

Analisis Regresi Logistik Biner Terhadap Data Indeks Kedalaman Kemiskinan Di Indonesia Tahun 2020

Regita Elza Fitri¹, Eri Setiawan^{1,*}, Mustofa Usman¹, dan Dorrah Aziz¹

¹Jurusan Matematika, Fakultas MIPA, Universitas Lampung
Jl. Prof. Dr. Ir. Sumantri Brojonegoro. Kota Bandar Lampung, Lampung.

*Email korespondensi: erstatis@gmail.com

Abstrak

Salah satu upaya guna mewujudkan tujuan pembangunan nasional adalah dengan mengembangkan wilayah untuk mengurangi ketimpangan dan menjamin pemerataan. Namun pengembangan wilayah berimplikasi terhadap terjadinya ketimpangan antarwilayah sehingga menciptakan adanya wilayah maju dan wilayah tertinggal. Ketimpangan antarwilayah masih menjadi tantangan dalam pembangunan nasional di Indonesia. Pada penelitian ini akan membahas penerapan pendugaan parameter menggunakan pendekatan metode Bayesian untuk mengetahui model regresi logistik biner pada data daerah tertinggal di Indonesia Tahun 2020 dan mengetahui faktor-faktor yang mempengaruhi pengklasifikasian daerah tertinggal di Indonesia tahun 2020. Pendugaan parameter pada regresi logistik biner dilakukan dengan pendekatan Bayesian yang menggabungkan informasi awal (prior) dengan informasi dari sampel yang dalam hal ini dinyatakan dengan fungsi likelihood, kemudian hasilnya dinyatakan sebagai distribusi posterior. Berdasarkan analisis diperoleh kesimpulan bahwa variabel indeks pembangunan manusia, persentase penduduk miskin, dan persentase rumah tangga dengan akses air minum layak berpengaruh secara signifikan terhadap pengklasifikasian daerah tertinggal di Indonesia tahun 2020.

Kata kunci: Indeks kedalaman kemiskinan, regresi logistic biner

Abstract

Poverty is a serious problem that occurs in many countries, both developing and developed countries. This issue needs to be addressed by the government, especially in countries with large and dense populations such as Indonesia. Poverty inequality as measured by the poverty depth index shows a number that tends to be stable from year to year. Therefore, it's necessary know the causal factors that affect the depth of poverty in Indonesia. This study discusses the factors that affect the poverty depth index in Indonesia in 2020 using binary logistic regression analysis to determine the best binary logistic regression model and find out the magnitude of the classification accuracy of what factors affect the poverty depth index in 34 provinces in Indonesia in 2020. This problem can be overcome by using binary logistic regression because the response variable only consists of two categories, namely high and low poverty depth. Based on the analysis, it can be concluded that the open unemployment rate variable and the average expenditure per capita for one month for food have a significant effect on the classification of the poverty depth index in Indonesia 2020.

Keywords: the poverty depth index, binary logistic regression

1. Pendahuluan

Regresi logistik biner adalah suatu metode analisis data yang dapat digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat biner (dichotomus) dengan variabel prediktor (x) yang bersifat kategorik atau kontinu [1].

Kemiskinan merupakan salah satu persoalan serius yang dihadapi dan perlu ditanggulangi oleh pemerintah di suatu negara baik negara berkembang maupun negara maju terutama di negara-negara dengan jumlah penduduk besar dan padat seperti Indonesia. Kemiskinan merupakan ketidakmampuan masyarakat memenuhi kebutuhan primer dan kebutuhan sekunder sebagai standar minimum kebutuhan dasar yang harus terpenuhi [2]. Banyaknya jumlah penduduk miskin pada September 2020 sebesar 27,55 juta orang, meningkat 2,76 juta orang terhadap September 2019. Angka ini justru mengalami peningkatan seiring berjalannya waktu dari periode-periode sebelumnya. Selain itu, ketimpangan kemiskinan yang diukur dari indeks kedalaman kemiskinan menunjukkan angka yang cenderung stagnan dari tahun ke tahun. Oleh sebab itu, perlu diketahui faktor apa saja

yang mempengaruhi kedalaman kemiskinan yang ada di Indonesia. Untuk mengatasi masalah tersebut, digunakan regresi logistik biner karena variabel respon hanya terdiri dari dua kategori yaitu tingkat kedalaman kemiskinan tinggi dan rendah pada provinsi di Indonesia.

Penelitian menggunakan analisis regresi logistik biner pernah dilakukan oleh Ni Putu dan Maulidia pada tahun 2020 dalam penentuan ketepatan klasifikasi tingkat kedalaman kemiskinan provinsi-provinsi di Indonesia dengan variabel terikat pada penelitian ini adalah tingkat kedalaman kemiskinan provinsi-provinsi di Indonesia tahun 2019. Sedangkan variabel bebasnya adalah pengeluaran per kapita yang disesuaikan, rata-rata lama sekolah, serta harapan lama sekolah provinsi-provinsi di Indonesia tahun 2019 [4]. Oleh karena itu, pada penelitian ini akan dibahas mengenai faktor yang mempengaruhi indeks kedalaman kemiskinan di Indonesia tahun 2020 menggunakan analisis regresi logistik biner.

2. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) mengenai Indeks Kedalaman Kemiskinan (IKK) pada Provinsi di Indonesia tahun 2020. Indeks Kedalaman Kemiskinan di Indonesia sebesar 1,75%. IKK berperan sebagai variabel tak bebas, dimana IKK rendah dilambangkan dengan 0 dan IKK tinggi dilambangkan dengan 1. Indeks kedalaman kemiskinan di Indonesia $>$ Indeks kedalaman kemiskinan di Provinsi dikategorikan 1 dan Indeks kedalaman kemiskinan di Indonesia $<$ Indeks kedalaman kemiskinan di Provinsi dikategorikan 0. Sementara untuk variabel-variabel bebas yang memiliki pengaruh terhadap indeks kedalaman kemiskinan adalah tingkat pengangguran terbuka (X_1), angka melek huruf penduduk umur ≥ 15 tahun (X_2), rata-rata lama sekolah penduduk umur ≥ 15 tahun (X_3), dan rata-rata pengeluaran per kapita sebulan untuk makanan (X_4). Penelitian ini dilakukan dengan bantuan *software* SAS 9.4.

Pada penelitian ini akan dibahas mengenai faktor yang mempengaruhi indeks kedalaman kemiskinan di Indonesia tahun 2020 menggunakan analisis regresi logistik biner. Regresi logistik biner adalah suatu metode analisis data yang dapat digunakan untuk mencari hubungan antara variabel respon (Y) yang bersifat biner (dichotomus) dengan variabel prediktor (X) yang bersifat kategorik atau kontinu [6]. Hasil dari variabel respon y memiliki dua kriteria, yaitu “sukses” dan “gagal” yang dinotasikan dengan $Y = 1$ mewakili kemungkinan sukses dengan probabilitas $\pi(X)$ dan $Y = 0$ mewakili kemungkinan gagal dengan probabilitas $1 - \pi(X)$, dimana variabel respon Y mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Model regresi logistik multivariat dengan p variabel prediktor adalah sebagai berikut :

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (1)$$

Estimasi parameter model regresi logistik biner dilakukan dengan maximum likelihood estimation. Metode tersebut mengestimasi parameter β dengan cara memaksimumkan fungsi likelihood dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu [7]. Pengujian estimasi parameter merupakan pengujian yang digunakan untuk menguji signifikansi koefisien β dari model. Pengujian ini dapat menggunakan uji secara simultan maupun parsial. Uji kesesuaian model adalah uji untuk mengetahui apakah model yang sudah dibentuk sudah sesuai atau belum, dengan kata lain tidak terdapat perbedaan yang berarti antara hasil pengamatan dengan hasil prediksi model [5]. Estimasi koefisien dari variabel prediktor menyatakan slope atau nilai perubahan variabel respon untuk setiap perubahan satu unit variabel prediktor. Interpretasi dari $\hat{OR}$ merupakan resiko kecenderungan terjadinya peristiwa $y = 1$ adalah sebesar $\exp(\beta_1)$ kali resiko atau kecenderungan terjadinya peristiwa $y = 1$ pada kategori $x = 0$ [3]. Ketepatan klasifikasi model digunakan untuk mengetahui apakah data diklasifikasikan dengan benar atau tidak [1].

Adapun langkah-langkah yang dilakukan pada penelitian adalah sebagai berikut:

- Melakukan pengumpulan data
- Melakukan pengolahan dan penyajian data dengan statistika deskriptif.
- Melakukan identifikasi adanya multikolinearitas dengan nilai Variance Inflation Factor (VIF).
- Menguji estimasi parameter secara simultan dengan menggunakan uji *Likelihood Ratio*.
- Menguji estimasi parameter secara parsial dengan menggunakan uji *Wald*.
- Melakukan Uji kesesuaian model menggunakan *Hosmer-Lemeshow Goodness of Fit* untuk mengetahui apakah model yang sudah dibentuk sudah sesuai atau belum.
- Menginterpretasikan model akhir menggunakan odds ratio
- Menghitung ketepatan klasifikasi dan sensitivitas pada model yang dihasilkan.

3. Hasil dan Pembahasan

3.1 Analisis Deskriptif

Analisis statistika deskriptif untuk melihat gambaran secara umum mengenai data yang digunakan dalam penelitian. Dalam penelitian ini variabel tak bebas yang digunakan adalah Indeks Kedalaman Kemiskinan 34 Provinsi di Indonesia pada tahun 2020. Statistika deskriptif untuk data kontinu menggunakan tabel deskriptif yang meliputi jumlah keseluruhan data, rata-rata, standar deviasi, nilai minimum dan nilai maksimum.

Table 1. Statistika Deskriptif

Kelompok	Variable	N	Mean	Std Dev	Minimum	Maximum
Rendah (Y=0)	X_1	19	6,88	2,12	4,58	10,95
	X_2	19	99,89	0,09	99,6	99,97
	X_3	19	9,26	0,74	7,90	11,17
	X_4	19	669252,79	112488,18	496173	944687
Tinggi (Y=1)	X_1	15	4,96	1,23	3,32	7,57
	X_2	15	99,18	2,34	90,78	99,98
	X_3	15	8,85	0,90	6,96	10,2
	X_4	15	548440,73	78074,09	442700	723821

Variabel independen dalam penelitian ini berjumlah 4 variabel, yaitu variabel tingkat pengangguran terbuka (X_1), angka melek huruf penduduk umur ≥ 15 tahun (X_2), rata-rata lama sekolah penduduk umur ≥ 15 (X_3), dan rata-rata pengeluaran per kapita sebulan untuk makanan (X_4).

3.2 Menguji Asumsi Multikolinearitas

Pengujian multikolinearitas dilakukan untuk mengetahui korelasi antar variable predictor dengan cara melihat nilai VIF (Variance Inflation Factor) seperti yang disajikan pada Tabel 2.

Tabel 2. Uji Multikolinearitas

Variabel	Variance Inflation (VIF)
Tingkat pengangguran terbuka (X_1)	1.79863
Angka melek huruf penduduk umur ≥ 15 tahun (X_2)	1.93811
Rata-rata lama sekolah penduduk umur ≥ 15 tahun (X_3)	2.36139
Rata-rata pengeluaran per kapita sebulan untuk makanan (X_4)	1.29401

Berdasarkan Tabel 2 pengujian multikolinearitas di atas, dapat disimpulkan bahwa tidak terdapat multikolinearitas antar variabel prediktor. Hal ini berarti tidak ada korelasi antar variabel prediktor dalam model. Sehingga asumsi multikolinearitas terpenuhi dan dapat dilanjutkan ketahap pembentukan model.

3.3 Pemodelan Regresi Logistik Biner

3.3.1 Model Awal Regresi Logistik Biner

Model regresi ini dikatakan model awal karena perlu dilakukan pengujian terlebih dahulu baik secara simultan maupun secara parsial. Berikut adalah parameter logit yang diperoleh:

Tabel 3. Estimasi Parameter Fungsi Logit Awal

Variabel	Estimate
Y	527,7
X_1	-0,0435
X_2	-5,2345
X_3	0.1427
X_4	-0,00002

Model Awal regresi logistik biner yang terbentuk adalah sebagai berikut:

$$\pi(x) = \frac{\exp(527,7 - 0,0435X_1 - 5,2345X_2 + 0,1427X_3 - 0,00002X_4)}{1 + \exp(527,7 - 0,0435X_1 - 5,2345X_2 + 0,1427X_3 - 0,00002X_4)}$$

Dan jika model regresi logistik biner sementara ditulis dalam bentuk logit menjadi:

$$g(x) = 527,7 - 0,0435X_1 - 5,2345X_2 + 0,1427X_3 - 0,00002X_4$$

Kemudian akan dilakukan pengujian terhadap model secara keseluruhan maupun secara parsial

a. Uji Simultan

Uji signifikansi parameter secara simultan menggunakan uji G (*Likelihood Ratio*). Adapun hasil uji estimasi parameter secara simultan model regresi logistik biner disajikan pada Tabel 4 :

Tabel 4. Hasil Uji Simultan

Test	Chi-Square	DF	Pr > Chi-Sq
Likelihood Ratio	23,8095	4	< 0.0001

Berdasarkan Tabel 4, keputusan adalah tolak H_0 karena $p - value < \alpha$ dengan tingkat kepercayaan 95% sehingga dapat disimpulkan bahwa terdapat minimal satu variabel prediktor yang berpengaruh terhadap indeks kedalaman kemiskinan.

b. Uji Wald

Menguji signifikansi parameter secara parsial digunakan untuk mengetahui apakah masing-masing variabel prediktor berpengaruh signifikan terhadap pembentukan model atau tidak. Uji signifikansi parameter secara parsial menggunakan uji Wald.

Tabel 5. Hasil Uji Parsial

Parameter	DF	Estimate	Standard Error	Wald Chi - Square	Pr > ChiSq	Keputusan
Y	1	8.0586	5.0647	2.5317	0.1116	Tidak tolak H_0
X_1	1	-0.0435	0.0224	3.7720	0.0491	Tolak H_0
X_2	1	-5.2345	5.1009	1.0531	0.3048	Tidak tolak H_0
X_3	1	0.1427	0.0882	2.6181	0.1057	Tidak tolak H_0
X_4	1	-0.00002	0.000010	5.9719	0.0145	Tolak H_0

Dengan menggunakan taraf signifikansi $\alpha = 0,05$ dari tabel χ^2 maka diperoleh nilai $\chi^2(0,05; 1)$ yaitu sebesar 3,84146 dengan kriteria pengambilan keputusan yaitu tolak H_0 jika nilai $W^2 > \chi^2(0,05; 1)$ atau P-value $< \alpha$. Berdasarkan tabel di atas maka terdapat dua variabel yaitu variabel X_1 dan X_4 yang signifikan terhadap model. Sehingga dengan hasil keputusan diatas dapat disimpulkan bahwa variabel X_1 (Tingkat Pengangguran Terbuka) dan X_4 (Rata-Rata Pengeluaran Per Kapita Sebulan Untuk Makanan) berpengaruh signifikan terhadap indeks kedalaman kemiskinan.

3.3.2 Model Akhir Regresi Logistik Biner

Model regresi logistik biner berikut merupakan model setelah dilakukan eliminasi variabel-variabel bebas yang tidak memberikan pengaruh signifikan terhadap variabel bebas. Berikut adalah *output* fungsi logit yang diperoleh:

Tabel 6. Estimasi Parameter Model Terbaik

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi - Square	Pr > ChiSq
Intercept	1	11.4455	4.2315	7.3163	0.0438
X_1	1	-0.0438	0.0217	4.0926	0.0431
X_4	1	-0.00001	5.696E-6	4.6570	0.0309

Berdasarkan Tabel 6, maka model regresi logistik terbaik yang terbentuk adalah sebagai berikut:

$$\pi(x) = \frac{\exp(11.4455 - 0,0438X_1 - 0,00001X_4)}{1 + \exp(11.4455 - 0,0438X_1 - 0,00001X_4)}$$

Dan jika model regresi logistik biner sementara ditulis dalam bentuk logit menjadi:

$$g(x) = 11.4455 - 0.0438X_1 - 0.00001X_4$$

Kemudian akan dilakukan pengujian terhadap model secara keseluruhan maupun secara parsial

a. Uji Simultan

Uji signifikansi parameter secara simultan menggunakan uji G (*Likelihood Ratio*). Adapun hasil uji estimasi parameter secara simultan model regresi logistik biner terbaik disajikan pada Tabel 7 :

Tabel 7. Hasil Uji Simultan Terbaik

Test	Chi-Square	DF	Pr > Chi-Sq
Likelihood Ratio	16,2694	2	0.0003

Berdasarkan Tabel 7, kriteria pengujian dengan pengambilan taraf signifikansi 5% dan $df = 2$ dari table distribusi Chi-Square diperoleh $\chi^2_{(0,05;2)} = 5,9915$. Karena nilai statistic $G(16,294) > \chi^2_{(0,05;2)}$ maka keputusan tolak H_0 yang artinya dengan tingkat kepercayaan 95% dapat disimpulkan bahwa terdapat minimal satu variabel prediktor yang berpengaruh terhadap indeks kedalaman kemiskinan.

b. Uji Wald

Menguji signifikansi parameter secara parsial digunakan untuk mengetahui apakah masing-masing variabel prediktor berpengaruh signifikan terhadap pembentukan model tau tidak. Uji signifikansi parameter secara parsial menggunakan uji Wald.

Tabel 8. Hasil Uji Parsial Terbaik

Parameter	DF	Estimate	Standard Error	Wald Chi - Square	Pr > ChiSq	Keputusan
Y	1	1.4455	0.5269	4.3163	0.0438	Tolak H_0
X_1	1	-0.0438	0.0217	4.0926	0.0431	Tolak H_0
X_4	1	-0.00001	5.696E-6	4.6570	0.0309	Tolak H_0

Dengan menggunakan taraf signifikansi $\alpha = 0,05$ dari tabel χ^2 maka diperoleh nilai $\chi^2(0,05; 1)$ yaitu sebesar 3,84146 dengan kriteria pengambilan keputusan yaitu tolak H_0 jika nilai $W^2 > \chi^2(0,05; 1)$ atau P-value $< \alpha$. Berdasarkan Tabel 8 maka semua parameter X_1 dan X_4 signifikan terhadap model. Sehingga dapat disimpulkan bahwa variabel X_1 (Tingkat Pengangguran Terbuka) dan X_4 (Rata-Rata Pengeluaran Per Kapita Sebulan Untuk Makanan) berpengaruh signifikan terhadap indeks kedalaman kemiskinan.

3.3.3 Interpretasi Model

Interpretasi Model Regresi Logistik Biner menggunakan nilai *odds ratio* yang bertujuan untuk mengetahui seberapa besar pengaruh (kecenderungan) masing-masing variabel prediktor terhadap indeks kedalaman kemiskinan 34 Provinsi di Indonesia. Hasil nilai *odds ratio* dapat dilihat pada tabel sebagai berikut:

Tabel 9. Nilai Odds Ratio

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X_1	0,957	0,917	0,999
X_4	1,000	1,000	1,000

Untuk variabel X_1 sebesar 0,957 dan variabel X_4 sebesar 1,000 menunjukkan bahwa apabila nilai persentase X_1 dan X_4 bertambah sebesar satu satuan, maka kecenderungan Indeks Kedalaman Kemiskinan menjadi 0,957 kali untuk X_1 dan 1,000 kali untuk X_4 dengan asumsi semua variabel lain konstan. Dengan kata lain, semakin tinggi X_1 dan X_4 maka akan semakin tinggi indeks kedalaman kemiskinan di 34 Provinsi di Indonesia.

3.3.4 Ketepatan Klasifikasi

Ketepatan klasifikasi model digunakan untuk mengetahui besarnya klasifikasi yang dihasilkan, dengan cara melihat apakah data diklasifikasikan dengan benar atau tidak. Hasil perhitungan ketepatan klasifikasi dapat dilihat sebagai berikut.

Tabel 10. Ketepatan Klasifikasi Model

Aktual		Prediksi		Total
		Y		
		0	1	
Y	0	11	4	15
	1	4	15	19
Total		15	19	34

Berdasarkan Tabel 10, proporsi salah klasifikasi dapat diketahui menggunakan nilai APER yang diberikan oleh

$$APER = \frac{n_{12} + n_{21}}{n} = \frac{4 + 4}{34} = 0,235$$

Dengan demikian proporsi salah klasifikasi adalah sebesar 23,5%, sehingga ketepatan klasifikasi yang dihasilkan adalah 76,5%.

4. Kesimpulan

Berdasarkan hasil dan pembahasan, diperoleh kesimpulan bahwa faktor yang mempengaruhi indeks kedalaman kemiskinan 34 Provinsi di Indonesia tahun 2020 adalah variabel tingkat pengangguran terbuka. Berdasarkan analisis regresi logistik biner yang diperoleh model regresi yaitu:

$$\pi(x) = \frac{\exp(11.4455 - 0,0438X_1 - 0,00001X_4)}{1 + \exp(11.4455 - 0,0438X_1 - 0,00001X_4)}$$

Berdasarkan hasil pemodelan regresi logistik biner, variabel prediktor yang memberikan pengaruh terhadap indeks kedalaman kemiskinan di Indonesia tahun 2020 adalah tingkat pengangguran terbuka (X_1) dan rata-rata pengeluaran per kapita sebulan untuk makanan (X_4). Lebih jauh analisis regresi logistik biner berhasil digunakan untuk memodelkan indeks kedalaman kemiskinan 34 Provinsi di Indonesia tahun 2020 dimana dengan menggunakan data sekunder diperoleh bahwa variabel persentase tingkat pengangguran terbuka dan rata-rata pengeluaran per kapita untuk makanan terbukti secara signifikan memberikan pengaruh dengan tingkat ketepatan klasifikasi model sebesar 76,5 % artinya bahwa terdapat 23,5% dari kategori variabel Y antara nilai aktual dan prediksi yang kurang tepat.

Daftar Pustaka:

- [1] Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. 2nd Edition. John Wiley & Sons, New Jersey.
- [2] BPS. 2021. *Indeks Kedalaman Kemiskinan 2020*. Badan Pusat Statistik, Jakarta.
- [3] Djalal, N & Usman, H. 2002. *Penggunaan Teknik Ekonometri*. PT. Raja Grafindo Persada, Jakarta.
- [4] Hendayanti, N.P.N. & Nurhidayati, M. 2020. Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia. *Jurnal Sains Dan Teknologi*. 12(2): 63-70.
- [5] Hosmer, D.W., & Lemeshow. 2000. *Applied Logistic Regression*. 2th Edition. John Wiley & Sons, USA.
- [6] Hosmer, D.W., & Lemeshow, S. & Sturdivant, X. R. 2013. *Applied Logistic Regression*. 3rd Edition. John Wiley & Sons, New York
- [7] Kleinbaum, D.G. & Klein, M. 2010. *Logistic Regression A Self-Learning Text*. 3rd Edition . Springer, New York.