiltonian Monte Carlo* (HMC) melalui program RStan.

Dalam perbandingan kinerja model, model WLR dan *Naive Bayes* menunjukkan kemampuan yang lebih baik dalam mendeteksi rumah tangga miskin dibandingkan model Regresi Binomial dengan fungsi tautan glogit, meskipun model Regresi Binomial dengan fungsi tautan glogit menunjukkan akurasi keseluruhan yang lebih tinggi. Model Regresi Binomial dengan fungsi tautan glogit cenderung bias terhadap kelas mayoritas dan gagal mengidentifikasi rumah tangga miskin, yang tercermin dalam sensitivitas model yang sangat rendah.

Setiap model memiliki kemampuan yang berbeda dalam mengidentifikasi rumah tangga miskin. Namun, analisis menunjukkan bahwa beberapa faktor yang berpengaruh signifikan terhadap status kemiskinan tetap konsisten di ketiga model tersebut. Faktor-faktor yang signifikan tersebut meliputi jumlah anggota rumah tangga, lokasi tempat tinggal, tingkat pendidikan kepala rumah tangga, status pekerjaan kepala rumah tangga, usia kepala rumah tangga, dan *dependency ratio*. Secara spesifik, rumah tangga dengan lima anggota atau lebih, yang tinggal di perdesaan, dengan kepala rumah tangga berpendidikan SLTP ke bawah, yang kepala rumah tangganya tidak bekerja atau bekerja di sektor pertanian, yang kepala rumah tangganya lebih muda, serta rumah tangga dengan lebih banyak anggota non-produktif (seperti anak-anak dan orang tua) dibandingkan dengan anggota produktif, memiliki probabilitas kemiskinan yang lebih tinggi.

Temuan dalam penelitian ini menunjukkan bahwa model WLR dan *Naive Bayes* lebih efektif dalam menangani data yang tidak seimbang dalam konteks identifikasi rumah tangga miskin di Kalimantan Timur. Selain itu, hasil penelitian memberikan wawasan penting bahwa ketiga model mampu mengidentifikasi faktor-faktor yang mempengaruhi kemiskinan, meskipun mereka memiliki kemampuan prediksi yang berbeda dalam hal mendeteksi rumah tangga miskin.

DAFTAR PUSTAKA

- Aderounmu, B., Azuh, D., Onanuga, O., Oluwatomisin, O., Ebenezer, B. and Azuh, A. (2021), "Poverty drivers and Nigeria's development: Implications for policy intervention", *Cogent Arts & Humanities*, Vol. 8, No. 1.
- Biyase, M. and Zwane, T. (2018), "An Empirical Analysis of The Determinants of Poverty and Household Welfare in South Africa", *The Journal of Developing Areas*, Vol. 52, No. 1, pp. 115-130.
- Collins, M. (2012), *The naive bayes model, maximum-likelihood estimation, and the em algorithm*, [Online] diakses dari: <http://web2.cs.columbia.edu/~mcollins/em.pdf>.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2014), *Bayesian Data Analysis*, 3rd ed., Chapman and Hall/CRC Press, Boca Raton, Fla.
- Gelman, A. and Tuerlinckx, F. (2000), "Type S error rates for classical and Bayesian single and multiple comparison procedures", *Computational Statistics*, Vol. 15, pp. 373-390.
- Ghosh, I. and Alzaatreh, A. (2018), "A new class of generalized logistic distribution", *Communications in Statistics-Theory and Methods*, Vol. 47, No. 9, pp. 2043-2055.
- He, H. and Garcia, E.A. (2009), "Learning from imbalanced data", *IEEE Transactions on knowledge and data engineering*, Vol. 21, No. 9, pp. 1263-1284.
- He, H. and Ma, Y. (2013), *Imbalanced Learning: Foundations, Algorithms, And Applications*, John Wiley & Sons.
- Iqbal, A., Siddiqui, A. and Zafar, M. (2020), "A Geographically Disaggregated Analysis of Multidimensional Poverty in Punjab", *International Journal of Social Economics*, Vol. 47, No. 3, pp. 365-383.
- King, G. and Zeng, L. (2001), "Logistic Regression in Rare Events Data", *Political Analysis*, Vol. 9, No. 2, pp. 137-163.

- Kuhn, M. and Johnson, K. (2013), *Applied predictive modeling*, Springer, New York.
- Murty, M.N. and Devi, V.S. (2011), *Pattern recognition: An Algorithmic Approach*, Springer Science & Business Media.
- Prasetyo, R.B., Kuswanto, H., Iriawan, N. and Ulama, B.S.S. (2020), "Binomial regression models with a flexible generalized logit link function", *Symmetry*, Vol. 12, No. 2, p. 221.
- Sulasih, D.E.A., Purnami, S.W. and Rahayu, S.P. (2015), "The theoretical study of rare event weighted logistic regression for classification of imbalanced data", *International Conference on Science, Technology and Humanity*.
- Ting, K.M. (2002), "An instance-weighting method to induce cost-sensitive trees", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 14, No. 3, pp. 659-665.