

Regresi Logistik dengan Metode Bayes untuk Pemodelan Indeks Pembangunan Manusia Kabupaten/Kota di Pulau Kalimantan*Logistic Regression with Bayesian Method for The Modelling of Human Development Index Districts/Cities in Kalimantan Island***Febriana Syafitri¹, Rito Goejantoro², Wasono³**¹Laboratorium Statistika Ekonomi dan Bisnis FMIPA Universitas Mulawarman²Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman³Laboratorium Matematika Komputasi FMIPA Universitas MulawarmanE-mail: febrianasyafitri14@gmail.com**ABSTRACT**

Human Development Index (HDI) is an indicator that can measure success in efforts to build the quality of human life. HDI is also a measure of the prosperity of a region which is observed based on three dimensions, namely health, education and economy. Based on HDI publication by the Central Statistics Agency in 2018, it showed that the scores of HDI for 56 districts/cities in Kalimantan Island only has two categories of HDI which are medium and high. Bayesian method is a parameter estimation technique that combines the likelihood and prior distribution functions. The estimation with Bayesian method was solved using Markov Chain Monte Carlo simulation (MCMC) with Gibbs Sampler algorithm. The aim of this study is to examine the modelling of the factors that influence the HDI of districts/cities in Kalimantan Island and determine the accuracy of the model classification using logistic regression with Bayesian method. The data used is the HDI of districts/cities in Kalimantan Island in 2018. Bayesian method is a parameter estimation technique that combines the likelihood and prior distribution functions. The estimation with Bayesian method was solved using Markov Chain Monte Carlo simulation (MCMC) with Gibbs Sampler algorithm. The results of modelling and analysis on districts/cities HDI data on Kalimantan Island showed that the factors that significantly influence HDI are the number of paramedic, the number of health facility and the participation rate of high school. The results of the classification accuracy of the model amounted to 82,14% which resulted in 37 districts/cities are categorized as the HDI medium category and 19 districts/cities are categorized as the HDI high category.

Keywords: Bayesian Method, HDI, Logistic Regression, MCMC

Pendahuluan

Pembangunan merupakan suatu langkah dalam membuat suatu perubahan yaitu membuat sesuatu menjadi lebih baik atau meningkat. Keberhasilan pembangunan khususnya pembangunan manusia dapat dinilai secara parsial dengan melihat seberapa besar permasalahan yang paling mendasar dapat teratasi. Permasalahan tersebut di antaranya adalah kemiskinan, pengangguran, buta huruf, ketahanan pangan dan penegakan demokrasi. Salah satu indikator yang dapat mengukur keberhasilan dalam upaya membangun kualitas hidup manusia yaitu Indeks Pembangunan Manusia (IPM) (Nurmalasari dkk, 2017).

Menurut BPS (2017), IPM adalah suatu indikator yang diterbitkan oleh *United Nations Development Programme* (UNDP) untuk mengukur kesuksesan pembangunan dan kesejahteraan suatu daerah. IPM adalah suatu tolak ukur angka kesejahteraan daerah yang dilihat berdasarkan tiga dimensi yaitu: kesehatan, pendidikan, dan ekonomi.

IPM di Pulau Kalimantan hanya memiliki dua kategori yaitu sedang dan tinggi. Berdasarkan data

BPS, pada tahun 2018 IPM Kalimantan Barat merupakan salah satu IPM kategori sedang dengan capaian sebesar 66,98% sedangkan di Provinsi Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur dan Kalimantan Utara termasuk ke dalam nilai IPM kategori tinggi dengan capaian berturut-turut yaitu 70,42%, 70,17%, 75,83% dan 70,56%. Capaian IPM di Pulau Kalimantan tahun 2018 terlihat sangat berbeda-beda di setiap provinsi, maka perlu dikaji faktor-faktor yang mempengaruhi perbedaan capaian IPM tersebut.

Regresi logistik tepat digunakan apabila variabel respon berbentuk kategorik dan variabel prediktornya diskrit, kontinu ataupun gabungan dari keduanya (Agesti, 2002). Estimasi parameter yang sering digunakan dalam pemodelan regresi logistik yaitu *Maximum Likelihood Estimation* (MLE), akan tetapi asumsi yang harus dipenuhi adalah tidak ada multikolinieritas dan tidak ada autokorelasi antar observasi. Salah satu estimasi parameter yang lain dalam regresi logistik adalah estimasi *Bayesian*. Menurut Teti, dkk (2014), metode Bayes memandang parameter sebagai variabel acak yang menggambarkan pengetahuan

awal tentang parameter sebelum pengamatan dilakukan dan dinyatakan dalam suatu distribusi yang disebut dengan distribusi *prior*. Informasi dalam distribusi *prior* dikombinasikan dengan informasi data sampel melalui teorema Bayes dan hasilnya dinyatakan dalam bentuk distribusi yang disebut *posterior* yang digunakan menjadi dasar untuk inferensi di dalam metode Bayes.

Berdasarkan uraian di atas, penulis tertarik untuk melakukan penelitian dengan judul Regresi Logistik dengan Metode Bayes untuk Pemodelan Indeks Pembangunan Manusia Kabupaten/Kota di Pulau Kalimantan.

Regresi Logistik

Regresi logistik merupakan suatu teknik analisis data dalam statistika yang bertujuan untuk mengetahui hubungan antara beberapa variabel di mana variabel responnya adalah bersifat kategorik, baik nominal maupun ordinal dengan variabel prediktornya dapat bersifat kategorik atau kontinu. Regresi logistik tidak memodelkan secara langsung variabel respon (Y) dengan variabel prediktor (X), melainkan melalui transformasi variabel respon ke variabel logit yang merupakan *natural log* dari *odds* rasio (Ramli dkk, 2013). Untuk mencari persamaan logistiknya, model yang dipakai adalah:

$$\pi(x) = \frac{e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}} \quad (1)$$

di mana $\pi(x)$ merupakan nilai peluang dari $0 \leq \pi(x) \leq 1$, yang berarti bahwa regresi logistik menggambarkan suatu peluang. Untuk mempermudah menaksir parameter regresi, maka $\pi(x)$ pada persamaan (1) ditransformasikan dengan menggunakan transformasi logit $g(\pi(x))$. Uraian transformasi tersebut adalah sebagai berikut:

$$1 - \pi(x) = 1 - \frac{e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}}$$

$$= \frac{1}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}}$$

Sehingga $\frac{\pi(x)}{1 - \pi(x)}$ sebagai berikut:

$$\frac{\pi(x)}{1 - \pi(x)} = \frac{e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}} \times \frac{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}}{1}$$

$$= e^{\beta_0 + \sum_{j=1}^p \beta_j X_j}$$

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \ln(e^{\beta_0 + \sum_{j=1}^p \beta_j X_j})$$

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \sum_{j=1}^p \beta_j X_j$$

Penaksiran Parameter Model Regresi Logistik

Prosedur mengestimasi nilai parameter dalam regresi logistik yaitu dengan menggunakan metode kemungkinan maksimum atau biasa disebut metode *Maximum Likelihood Estimation* (MLE) yang digunakan untuk penaksiran parameter di mana koefisien memaksimalkan kemungkinan (*likelihood*) dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. (Hajarisman, 2010).

Pengujian Signifkansi Parameter Model Regresi Logistik

Pengujian signifikansi parameter dalam regresi penting untuk dilakukan. Hal ini dikarenakan pengujian tersebut digunakan untuk menentukan apakah pengaruh variabel prediktor dalam model signifikan terhadap variabel respon. Pengujian signifikansi parameter terdiri dari dua uji yaitu sebagai berikut

1) Pengujian Signifkansi Parameter Secara Simultan

Uji simultan bertujuan untuk mengetahui pengaruh variabel prediktor secara serentak terhadap variabel respon (Hosmer dkk., 2013).

Hipotesis:

- $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$
(semua variabel prediktor tidak berpengaruh terhadap variabel respon)
- $H_1 : \text{Paling sedikit ada satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$
(minimal ada satu variabel prediktor yang berpengaruh terhadap variabel respon)

Statistik Uji:

Statistik uji yang digunakan adalah uji *G Likelihood Ratio*

$$G = -2 \ln \left(\frac{\left(\frac{n_1}{n_0}\right)^{n_1} \left(\frac{n_0}{n_1}\right)^{n_0}}{\prod_{i=1}^n \pi^{y_i} (1-\pi)^{(1-y_i)}} \right) \quad (2)$$

Kriteria pengujian:

H_0 ditolak jika $G \geq \chi^2_{(\alpha, j)}$ atau nilai probabilitas $p \leq \alpha$.

2) Pengujian Signifkansi Parameter Secara Parsial

Uji parsial dilakukan untuk menguji setiap variabel β_j secara individual. Hasil pengujian secara individual akan menunjukkan apakah suatu variabel prediktor layak untuk masuk model atau tidak (Hosmer dkk., 2013).

Hipotesis:

- $H_0 : \beta_j = 0$
(tidak ada pengaruh variabel prediktor ke-j terhadap variabel respon), di mana $j = 1, 2, \dots, p$
- $H_1 : \beta_j \neq 0, \text{ untuk suatu } \beta_j$
(ada pengaruh variabel prediktor ke-j terhadap variabel respon) di mana $j = 1, 2, \dots, p$

Statistik Uji:

Statistik uji yang digunakan adalah uji Wald

$$W^2 = \left[\frac{\beta_j}{SE(\beta_j)} \right]^2 \quad (3)$$

Kriteria Pengujian:

H_0 ditolak jika $W^2 \geq \chi^2_{(\alpha,1)}$ atau nilai probabilitas $p \leq \alpha$.

Ketepatan Klasifikasi Regresi Logistik

Menurut Prasetyo (2014), sebuah sistem klasifikasi perlu diukur kinerjanya untuk dapat melakukan klasifikasi semua *set* data dengan benar. Umumnya, pengukuran ketepatan klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*), seperti pada Tabel 1.

Tabel 1. Matriks Konfusi Klasifikasi Dua Kelas

Hasil Observasi	Prediksi	
	y_0	y_1
y_0	f_{00}	f_{01}
y_1	f_{10}	f_{11}

Untuk menghitung akurasi digunakan persamaan sebagai berikut:

$$\text{Ketepatan klasifikasi} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \times 100\% \quad (4)$$

Teorema Bayes

Teorema Bayes merupakan salah satu metode pemilihan model terbaik yang melibatkan ketidakpastian model yaitu dengan merata-ratakan distribusi posterior dari semua model yang mungkin. Tujuan dari metode Bayes adalah menggabungkan model-model yang tidak pasti sehingga didapat suatu model yang terbaik. Misalnya peristiwa-peristiwa A_1, A_2, \dots, A_k membentuk suatu partisi di dalam ruang sampel S sedemikian sehingga $P(A_i) > 0$ dengan i bernilai 1, 2, ..., k dan B sembarang peristiwa sedemikian sehingga $P(B) > 0$, maka

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^k P(A_i)P(B|A_i)} \quad (5)$$

Teorema Bayes memberikan aturan sederhana untuk menghitung probabilitas bersyarat peristiwa A_i jika B telah terjadi, yaitu jika masing-masing probabilitas tak bersyarat A_i dan probabilitas B dengan A_i diketahui (Soejoeti dan Soebanar, 1988).

Prior

Untuk memperoleh estimasi Bayes suatu parameter perlu diketahui bentuk distribusi *prior*-nya. Berikut adalah beberapa *prior* dalam *Bayesian* yaitu:

1. *Prior* konjugat atau non-konjugat yang ditentukan berdasarkan pola *likelihood* dari data atau tidak.

2. *Prior* informatif dan non-informatif merupakan *prior* yang ditentukan berdasarkan pada ketersediaan informasi sebelumnya mengenai pola distribusi data. (Lancaster, 2004).
3. *Pseudo prior* adalah *prior* terkait dengan pemberian nilai yang disetarakan dengan hasil elaborasi dari *frequentist* (Carlin dan Chib, 1995).

Posterior

Menurut Box dan Tiao (1973), hal terpenting dalam Bayes terletak pada penentuan distribusi *posterior* $f(\theta|y)$. Berdasarkan aturan probabilitas dalam teorema Bayes, distribusi *posterior* dari parameter θ dapat dinyatakan dalam persamaan berikut

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \quad (6)$$

Uji Hipotesis

Dalam statistika *Bayesian*, *credible interval* merupakan interval di dalam domain dari distribusi probabilitas *posterior* yang digunakan untuk penaksiran interval. Salah satu metode untuk mengestimasi *credible interval* yang paling mudah digunakan adalah interval kredibel dua ekor. *Credible interval* dua ekor disusun dengan menemukan kuantil $\alpha/2$ dan $1 - \alpha/2$ dengan tingkat signifikansi α (Johnson, 2009).

Markov Chain Monte Carlo

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. Untuk melihat konvergensi dari hasil estimasi parameter posterior bisa dilihat dari:

1. *Trace Plot*
Salah satu cara pendugaan *burn-in period* adalah memeriksa *trace plot* nilai simulasi dari komponen atau beberapa fungsi lainnya dari x terhadap jumlah iterasi. *Trend* naik turun pada nilai parameter pada *trace plot* menunjukkan bahwa *burn-in period* belum tercapai. Jika semua nilai-nilai berada dalam sebuah daerah tanpa periodik yang kuat cenderung dapat dikatakan konvergen.
2. *Autocorrelation Plot*
Nilai simulasi x pada iterasi ke- $(t+1)$ bergantung pada nilai simulasi pada iterasi

ke-t. Jika pada *autocorrelation plot* pada lag pertama mendekati satu dan selanjutnya nilai-nilainya terus berkurang menuju 0 dapat dikatakan bahwa iterasi telah konvergen.

3. *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.

Gibbs Sampler

Gibbs Sampler sebagai suatu teknik simulasi untuk membangkitkan variabel random dari suatu distribusi tertentu secara tidak langsung, tanpa harus menghitung fungsi densitas dari suatu distribusi data. *Gibbs Sampler* dapat diterapkan apabila distribusi probabilitas bersama (*joint probability distribution*) tidak diketahui secara eksplisit, tetapi distribusi bersyarat dari tiap-tiap variabel diketahui (Casella dan George, 1992).

Menurut Teti, dkk (2014) langkah-langkah dari algoritma *Gibbs Sampler* adalah:

- Langkah 1 : Menentukan nilai awal
 $\beta^0 = \{\beta_0^0, \beta_1^0, \beta_2^0, \dots, \beta_p^0\}$
 - Langkah 2 : Membangkitkan tiap komponen dari $\beta^t = \{\beta_0^t, \beta_1^t, \beta_2^t, \dots, \beta_p^t\}$ untuk iterasi $t = 1, 2, \dots, T$
- Di mana
- Nilai β_0^t berasal dari $f(\beta_0 | \beta_1^{t-1}, \beta_2^{t-1}, \dots, \beta_p^{t-1}, \mathbf{x})$
 - Nilai β_1^t berasal dari $f(\beta_1 | \beta_0^t, \beta_2^{t-1}, \dots, \beta_p^{t-1}, \mathbf{x})$
 - Nilai β_2^t berasal dari $f(\beta_2 | \beta_0^t, \beta_1^t, \dots, \beta_p^{t-1}, \mathbf{x})$
 - ⋮
 - Nilai β_p^t berasal dari $f(\beta_p | \beta_0^t, \beta_1^t, \beta_2^t, \dots, \beta_{p-1}^t, \mathbf{x})$

Indeks Pembangunan Manusia

Indeks pembangunan manusia (IPM) merupakan indikator yang digunakan untuk melihat perkembangan manusia, terdapat dua aspek yang perlu diperhatikan, yaitu kecepatan dan status pencapaian. IPM digunakan untuk mengklasifikasikan apakah sebuah negara adalah negara maju, negara berkembang atau negara terbelakang dan juga untuk mengukur pengaruh dari kebijaksanaan ekonomi terhadap kualitas hidup. Perkembangan angka IPM memberikan indikasi peningkatan atau penurunan kinerja pembangunan manusia. IPM disusun dari tiga dimensi dasar yaitu umur panjang dan hidup sehat (*a long and a healthy life*), pengetahuan (*knowledge*) dan standar hidup yang layak (*decent standard of living*) (BPS, 2017).

Hasil dan Pembahasan

Data penelitian terdiri dari data variabel respon dan variabel prediktor. Data variabel respon yaitu data Indeks Pembangunan Manusia (IPM) Kabupaten/Kota di Pulau Kalimantan tahun 2018 yang dinotasikan (*Y*) dengan sampel sebanyak 56 data. Data variabel prediktor terdiri dari data jumlah tenaga kesehatan (X_1), jumlah sarana kesehatan (X_2), angka partisipasi sekolah (APS) SMA (X_3), tingkat pengangguran terbuka (TPT) (X_4) dan persentase penduduk miskin (X_5).

Deskripsi Data Penelitian

Deskripsi dari variabel prediktor disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Data Variabel Prediktor

Variabel	Mean	Max	Min	Variansi
X_1	1.124	4.013	189	553.506
X_2	360	1.522	6	60.533
X_3	70,98	85,72	44,37	72,03
X_4	4,43	10,37	1,58	3,51
X_5	6,27	12,83	2,64	5,91

Berdasarkan Tabel 2. terlihat bahwa rata-rata jumlah tenaga kesehatan kabupaten/kota di Pulau Kalimantan adalah 1.124 orang dengan variansi sebesar 553.506. Jumlah tenaga kesehatan tertinggi adalah sebanyak 4.013 orang dan jumlah tenaga kesehatan terendah adalah sebanyak 189 orang. Rata-rata jumlah sarana kesehatan kabupaten/kota di Pulau Kalimantan adalah 360 unit dengan variansi sebesar 60.533. Jumlah sarana kesehatan tertinggi adalah sebanyak 1.522 unit dan jumlah sarana kesehatan terendah sebanyak 6 unit. Rata-rata APS SMA kabupaten/kota di Pulau Kalimantan adalah 70,98 persen dengan variansi sebesar 72,03. APS SMA tertinggi adalah kabupaten Penajam Paser Utara sebesar 85,72 persen dan APS SMA terendah adalah sebesar 44,37 persen. Rata-rata TPT kabupaten/kota di Pulau Kalimantan adalah 4,43 persen dengan variansi sebesar 3,51. TPT tertinggi adalah sebesar 10,37 persen dan TPT terendah adalah sebesar 1,58 persen. Rata-rata persentase penduduk miskin 56 kabupaten/kota di Pulau Kalimantan adalah 6,27 persen dengan variansi sebesar 5,91. Persentase penduduk miskin tertinggi adalah sebesar 12,83 persen dan persentase penduduk miskin terendah adalah sebesar 2,64 persen.

Pemodelan Regresi Logistik Biner

Pada penelitian ini model regresi logistik yang digunakan yaitu

$$g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Berikut adalah hasil pengujian signifikansi parameter secara simultan untuk mengetahui apakah variabel prediktor secara serentak

mempunyai pengaruh yang signifikan terhadap variabel respon.

Hipotesis pengujian secara simultan adalah

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1 : \text{Paling sedikit satu } \beta_j \neq 0, j = 1, 2, 3, 4, 5$$

Hasil perhitungan uji signifikansi parameter secara simultan dapat dilihat pada Tabel 3.

Tabel 3. Pengujian Signifikansi Parameter Secara Simultan

G_{hitung}	$p\text{-value}$
32,602	0,000

Berdasarkan hasil perhitungan pada Tabel 3. keputusan uji adalah menolak H_0 pada taraf signifikansi 0,05, hal ini ditunjukkan oleh $p\text{-value} = 0,000 < \alpha = 0,05$. Kesimpulan uji hipotesis simultan adalah minimal ada satu variabel prediktor yang berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan.

Setelah dilakukan pengujian signifikansi parameter secara simultan maka dilanjutkan dengan pengujian signifikansi parameter secara parsial. Pengujian signifikansi parameter secara parsial bertujuan untuk mengetahui apakah variabel prediktor secara individual berpengaruh terhadap variabel respon. Hipotesis pengujian secara parsial untuk j tertentu yaitu $j = 1, 2, 3, 4, 5$

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Hasil perhitungan uji signifikansi parameter secara parsial dapat dilihat pada Tabel 4

Tabel 4. Pengujian Signifikansi Parameter Secara Parsial

Variabel	Taksiran	W^2	$p\text{-value}$
$X_1(\beta_1)$	0,004288	4,2	0,041
$X_2(\beta_2)$	-0,007715	3,4	0,064
$X_3(\beta_3)$	0,078468	1,9	0,170
$X_4(\beta_4)$	0,385794	1,6	0,210
$X_5(\beta_5)$	-0,105569	0,32	0,570
β_0	-8,601579	-	-

Berdasarkan hasil perhitungan pada Tabel 4. diperoleh bahwa jumlah tenaga kesehatan (X_1) secara individual berpengaruh terhadap IPM kabupaten/kota di pulau Kalimantan, hal ini ditunjukkan oleh $p\text{-value} = 0,041 < \alpha = 0,05$. Jumlah sarana kesehatan (X_2), angka partisipasi sekolah (APS) SMA (X_3), tingkat pengangguran terbuka (TPT) (X_4) dan persentase penduduk miskin (X_5). masing-masing secara individual tidak berpengaruh terhadap IPM kabupaten/kota di pulau Kalimantan, hal ini ditunjukkan oleh $p\text{-value}$ masing-masing variabel tersebut lebih dari $\alpha = 0,05$. Oleh karena itu, perlu dilakukan pemilihan model regresi terbaik untuk mendapatkan model regresi dengan keseluruhan

parameter telah signifikan terhadap variabel respon IPM.

Model terbaik didapat dengan metode *backward elimination* yang berhenti pada iterasi terakhir yaitu *step 3*. Hasil pengujian signifikansi parameter untuk model terbaik dapat dilihat pada Tabel 5.

Tabel 5. Pengujian Signifikansi Parameter Secara Parsial (Model Terbaik)

Variabel	Taksiran	W^2	$p\text{-value}$
$X_1(\beta_1)$	0,004188	7,1	0,0079
$X_2(\beta_2)$	-0,006751	5,2	0,023
$X_3(\beta_3)$	0,112094	4,0	0,047
β_0	0,112094	-	-

Berdasarkan hasil perhitungan pada Tabel 5. diperoleh bahwa jumlah tenaga kesehatan (X_1), jumlah sarana kesehatan (X_2) dan angka partisipasi sekolah (APS) SMA (X_3) secara individual berpengaruh terhadap IPM kabupaten/kota di pulau Kalimantan, hal ini ditunjukkan oleh $p\text{-value}$ masing-masing variabel tersebut kurang dari $\alpha = 0,05$. Berdasarkan model terbaik dengan analisis regresi logistik biner menggunakan metode *backward elimination*, maka model logit yang terbentuk adalah $g(x) = -0,09289 + 3,11595X_1 - 1,66107X_2 + 0,95133X_3$

Pemodelan Regresi Logistik dengan Metode Bayes

Proses estimasi dilakukan setelah menentukan *prior* parameter β dan σ^2 . Penentuan *prior* β dilakukan dengan mengambil parameter β yang didapat melalui cara frekuentis (*pseudo prior*), sedangkan untuk nilai σ^2 adalah 1.

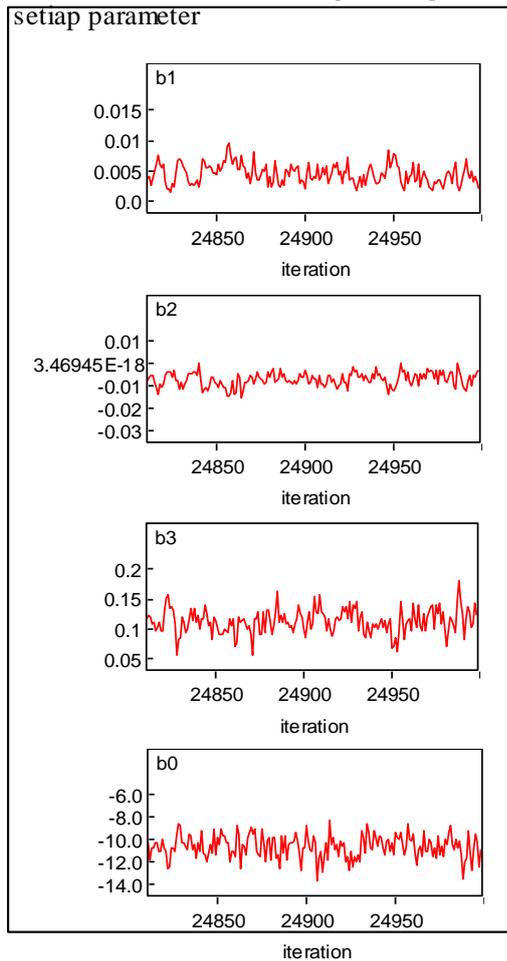
$$\begin{aligned} \beta_0 &\sim \text{dnorm}(-10,324539, 1) \\ \beta_1 &\sim \text{dnorm}(0,004188, 1) \\ \beta_2 &\sim \text{dnorm}(-0,006751, 1) \\ \beta_3 &\sim \text{dnorm}(0,112094, 1) \end{aligned}$$

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

$$f(\beta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{\left\{-\frac{1}{2}\left(\frac{\beta_j - \mu_j}{\sigma_j}\right)^2\right\}}$$

Proses estimasi dilakukan dengan simulasi MCMC melalui algoritma *Gibbs Sampler* dengan bantuan *software* WinBUGS. Banyaknya *update* yang dijalankan pada simulasi adalah $N=25.000$ dengan *thin*=10. Hasil yang diperoleh dari proses estimasi menunjukkan bahwa proses estimasi yang dilakukan telah mencapai kondisi konvergen. Kesimpulan ini didapatkan dari indikasi yang diperlihatkan oleh MCMC *diagnostic plot* yang terdiri dari *trace plot*,

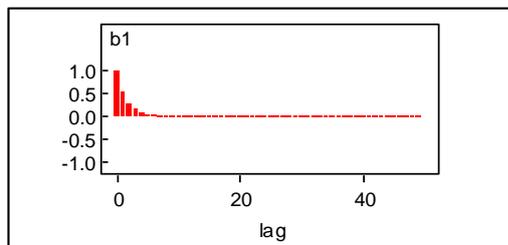
autocorrelation plot dan running quantiles plot. Berikut adalah MCMC diagnostic plot untuk setiap parameter



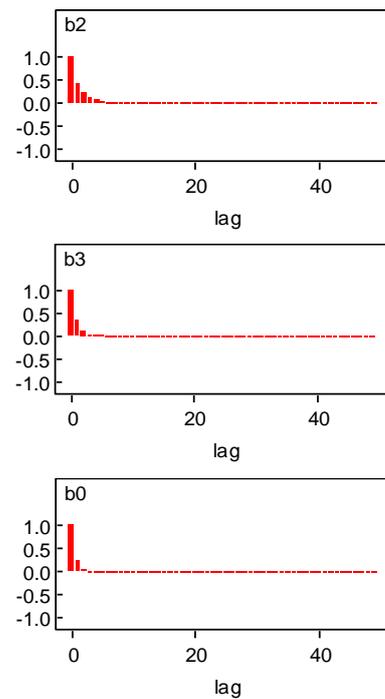
Gambar 1. Trace plot

Pada Gambar 1. dapat dilihat bahwa trace plot tidak membentuk pola tertentu hal ini dapat dikatakan iterasi sudah mencapai konvergen.

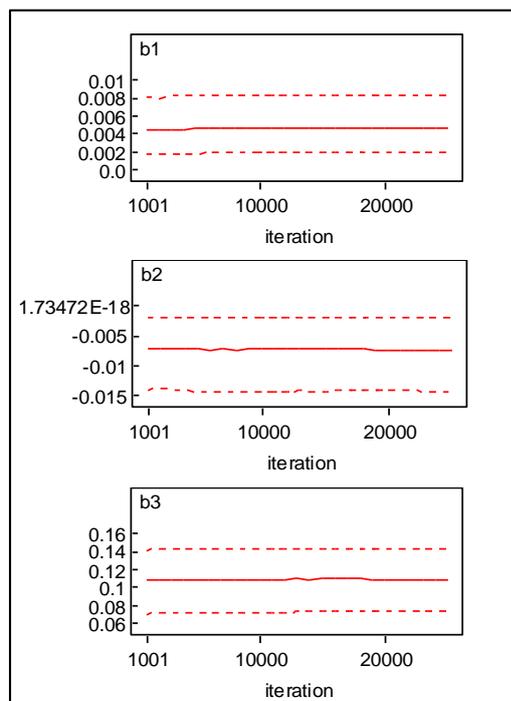
Pada Gambar 2. dapat dilihat bahwa nilai-nilai autokorelasi pada lag pertama mendekati satu dan selanjutnya nilai-nilai terus berkurang menuju 0 sehingga dapat dikatakan iterasi telah konvergen.



Gambar 2. Autocorrelation plot

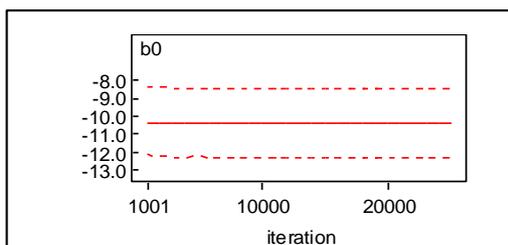


Gambar 2. Autocorrelation plot (lanjutan)



Gambar 3. Running quantiles plot

Gambar 3. menunjukkan bahwa nilai ergodic mean parameter yang diperoleh sudah mencapai nilai yang stabil dan berada dalam credible interval.



Gambar 3. Running quantiles plot (lanjutan)

Estimasi parameter yang digunakan adalah rata-rata dari nilai sampel hasil simulasi yang dihitung. Berikut ini merupakan hasil estimasi regresi logistik dengan metode Bayes.

Tabel 6. Hasil Estimasi Parameter Regresi Logistik dengan Metode Bayes

Variabel	Mean	Interval	
		Batas Atas (2,5%)	Batas Bawah (97,5%)
$X_1(\beta_1)$	0,00475	0,001965	0,008318
$X_2(\beta_2)$	-0,007509	-0,01422	-0,001929
$X_3(\beta_3)$	0,1092	0,0732	0,1437
β_0	-10,39	-12,29	-8,482

Berdasarkan Tabel 6. diperoleh *credible interval* 95% untuk masing-masing parameter adalah

$$\beta_1 = (0,001965; 0,008318),$$

$$\beta_2 = (-0,01422; -0,001929) \text{ dan}$$

$$\beta_3 = (0,0732; 0,1437),$$

Hal ini menunjukkan bahwa setiap taksiran parameter β_1 , β_2 dan β_3 signifikan pada iterasi ke-25000. Signifikansi terlihat dari nilai *quantile* pada *credible interval* antara 2,5% dan 97,5% yang tidak memuat nilai nol. Sehingga dapat disimpulkan bahwa jumlah tenaga kesehatan (X_1), jumlah sarana kesehatan (X_2) dan angka partisipasi sekolah (APS) SMA (X_3) berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan. Nilai yang digunakan dalam model adalah nilai *mean* pada Tabel 6, maka model logit yang terbentuk adalah

$$g(x) = 10,39 + 0,00475X_1 - 0,007509X_2 + 0,1092X_3$$

Kemudian fungsi probabilitas dari model logit yang terbentuk (1) adalah sebagai berikut

$$\pi(x) = \frac{e^{10,39+0,00475X_1-0,007509X_2+0,1092X_3}}{1 + e^{10,39+0,00475X_1-0,007509X_2+0,1092X_3}}$$

Setiap penambahan 1 orang tenaga kesehatan maka akan meningkatkan probabilitas IPM kabupaten/kota di Pulau Kalimantan tinggi menjadi 1,005 kali, lalu setiap penambahan 1 unit sarana kesehatan maka akan meningkatkan

probabilitas IPM kabupaten/kota di Pulau Kalimantan tinggi menjadi 1,008 kali dan setiap penambahan satu persen APS SMA maka akan meningkatkan probabilitas IPM kabupaten/kota di Pulau Kalimantan tinggi menjadi 1,115 kali.

Persentase ketepatan klasifikasi adalah rasio antara jumlah observasi-observasi yang diklasifikasikan secara tepat oleh model (sesuai dengan kelompok yang sebenarnya) dengan jumlah seluruh observasi. Hasil persentase ketepatan klasifikasi menggunakan regresi logistik dengan metode Bayes adalah sebagai berikut

Tabel 7. Ketepatan Klasifikasi Model

IPM	Prediksi		Total
	Sedang	Tinggi	
Sedang	30	3	33
Tinggi	7	16	23
Total	37	19	56

$$\text{Ketepatan klasifikasi} = \frac{30+16}{56} \times 100\% = 82,14\%$$

Kesimpulan

Berdasarkan hasil analisis dan pembahasan, maka diperoleh kesimpulan sebagai berikut

1. Model regresi logistik dengan metode Bayes yang digunakan dalam pemodelan IPM kabupaten/kota di Pulau Kalimantan adalah sebagai berikut

$$g(x) = 10,39 + 0,00475X_1 - 0,007509X_2 + 0,1092X_3$$
2. Faktor-faktor yang berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan adalah jumlah tenaga kesehatan (X_1), jumlah sarana kesehatan (X_2) dan angka partisipasi sekolah (APS) SMA (X_3).
3. Hasil ketepatan klasifikasi IPM kabupaten/kota di Pulau Kalimantan menggunakan regresi logistik dengan metode Bayes sebesar 82,14%.

Daftar Pustaka

- Agresti, A. (2002). *Categorical Data Analysis, Second Edition*. New York: John Wiley and Sons.
- Badan Pusat Statistik. (2017). *Indeks Pembangunan Manusia BPS Tahun 2016*. Samarinda: BPS.
- Box, G. E. P., and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons.
- Carlin, B. P., and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of The Royal Statistical Society*, 57 (3), 473-484.
- Casella, G., and George, E. I. (1992). Explaining The Gibbs Sampler. *Journal of The*

- American Statistical Association*, 46 (3), 167-174.
- Hajarisman, N. (2010). *Analisis Data Kategorik*. Bandung: Pustaka Ceria Yayasan PENA.
- Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied Logistic Regression*. New Jersey: John Wiley and Sons
- Johnson, M. S. (2009). *Introduction to Bayesian Statistics with WinBUGS*. New York: Columbia University.
- Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. New Jersey: Wiley-Blackwell.
- Ntzoufras, I. (2009). *Bayesian Modelling Using WinBUGS*. New Jersey: John Wiley and Sons.
- Nurmalasari, R., Ispriyanti, D., dan Sudarno. (2017). Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia (IPM) Menggunakan Metode Regresi Logistik Ordinal dan Regresi Probit Ordinal. *Jurnal Gaussian*, 6 (1), 111-120.
- Prasetyo, E. (2014). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Ramli., Yuniarti, D. dan Goejantoro, R. (2013). Perbandingan Metode Klasifikasi Regresi Logistik dengan Jaringan Saraf Tiruan. *Jurnal Eksponensial*, 4 (1), 17-24.
- Soejoeti, Z dan Soebanar. (1988). *Inferensi Bayesian*. Jakarta: Karunika Universitas Terbuka.
- Teti, M. S., Yanuar, F., dan Yozza, H. (2014). Analisis Regresi Logistik Dengan Metode Penduga Bayes Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Kejadian Bayi Berat Badan Lahir Rendah. *Jurnal Matematika UNAND*, 4 (1), 53-60.