p-ISSN 2085-7829
e-ISSN 2798-3455

Analisis Regresi Logistik Multinomial Bayes untuk Pemodelan Minat Peserta Didik MAN 2 Samarinda Tahun Ajaran 2018/2019

Bayesian Multinomial Logistic Regression Analysis for Modeling The Interest of MAN 2 Samarinda Students in the 2018/2019 Academic Year

Era Tri Cahyani[1], Rito Goejantoro[2], Meiliyani Siringoringo[3]
[1,2] Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman
[3]Laboratorium Statistika Ekonomi dan Bisnis FMIPA Universitas Mulawarman
E-mail: [1]eradochi74@gmail.com

ABSTRACT

Currently, Senior High School and Madrasah Aliyah have implemented student specialization. The specialization includes Natural Science, Social Science and Language. There are several criteria for determining interest in Senior High School and Madrasah Aliyah which include academic scores, student interests and IQ. The multinomial logistic regression model is used to examine these factors because the dependent variable has more than 2 categories. Bayes method is used to estimate the parameters of the multinomial logistic regression. The Bayesian method is a parameter estimation technique that combines the likelihood and prior distribution function. The estimation with the Bayesian method was solved using Markov Chain Monte Carlo simulation (MCMC) with the Gibbs Sampler algorithm. The data used were new students at MAN 2 Samarinda on 2018/2019 with the results of interest namely Natural Science, Social Science and Language. Independent variables were used, namely the score of the Junior High School in subjects Natural Science, Social Science, Language and the rate of National Test. The results of modeling and analysis showed that the factors that significantly influenced were the score of the junior high school in the subject of Natural Science and the rate of National Test. The classification accuracy of the model was 63,10%.

Keywords: *Bayesian Method, Multinomial Logistic Regression, Specialization of Students*

Pendahuluan

Regresi logistik merupakan pendekatan model matematis yang dapat digunakan untuk menyatakan hubungan dari beberapa variabel bebas terhadap variabel terikatnya yang bersifat kategori (Kleinbaum, 1994). Regresi logistik terbagi menjadi dua yaitu regresi logistik biner dan regresi logistik multinomial. Regresi logistik biner adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara sekumpulan variabel bebas dengan variabel terikat, di mana variabel terikat bersifat biner atau dikotomus. Variabel kategorik yang tidak memiliki urutan disebut sebagai variabel nominal, sedangkan yang memiliki urutan disebut variabel ordinal. Kedua jenis variabel ini, baik nominal maupun ordinal sering disebut juga sebagai variabel multinomial (Agresti, 2002).

Pada penelitian sebelumnya oleh Arieska (2012) yang membandingkan metode *maximum likelihood estimation* (MLE) dengan metode Bayes pada regresi logistik multinomial dan diperoleh hasil bahwa misklasifikasi pada regresi logistik multinomial klasik lebih besar dibandingkan pada regresi logistik multinomial dengan metode Bayes. Misklasifikasi pada metode klasik sebesar 46,1% sedangkan pada metode Bayes sebesar 39,5%. Sehingga regresi logistik multinomial Bayes lebih baik dalam pemodelan jurusan di SMAN 1 Grati Pasuruan jika dibandingkan dengan regresi logistik multinomial klasik. Variabel bebas yang digunakan yaitu tuntas IPA, tuntas IPS, tuntas Bahasa, dan IQ di mana variabel terikatnya yaitu minat dengan tiga kategori yaitu IPA, IPS, dan Bahasa. Diperoleh hasil bahwa variabel yang berpengaruh secara signifikan yaitu tuntas Bahasa dan IQ.

Selain misklasifikasi yang lebih kecil dibandingkan dengan metode klasik, metode Bayes juga memberikan estimasi parameter dengan sifat-sifat statistik yang baik, deskripsi parsimonius data terobservasi, prediksi data hilang dan peramalan data yang akan datang, serta kerangka komputasi untuk estimasi model, pemilihan, dan validasi. (Subanar, 2019)

Penggunaan metode regresi logistik multinomial diterapakan pada minat peserta didik MAN 2 Samarinda. Variabel bebas yang digunakan adalah nilai IPA SMP semester 5, nilai IPS SMP semester 5, nilai Bahasa SMP semester 5 dan rata-rata nilai UN SMP. Variabel terikat yang digunakan adalah minat peserta didik dengan 3 kategori yaitu IPA, IPS dan Bahasa. *Prior* yang digunakan adalah *pseudo prior* dan mengikuti distribusi normal dengan nilai $\sigma^2 = 1$. Penetapan jurusan SMA ini akan dianalisis menggunakan regresi logistik multinomial dengan metode Bayes untuk mengetahui model yang terbentuk dan variabel apa saja yang berpengaruh secara signifikan.

Regresi Logistik Multinomial

Regresi logistik terbagi menjadi dua yaitu regresi logistik biner dan regresi logistik multinomial. Regresi logistik biner adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara sekumpulan variabel bebas dengan variabel terikat, di mana variabel terikat bersifat biner atau dikotomus. Variabel kategorik yang tidak memiliki urutan disebut sebagai variabel nominal, sedangkan yang memiliki urutan disebut variabel ordinal. Kedua jenis variabel ini, baik nominal maupun ordinal sering disebut juga sebagai variabel multinomial. Regresi logistik multinomial juga dapat diterapkan untuk meneliti sebuah variabel ordinal namun memanfaatkan sifat ordinal data dapat meningkatkan kesederhanaan dan kekuatan model. (Agresti, 2002)

Kita asumsikan variabel terikat *Y* mempunyai tiga kategori dan diberikan kode 0, 1, atau 2. Pada model regresi logistik biner, parameter terikat mempunyai fungsi logit *Y*=1 banding *Y*=0. Dalam regresi logistik multinomial dibutuhkan dua fungsi logit, di mana salah satu kategori dijadikan sebagai pembanding. Kita gunakan *Y*=0 sebagai pembanding *Y*=1 dan *Y*=2. (Hosmer & Lemeshow, 2000)

Dua fungsi logit ditunjukkan pada persamaan berikut. (Hosmer & Lemeshow, 2000)

$$\begin{aligned} g_1(\mathbf{x}) &= ln\left[\frac{P(Y=1|\mathbf{x})}{P(Y=0|\mathbf{x})}\right] \\ &= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p \\ &= \mathbf{x}'\boldsymbol{\beta}_1 \end{aligned} \quad (1)$$

dan

$$\begin{aligned} g_2(\mathbf{x}) &= ln\left[\frac{P(Y=2|\mathbf{x})}{P(Y=0|\mathbf{x})}\right] \\ &= \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots + \beta_{2p}x_p \\ &= \mathbf{x}'\boldsymbol{\beta}_2 \end{aligned} \quad (2)$$

Maka peluang dari masing-masing kategori diberikan oleh vektor kovariat

$$P(Y=0|\mathbf{x}) = \frac{1}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}} \quad (3)$$

$$P(Y=1|\mathbf{x}) = \frac{e^{g_1(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}} \quad (4)$$

dan

$$P(Y=2|\mathbf{x}) = \frac{e^{g_2(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}} \quad (5)$$

Teorema Bayes

Diberikan $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ adalah vektor dari *n* observasi di mana distribusi peluang $p(\mathbf{y}|\boldsymbol{\theta})$ bergantung dari nilai parameter k, di mana $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_k)$. Anggap bahwa $\boldsymbol{\theta}$ memiliki distribusi peluang $p(\boldsymbol{\theta})$. Maka,

$$p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\mathbf{y},\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})p(\mathbf{y}) \quad (6)$$

Diberikan data observasi $\mathbf{y}$, peluang bersyarat dari $\boldsymbol{\theta}$ adalah:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \quad (7)$$

Dapat dituliskan juga dengan

$$p(\mathbf{y}) = E[p(\mathbf{y}|\boldsymbol{\theta})] = c^{-1} = \begin{cases} \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}, \boldsymbol{\theta} \text{ kontinu} \\ \sum p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \boldsymbol{\theta} \text{ diskrit} \end{cases} \quad (8)$$

Di mana jumlah atau integral diambil atas kisaran yang dapat diterima dari $\boldsymbol{\theta}$, dan $E[f(\boldsymbol{\theta})]$ ekspektasi matematis dari $f(\boldsymbol{\theta})$ sehubungan dengan distribusi $p(\boldsymbol{\theta})$. Kita dapat menuliskan persamaan 7 sebagai berikut.

$$p(\boldsymbol{\theta}|\mathbf{y}) = cp(y|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (9)$$

Persamaan 7 ataupun persamaan 9 biasa disebut dengan Teorema Bayes. Dalam situasi ini $p(\boldsymbol{\theta})$ disebut distribusi *prior* $\boldsymbol{\theta}$. Sejalan dengan itu $p(\boldsymbol{\theta}|\mathbf{y})$ disebut distribusi *posterior* $\boldsymbol{\theta}$. Kuantitas c hanyalah konstanta normalisasi yang diperlukan untuk memastikan bahwa distribusi *posterior* $p(\boldsymbol{\theta}|\mathbf{y})$ terintegrasi atau dijumlahkan menjadi satu. Kita lebih sering menyebut distribusi *prior* dan distribusi *posterior* secara sederhana yaitu *prior* dan *posterior* secara spontan.

(Box & Tiao, 1973)

Distribusi *Prior*

Permasalahan pokok agar *prior* dapat interpretatif adalah bagaimana memilih distribusi *prior* untuk suatu parameter yang tidak diketahui namun sesuai dengan permasalahan yang ada. Tingkat keberhasilan metode Bayes sangat bergantung pada informasi awal yang diberikan atau *prior* yang dipilih. (Ainul, 2018)

Menurut Box dan Tiao (1973) distribusi *prior* dapat dikelompokkan menjadi dua kelompok berdasarkan bentuk fungsi *likelihood*-nya:

a. Berkaitan dengan bentuk distribusi hasil identifikasi pola datanya
   1) *Prior* konjugat, mengacu pada acuan analisis model terutama dalam pembentukan fungsi *likelihood*-nya sehingga dalam penentuan *prior* konjugat selalu dipikirkan mengenai penentuan pola distribusi *prior* yang mempunyai bentuk konjugat dengan fungsi densitas peluang pembangun *likelihood*-nya.
   2) *Prior* non-konjugat, pemberian *prior* pada model tidak mempertimbangkan pola pembentuk fungsi *likelihood*-nya.

p-ISSN 2085-7829
e-ISSN 2798-3455

b. Berkaitan dengan penentuan masing-masing parameter pada pola distribusi *prior* tersebut
   1) *Prior* informatif, mengacu pada pemberian parameter dari distribusi *prior* yang telah dipilih baik distribusi *prior* konjugat atau tidak. Pemberian nilai parameter pada distribusi *prior* ini akan sangat mempengaruhi bentuk distribusi *posterior* yang akan didapatkan pada informasi data yang diperoleh.
   2) *Prior* non-informatif, pemilihan distribusi *prior*-nya tidak didasarkan pada data yang ada atau distribusi *prior* yang tidak mengandung informasi tentang parameter θ.
c. Distribusi *prior* yang dibedakan atas ada atau tidaknya bentuk tetap untuk setiap variabel acak t yaitu *prior* proper dan improper. *Prior* ini timbul bila (θ) bukan distribusi peluang yaitu $(\theta) \geq 0$ atau $\int g(\theta)d\theta \neq 1$ (Lancaster, 2003).
d. *Pseudo prior* adalah *prior* terkait dengan pemberian nilai yang disetarakan dengan hasil elaborasi dari *frequentist* (Carlin & Chip, 1995).

Markov *Chain* Monte Carlo (MCMC)

Menurut Ntzoufras (2009) ide dasar dari MCMC adalah membangkitkan data parameter sesuai proses. Markov *Chain* dengan menggunakan simulasi Monte Carlo secara iteratif hingga diperoleh distribusi *posterior* yang stasioner (*steady state*). Hal ini yang membedakan MCMC dengan teknik simulasi langsung (*direct simulation*) yang lebih menitikberatkan efektifitas perhitungan integrasi tertentu dan tidak dapat digunakan untuk membangkitkan sampel dari berbagai bentuk distribusi *posterior* yang ada. Sementara MCMC lebih bersifat umum dan fleksibel.

Markov *Chain* adalah suatu proses stokastik dari $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)}\}$ sedemikian sehingga

$$f(\theta^{(t+1)}|\theta^{(t)}, \dots, \theta^{(1)} = f(\theta^{(t+1)}|\theta^{(t)}) \quad (10)$$

dengan distribusi θ pada saat $t+1$ diketahui kondisi semua nilai θ sebelumnya pada saat $t, t-1, \dots, 1$ adalah hanya akan dipengaruhi oleh nilai pada saat t saja. Pada saat $t \to \infty$, distribusi dari $\theta^{(t)}$ akan konvergen menuju distribusi tertentu yang independen terhadap nilai awal dari rantai tersebut. Dengan demikian untuk mendapatkan sampel dari distribusi *posterior*, maka diperlukan struktur Markov *Chain* yang memiliki sifat $f(\theta^{(t+1)}|\theta^{(t)})$ dan mudah untuk dibangkitkan. MCMC dilakukan dengan cara membangkitkan Markov *Chain* yang konvergen terhadap distribusi target yaitu distribusi *posterior* dari parameter yang ditaksir. Kondisi ini dinamakan kondisi stasioner atau kondisi *equilibrium*. Selanjutnya, sampel parameter dalam Markov *Chain* diambil setelah kondisi *equilibrium* tercapai. Dengan demikian sampel yang diperoleh merupakan sampel dari distribusi target yaitu distribusi *posterior* dari parameter tersebut. Untuk melihat konvergensi dari hasil estimasi parameter *posterior* bisa dilihat dari: (Ntzoufraz, 2009)

a. *Trace Plot*
   Salah satu cara estimasi *burn-in period* adalah memeriksa *trace plot* nilai simulasi dari komponen atau beberapa fungsi lainnya dari x terhadap jumlah iterasi. *Trend* naik turun nilai parameter pada *trace plot* menunjukkan bahwa *burn-in period* belum tercapai. Jika semua nilai-nilai berada dalam sebuah daerah tanpa kepriodikan yang kuat cenderung dapat dikatakan konvergen.
b. Otokorelasi
   Nilai simulasi x pada iterasi ke-$(t+1)$ bergantung pada nilai simulasi pada iterasi ke-t. Jika pada grafik otokorelasi pada lag pertama mendekati satu dan selanjutnya nilai-nilainya terus berkurang menuju 0 dapat dikatakan bahwa iterasi sudah dapat diberikan. Seperti yang telah disebutkan, sampel akhir yang dihasilkan MCMC tidak independen. Oleh karena itu, kita perlu memantau otokorelasi dari nilai yang dihasilkan dan memilih sampling dengan lag pertama mendekati satu dan lag-lag selanjutnya mendekati nol. Memperbesar nilai *thin* dilakukan untuk menghemat ruang penyimpanan atau kecepatan komputasi dalam masalah dimensi tinggi. Nilai *thin* dapat diubah oleh peneliti setelah memantau otokorelasi dan kesalahan *Markov Chain* yang mengacu pada perhitungan ringkasan dan plot dan bukan pada kumpulan nilai yang diberikan (Ntzoufras, 2009; Link & Eaton, 2012).
c. *Ergodic Mean Plot*
   Ergodic mean adalah istilah yang menunjukkan nilai *mean* sampai *current iteration*. Jika setelah beberapa kali iterasi *ergodic mean* stabil, maka ini merupakan sebuah indikasi bahwa konvergensi telah tercapai.

(Ntzoufraz, 2009)

Gibbs Sampler

Gibbs sampler merupakan generator yang sangat efisien, sehingga sering digunakan sebagai generator variabel random pada analisis data yang menggunakan metode MCMC. Casella dan George (1992) mendefinisikan *gibbs sampler* sebagai suatu teknik simulasi untuk membangkitkan variabel random dari suatu

p-ISSN 2085-7829
e-ISSN 2798-3455

distribusi tertentu secara tidak langsung, tanpa harus menghitung fungsi densitas dari suatu distribusi data. *Gibbs sampler* dapat diterapkan apabila distribusi peluang bersama (*joint probability distribution*) tidak diketahui secara eksplisit, tetapi distribusi bersyarat dari tiap-tiap variabel diketahui. Menurut Ntzoufraz (2009) langkah-langkah dari algoritma *gibbs sampler* adalah sebagi berikut.

Langkah 1: Menentukan nilai awal $\boldsymbol{\beta}^{(0)}$

Langkah 2: Untuk $t = 1,2,\dots,T$, melakukan pengulangan sesuai langkah di bawah ini.

a. Menentukan $\boldsymbol{\beta}=\boldsymbol{\beta}^{(t-1)}$
b. Untuk $p = 1,2,\dots,P$, melakukan *update* β_p dari $\beta_p \sim f(\beta_p|\boldsymbol{\beta}_{\backslash p},\mathbf{y})$
c. Mengatur $\boldsymbol{\beta}^t = \boldsymbol{\beta}$ dan menyimpannya sebagai himpunan nilai yang dihasilkan pada iterasi $(t+1)$ pada algoritma

Oleh karena itu dengan keadaan rantai $\boldsymbol{\beta}^t$ tertentu, menghasilkan nilai parameter baru dengan langkah sebagai berikut.

$\beta_1^{(t)}$ dari $f(\beta_1|\beta_2^{(t-1)},\beta_3^{(t-1)},\dots,\beta_p^{(t-1)},\mathbf{y})$
$\beta_2^{(t)}$ dari $f(\beta_2|\beta_1^{(t)},\beta_3^{(t-1)},\dots,\beta_p^{(t-1)},\mathbf{y})$
$\beta_3^{(t)}$ dari $f(\beta_3|\beta_1^{(t)},\beta_2^{(t)},\beta_4^{(t-1)},\dots,\beta_p^{(t-1)},\mathbf{y})$
$\vdots$
$\beta_p^{(t)}$ dari $f(\beta_p|\beta_1^{(t)},\beta_2^{(t)},\dots,\beta_{p-1}^{(t)},\beta_{p+1}^{(t-1)},\dots,\beta_m^{(t-1)},\mathbf{y})$
$\vdots$
$\beta_m^{(t)}$ dari $f(\beta_m|\beta_1^{(t)},\beta_2^{(t)},\dots,\beta_{m-1}^{(t)},\mathbf{y})$

Menghasilkan nilai dari $f(\beta_p|\boldsymbol{\beta}_{\backslash p},\mathbf{y})=f(\beta_p|\beta_1^{(t)},\dots,\beta_{p-1}^{(t)},\beta_{p+1}^{(t-1)},\dots,\beta_m^{(t-1)},\mathbf{y})$ relatif mudah karena berbentuk distribusi univariat dan dapat ditulis sebagai $f(\beta_p|\boldsymbol{\beta}_{\backslash p},\mathbf{y}) \propto f(\boldsymbol{\beta}|\mathbf{y})$, di mana semua variabel kecuali β_p bernilai konstan.

Interval Kredibel dan Ketepatan Klasifikasi

Pengujian hipotesis terhadap parameter regresi dilakukan dengan pendekatan interval kredibel 95% dari masing-masing parameter. Interval kredibel 95% dihitung dengan batas bawah yaitu kuantil 2,5% dan batas atas adalah kuantil 97,5%. Parameter dinyatakan signifikan jika interval kredibel 95% dari parameter tidak memuat nilai nol. (Nadhifah, dkk., 2012)

Evaluasi ketepatan klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan yang dilakukan oleh suatu fungsi klasifikasi. Nilai APER (*Apparent Error Rate*) menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Penentuan ketepatan pengklasifikasian dapat diketahui melalui tabel klasifikasi sebagai berikut. (Johnson & Wichern, 1992)

Tabel 1. Perhitungan Ketepatan Pengklasifikasian

Kategori Aktual	Kategori Prediksi		
	y=1	*y*=2	*y*=3
y=1	n_{11}	n_{12}	n_{13}
y=2	n_{21}	n_{22}	n_{23}
y=3	n_{31}	n_{32}	n_{33}

Ketepatan klasifikasi = $1 - APER$ di mana nilai APER diperoleh dari persamaan berikut.

$$APER(\%) = \left(\frac{n_{12}+n_{13}+n_{21}+n_{23}+n_{31}+n_{32}}{n_{11}+n_{12}+n_{13}+n_{21}+n_{22}+n_{23}+n_{31}+n_{32}+n_{33}}\right) \times 100\% \quad (6)$$

Hakikat Peminatan

Peminatan adalah proses yang berkesinambungan untuk memfasilitasi peserta didik mencapai tujuan pendidikan nasional, dan oleh karena itu peminatan harus berpijak pada kaidah-kaidah dasar yang secara eksplisit dan implisit, terkandung dalam kurikulum. Pendalaman mata pelajaran merupakan aktivitas tambahan dalam belajar yang dilakukan oleh peserta didik yang memiliki kecerdasan dan bakat istimewa. Tujuan pendalaman mata pelajaran adalah untuk meluaskan dan memperdalam materi mata pelajaran tertentu sesuai dengan arah minatnya. Pendalaman mata pelajaran merujuk pada isi dan proses. Isi merujuk pada apa yang ada dalam materi yang diperkaya dan lebih sulit. Proses merujuk pada prosedur mental pemecahan masalah, pemikiran kreatif, pemikiran ilmiah, pemikiran kritis, perencanaan, analisis dan banyak keterampilan pemikiran lainnya. (Kemendikbud, 2013)

Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini merupakan peserta didik baru di MAN 2 Samarinda Tahun Ajaran 2018/2019. Dari data tersebut dilakukan pemodelan regresi logistik multinomial dengan penaksiran parameter menggunakan metode Bayes. Adapun variabel bebas yang digunakan adalah nilai IPA SMP semester 5, nilai IPS SMP semester 5, nilai Bahasa SMP semester 5 dan rata-rata nilai UN SMP. Serta yang menjadi variabel terikat adalah minat peserta didik.

Deskripsi Data Penelitian

Deskripsi data variabel bebas dapat dilihat pada Tabel 2.

Tabel 2. Statistik Deskriptif Variabel Bebas

Variabel	*Mean*	*Min*	*Max*	Simpangan Baku
X_1	83,98	60	98	5,92
X_2	84,47	78	98	4,09
X_3	86,22	72	94	4,90
X_4	65,85	43	88	9,65

p-ISSN 2085-7829
e-ISSN 2798-3455

Berdasarkan Tabel 2 diperoleh informasi bahwa rata-rata nilai IPA SMP semester 5 sebesar 83,98 dengan nilai minimum yang diperoleh peserta didik sebesar 60 dan nilai maksimumnya sebesar 98. Rata-rata nilai IPS SMP semester 5 sebesar 84,47 dengan nilai minimum yang diperoleh sebesar 78 dan nilai maksimumnya sebesar 98. Sedangkan nilai Bahasa SMP semester 5 memiliki rata-rata sebesar 86,22 dengan nilai minimum sebesar 72 dan nilai maksimum sebesar 94, serta nilai rata-rata untuk nilai UN SMP sebesar 65,85 dengan nilai minimum sebesar 43 dan nilai maksimum sebesar 88. Nilai simpangan baku untuk setiap variabel bebas berturut-turut adalah 5,92, 4,09, 4,90 dan 9,65. Nilai ini lebih kecil daripada nilai rata-rata untuk masing-masing variabel bebas yang menandakan bahwa sebaran data kecil dan kurang bervariasi.

Estimasi Parameter Model Regresi Logistik Multinomial dengan MLE

Langkah ini dilakukan untuk memperoleh nilai estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) tanpa dilakukan pengujian simultan maupun parsial. Hasil estimasi parameter regresi logistik disajikan pada Tabel 3.

Tabel 3. Estimasi Parameter Regresi Logistik dengan MLE

Variabel	Parameter	Estimasi
Konstanta	β_{10}	-22,28
Nilai IPA	β_{11}	0,13
Nilai IPS	β_{12}	0,05
Nilai Bahasa	β_{13}	0,01
Rata-rata UN	β_{14}	0,10
Konstanta	β_{20}	19,93
Nilai IPA	β_{21}	-0,05
Nilai IPS	β_{22}	-0,03
Nilai Bahasa	β_{23}	-0,07
Rata-rata UN	β_{24}	-0,13

Setelah diperoleh nilai estimasi parameter regresi logistik multinomial menggunakan metode MLE, maka nilai estimasi tersebut dapat digunakan sebagai nilai *prior* untuk etimasi parameter regresi logistik multinomial dengan metode Bayes.

Pemodelan Regresi Logistik Multinomial dengan Metode Bayes

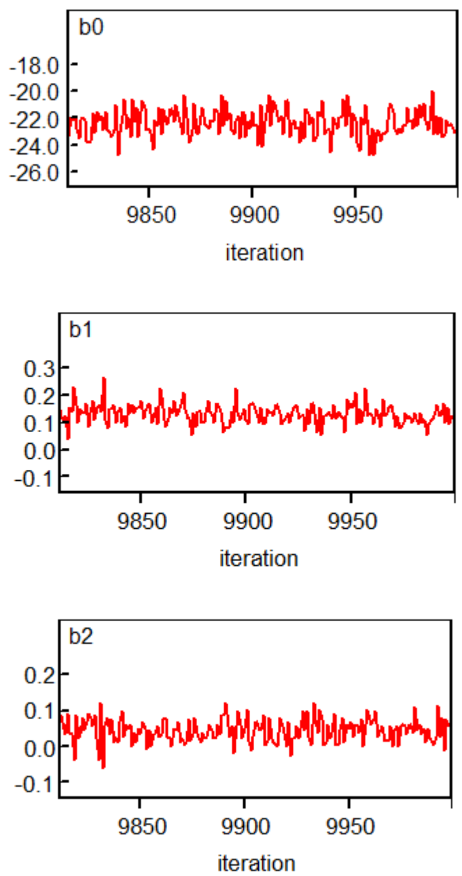
Langkah pertama yang dilakukan adalah menentukan *prior* parameter β_p. Distribusi *prior* untuk β_p adalah distribusi normal dengan nilai parameter μ_p adalah hasil estimasi parameter β_p yang diperoleh melalui cara frekuentis pada MLE dan σ_p^2 bernilai 1.

β_{10}~dnorm(-22,28, 1)
β_{11}~dnorm(0,13, 1)
β_{12}~dnorm(0,05, 1)
β_{13}~dnorm(0,01, 1)
β_{14}~dnorm(0,10, 1)
β_{20}~dnorm(19,93, 1)
β_{21}~dnorm(-0,05, 1)
β_{22}~dnorm(-0,03, 1)
β_{23}~dnorm(-0,07, 1)
β_{24}~dnorm(-0,13, 1)

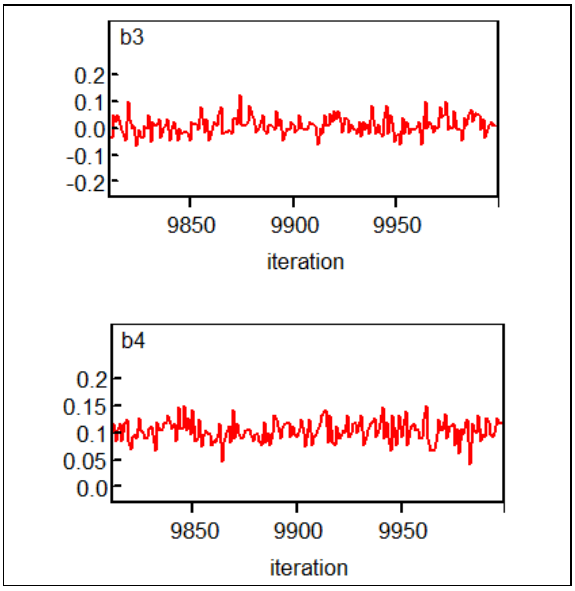
Distribusi *prior* yang digunakan adalah distribusi normal, sehingga fungsi densitasnya adalah

$$f(\beta_p) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left\{-\frac{1}{2}\left(\frac{\beta_p - \mu_p}{\sigma_p}\right)^2\right\}$$

Langkah selanjutnya adalah menjalankan simulasi MCMC melalui algoritma *gibbs sampler* dengan iterasi sebanyak 10.000 dengan *thin*=750 dan diperoleh hasil bahwa proses estimasi yang dilakukan telah mencapai kondisi yang konvergen. Berikut MCMC *diagnostic plot* untuk setiap parameter β_p.

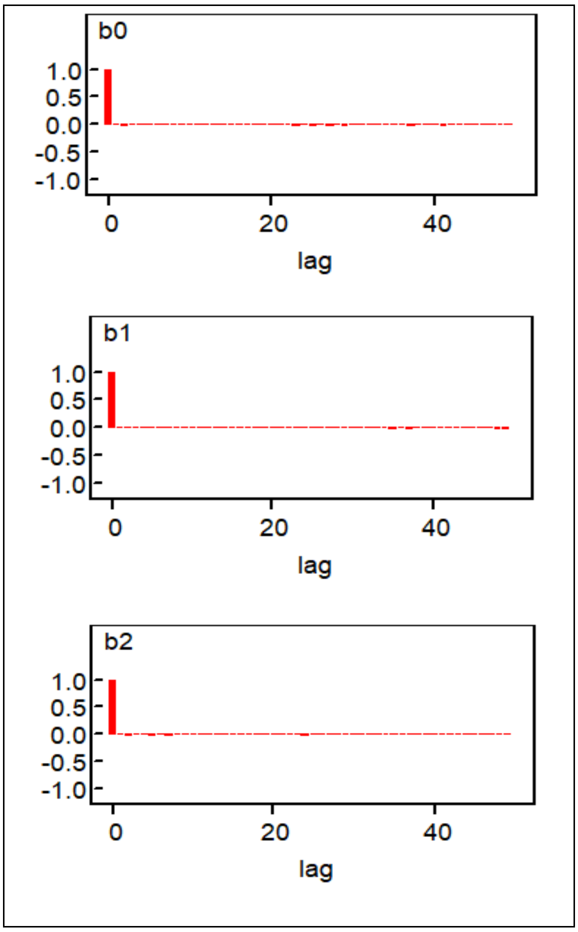


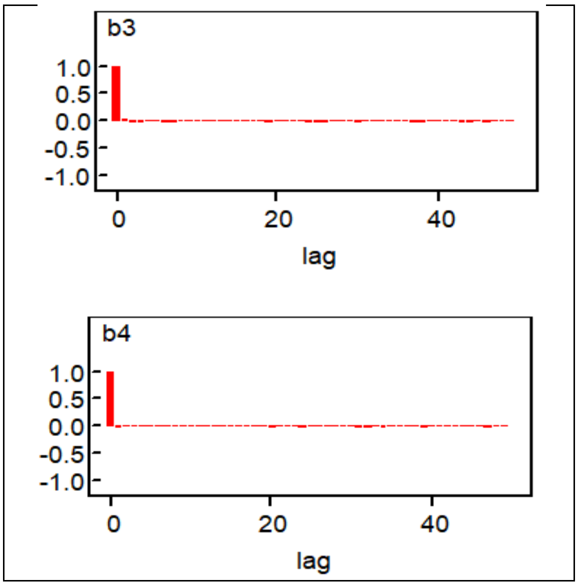
p-ISSN 2085-7829
e-ISSN 2798-3455



Gambar 1 *Trace plot*

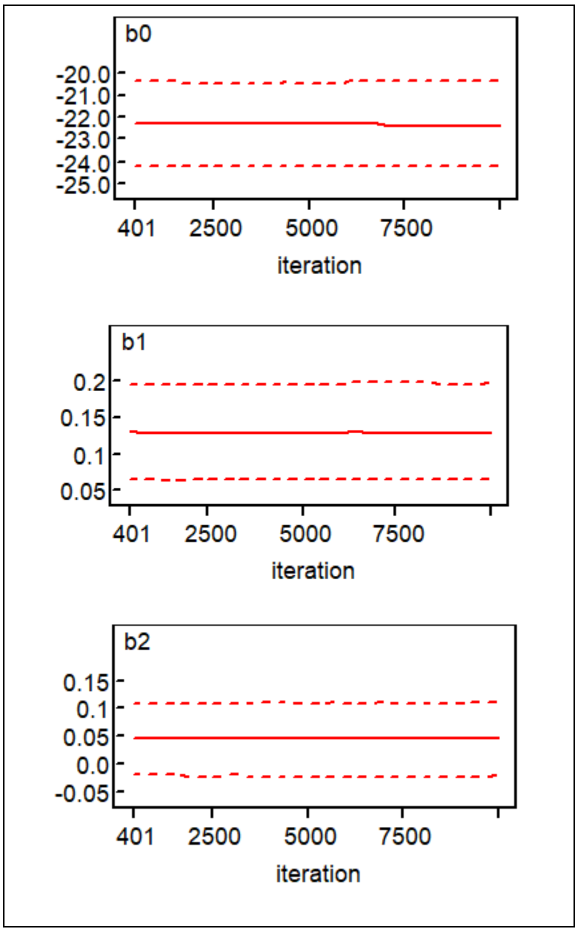
Pada Gambar 1 dapat dilihat bahwa *trace plot* untuk setiap parameter yang diamati terlihat tidak teratur dan tidak membentuk suatu pola tertentu, sehingga dapat dikatakan bahwa iterasi telah konvergen.



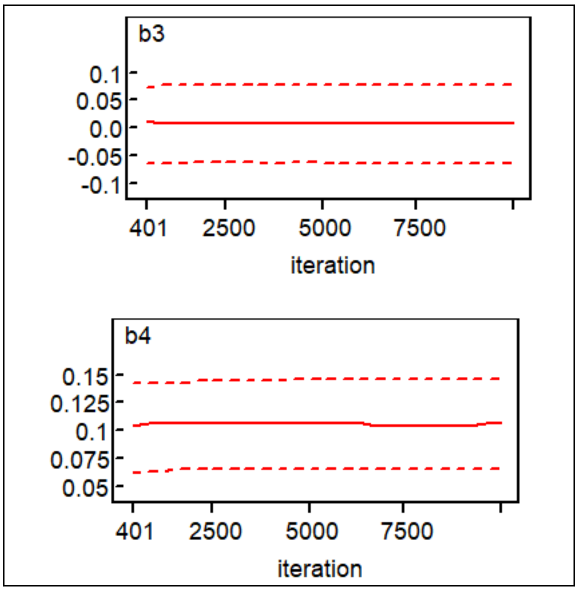


Gambar 2 *Autocorrelation plot*

Gambar 2 menunjukkan plot otokorelasi pada setiap parameter yang diamati. Terlihat bahwa nilai otokorelasi pada lag nol mendekati satu dan pada lag-lag selanjutnya (lag 1 hingga lag 50) mendekati nol sehingga dapat dikatakan bahwa iterasi telah konvergen.



p-ISSN 2085-7829
e-ISSN 2798-3455



Gambar 3 *Running quantiles plot*

Gambar 3 menunjukkan bahwa nilai median untuk setiap parameter yang diamati sudah mencapai nilai yang stabil (digambarkan dengan garis lurus) dan berada dalam interval kredibel. Interval kredibel mempunyai batas bawah yaitu kuantil 2,5% dan batas atas yaitu kuantil 97,5%.

Tabel 4. Estimasi Parameter Regresi Logistik Multinomial Bayes

Parameter	Estimasi	Kuantil (2,5%)	Kuantil (97,5%)
β_{10}	-22,32	-24,25	-20,37
β_{11}	0,13	0,07	0,20
β_{12}	0,05	-0,02	0,11
β_{13}	0,01	-0,06	0,08
β_{14}	0,11	0,07	0,15
β_{20}	19,96	18,04	21,84
β_{21}	-0,05	-0,12	0,02
β_{22}	-0,03	-0,10	0,05
β_{23}	-0,07	-0,14	$6{,}27\text{x}10^{-4}$
β_{24}	-0,13	-0,18	-0,09

Berdasarkan Tabel 4 diketahui variabel yang berpengaruh adalah X_1 dan X_4 di mana X_1 merupakan nilai IPA SMP semester 5 dan X_4 merupakan rata-rata nilai UN SMP. Variabel tersebut dikatakan berpengaruh karena memiliki nilai kuantil atau interval kredibel antara 2,5% dan 97,5% yang tidak memuat nilai nol. Persamaan model yang diperoleh adalah sebagai berikut.

Fungsi logit:

$$g_1(x) = -22{,}32 + 0{,}13X_1 + 0{,}05X_2 + 0{,}01X_3 + 0{,}11X_4$$

$$g_2(x) = 19{,}96 - 0{,}05X_1 - 0{,}03X_2 - 0{,}07X_3 - 0{,}13X_4$$

Sebelum dilakukan interpretasi, maka terlebih dahulu mencari nilai eksponensial untuk setiap parameter. Nilai eksponensial untuk setiap parameter disajikan pada Tabel 5 berikut.

Tabel 5. Nilai Eksponensial Setiap Parameter

Parameter	Estimasi	Nilai Eksponensial
β_{10}	-22,32	$2{,}02\text{x}10^{-4}$
β_{11}	0,13	1,14
β_{12}	0,05	1,05
β_{13}	0,01	1,01
β_{14}	0,11	1,11
β_{20}	19,96	466141596
β_{21}	-0,05	0,95
β_{22}	-0,03	0,97
β_{23}	-0,07	0,93
β_{24}	-0,13	0,88

Interpretasi fungsi logit pertama untuk setiap variabel bebas jika variabel bebas lainnya dianggap konstan yaitu setiap penambahan 1 nilai IPA maka akan meningkatkan peluang peserta didik 1,14 kali masuk jurusan IPA dibandigkan masuk jurusan Bahasa, dan setiap penambahan 1 rata-rata nilai UN maka akan meningkatkan peluang peserta didik 1,11 kali masuk jurusan IPA dibandingkan masuk jurusan Bahasa. Interpretasi fungsi logit kedua untuk setiap penambahan 1 rata-rata nilai UN maka akan meningkatkan peluang peserta didik 0,88 kali masuk jurusan IPS dibandingkan masuk jurusan Bahasa.

Setelah diperoleh nilai peluang untuk setiap kategori variabel terikat maka dapat dihitung ketepatan klasifikasinya untuk mengetahui peluang kesalahan yang dilakukan oleh model. Ketepatan klasifikasi yang diperoleh dapat dilihat pada hasil perhitungan berikut.

Tabel 6. Prediksi Ketepatan Klasifikasi Model

Aktual	Prediksi		
	IPA	IPS	Bahasa
IPA	111	17	16
IPS	7	35	7
Bahasa	25	21	13

Perhitungan nilai APER dan ketepatan klasifikasi sebagai berikut.

$$APER(\%) = \left(\frac{\begin{array}{c} n_{12} + n_{13} + n_{21} \\ +n_{23} + n_{31} + n_{32} \end{array}}{\begin{array}{c} n_{11} + n_{12} + n_{13} + n_{21} + n_{22} \\ +n_{23} + n_{31} + n_{32} + n_{33} \end{array}} \right) \times 100\%$$

$$= \left(\frac{\begin{array}{c} 17 + 16 + 7 \\ +7 + 25 + 21 \end{array}}{\begin{array}{c} 111 + 17 + 16 + 7 + 35 \\ +7 + 25 + 21 + 13 \end{array}} \right) \times 100\%$$

$$= \frac{93}{252} \times 100\% = 36{,}90\%$$

Maka ketepatan klasifikasinya

$$\begin{aligned} 1 - APER &= 1 - 36{,}90\% \\ &= 100\% - 36{,}90\% \\ &= 63{,}10\% \end{aligned}$$

Ketepatan klasifikasi berdasarkan model yang telah terbentuk yaitu sebesar 63,10% yang berarti banyaknya prediksi yang tepat diklasifikasikan sesuai dengan observasi adalah sebesar 63,10%

p-ISSN 2085-7829
e-ISSN 2798-3455

dan kesalahan klasifikasi yang dihasilkan adalah sebesar 36,90%.

Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan maka diperoleh kesimpulan sebagai berikut.

1. Model regresi logistik multinomial Bayes pada data minat peserta didik MAN 2 Samarinda Tahun Ajaran 2018/2019 adalah sebagai berikut.
   a. Model fungsi logit minat peserta didik kategori IPA

$$g_1(x) = -22{,}32 + 0{,}13X_1 + 0{,}05X_2 + 0{,}01X_3 + 0{,}11X_4$$

   b. Model fungsi logit minat peserta didik kategori IPS

$$g_2(x) = 19{,}96 - 0{,}05X_1 - 0{,}03X_2 - 0{,}07X_3 - 0{,}13X_4$$

2. Variabel yang berpengaruh secara signifikan terhadap minat peserta didik MAN 2 Samarinda Tahun Ajaran 2018/2019 adalah nilai IPA SMP semester 5 dan rata-rata nilai ujian nasional (UN).

Daftar Pustaka

Agresti, Alan. (2002). *Categorical Data Analysis*. New York: John Wiley and Sons Inc.

Ainul, A. M. (2018). Penerapan Model Analisis Regresi Linier Berganda Dengan Pendekatan Bayesian Pada Data Aset Bank di Indonesia. *Jurnal Keteknikan dan Sains (JUTEKS)*, 1 (1), 41-47.

Arieska. (2012). Perbandingan Metode *Maximum Likelihood Estimation (MLE)* Dengan *Bayesian* Pada Regresi Logistik Multinomial. *Jurnal Statistika*, 4 (1), 7-14.

Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. London: Addison-Wesley Publishing Company.

Carlin, B. P. & Chip, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of The Royal Statistical Association*, 91 (434), 473-484.

Casella, G. & George, E. I. (1992). Explaining Gibbs Sampler. *Journal of The American Statistical Association*, 46 (3), 167–174.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons.

Johnson, R. A. & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Kemendikbud. (2013). *Materi Diklat Peminatan Peserta Didik*. Diambil 08 Desember, 2020, dari https://akhmadsudrajat.word-press.com/2013/07/15/-download-materi-diklat-peminatan-peserta-didik/comment-page-1/.html

Kleinbaum, David G. (1994). *Logistic Regression, a self Learning Text*. New York: Springer Verlag.

Lancaster, Tony. (2003). *An Introduction to Bayesian Econometrics*. Lanchasire: Lancaster University.

Link, W. A. & Eaton, M. J. (2012). On Thinning of chain in MCMC. *Method in Ecology and Evolution*, 3, 112-115.

Nadhifah, L., Yasin, H. & Sugito. (2012). Analisis Faktor-Faktor Yang Mempengaruhi Bayi Berat Lahir Rendah Dengan Model Regresi Logistik Biner Menggunakan Metode Bayes. *Jurnal Gaussian*, 1 (1), 125-134.

Ntzoufraz, I. (2009). *Bayesian Modelling Using WinBUGS*. New Jersey: John Wiley and Sons, Inc.

Subanar. (2019). *Inferensi Bayesian Dengan R*. Yogyakarta: Gadjah Mada University Press.