

**PEMODELAN REGRESI LOGISTIK DAN REGRESI PROBIT
PADA INDEKS PEMBANGUNAN MANUSIA
KABUPATEN/KOTA DI PULAU KALIMANTAN**

**M. Fathurahman^{1*}, Meiliyani Siringoringo¹, Andi M. Ade Satriya¹,
Nariza Wanti Wulan Sari¹**

¹Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas
Mulawarman, Indonesia

Corresponding author: fathur@fmipa.unmul.ac.id

Abstrak. Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk). IPM dapat menentukan peringkat atau level pembangunan suatu wilayah/negara. Selain itu, IPM dapat juga digunakan untuk mengukur kinerja pemerintah baik di tingkat pusat maupun daerah. Berdasarkan publikasi peringkat IPM oleh Badan Pusat Statistik tahun 2017 menunjukkan bahwa peringkat IPM untuk 56 kabupaten/kota di Pulau Kalimantan sangat bervariasi. Penelitian ini bertujuan melakukan pemodelan IPM kabupaten/kota di Pulau Kalimantan tahun 2017 dengan regresi logistic dan regresi probit, dan mendapatkan faktor-faktor yang signifikan berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan tahun 2017. Hasil penelitian menunjukkan bahwa model regresi terbaik untuk memodelkan IPM kabupaten/kota di Pulau Kalimantan tahun 2017 adalah model regresi logistik. Faktor-faktor yang signifikan adalah jumlah tenaga kesehatan dan tingkat pengangguran terbuka.

Kata Kunci: IPM, regresi logistik, regresi probit, data kategorik.

1 PENDAHULUAN

Regresi logistik dan regresi probit merupakan model regresi yang sering digunakan untuk pemodelan dan analisis data kategorik. Kedua model tersebut termasuk dalam model linier tergeneralisasi (*generalized linear models*) dan dibedakan oleh fungsi penghubung (*link function*) logit dan probit [1]. Model regresi logistik dan regresi probit merupakan pendekatan alternatif untuk memodelkan hubungan antara variabel respon kategorik dan variabel bebas, dimana variabel respon berdistribusi Bernoulli atau multinomial [2, 3, 4].

Beberapa penelitian yang mengkaji pemodelan regresi logistik dan regresi probit diantaranya adalah estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE), pendekatan numerik untuk estimasi parameter, dan diagnosis terhadap model [5]. Penaksiran parameter model dengan metode MLE, pengujian hipotesis parameter model dengan metode LRT dan uji Wald, dan pemilihan model terbaik dengan kriteria *Akaike's Information Criterion* (AIC) dan *Cp Mallows* [6]. Pengujian hipotesis parameter *robust* [7]. Estimasi parameter model dengan metode *Least Absolute Deviation* (LAD) [8]. Matriks varians kovarians parameter model [9].

Penelitian yang mengkaji regresi probit diantaranya adalah pemilihan model regresi probit terbaik [10]. Variabel pengontrol hubungan antara variabel bebas dan variabel respon pada regresi probit [11]. Pemodelan regresi probit dan penerapannya untuk data mining [12]. pemodelan dan penerapan regresi probit pada ketahanan pangan [13]. Pemodelan dan prediksi kebangkrutan bank di Slovakia dengan regresi probit [14] dan [15]. Pengembangan paket *software* R yang diberi nama *SimCorMultRes* untuk simulasi dan analisis regresi probit [16]. Pemodelan faktor-faktor yang berpengaruh terhadap penderita diabetes dan HIV/AIDS dengan regresi probit semiparametrik [17].

Pemodelan regresi logistik dan regresi probit dalam penelitian ini diimplementasikan pada pemodelan Indeks Pembangunan Manusia (IPM) kabupaten/kota di pulau Kalimantan tahun 2017. IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. IPM dapat menentukan peringkat atau level pembangunan suatu wilayah [18]. Penelitian ini bertujuan mendapatkan model regresi terbaik pada pemodelan IPM kabupaten/kota di Pulau Kalimantan tahun 2017 dan mendapatkan faktor-faktor yang signifikan berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan tahun 2017.

2 TINJAUAN PUSTAKA

2.1 Regresi Logistik

Regresi logistik merupakan metode regresi yang dapat menjelaskan hubungan antara variabel respon kategorik dan satu atau lebih variabel bebas. Hasil pengamatan variabel random respon (Y) mempunyai dua kategori yaitu 0 dan 1, sehingga mengikuti distribusi Bernoulli dengan fungsi probabilitas [3].

$$P(Y = y) = \pi^y (1 - \pi)^{1-y}, y = 0,1 \quad (1)$$

dengan $P(Y = 1) = \pi$ menyatakan probabilitas untuk variabel respon dengan kategori $y = 1$ dan $P(Y = 0) = 1 - \pi$ menyatakan probabilitas untuk variabel respon dengan kategori $y = 0$.

Model regresi logistik dengan k variabel bebas dapat ditulis sebagai berikut [3].

$$\pi(\mathbf{x}) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})} \quad (2)$$

dengan $\boldsymbol{\beta}^T = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]$ adalah vektor parameter dan $\mathbf{x} = [x_0 \ x_1 \ x_2 \ \dots \ x_k]^T$ adalah vektor variabel bebas dengan $x_0 = 1$. Untuk mendapatkan model regresi logistik yang linier dalam parameter dilakukan transformasi logit terhadap Persamaan (2) dan diperoleh model regresi logistik seperti Persamaan (3).

$$g(\mathbf{x}) = \ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \boldsymbol{\beta}^T \mathbf{x} \quad (3)$$

Variabel bebas pada Persamaan (3) tidak saling berkorelasi atau bebas multikolinieritas. Untuk mengetahui adanya multikolinieritas digunakan nilai *Variance Inflation Factor* (VIF). Jika nilai VIF antar variabel bebas kurang dari 10, maka tidak ada multikolinieritas [19].

Estimasi parameter dalam regresi logistik dilakukan dengan metode *maximum likelihood*. Metode tersebut mengestimasi parameter dengan cara memaksimalkan fungsi *likelihood* dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Berdasarkan Persamaan (1) dan (2) dibentuk fungsi *likelihood* seperti Persamaan (4).

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

Dalam bentuk $\log l(\boldsymbol{\beta})$ menjadi dan dinyatakan dalam bentuk $L(\boldsymbol{\beta})$ dengan persamaan sebagai berikut.

$$L(\boldsymbol{\beta}) = \sum_{j=0}^p \left(\sum_{i=1}^n y_i x_{ij} \right) \beta_j - \sum_{i=1}^n \log \left(1 + \exp \left(\sum_{j=0}^p \beta_j x_{ij} \right) \right) \quad (5)$$

Nilai $\boldsymbol{\beta}$ maksimum diperoleh melalui turunan pertama $L(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ dan hasilnya sama dengan nol. Untuk mendapatkan taksiran $\boldsymbol{\beta}$ dari turunan pertama fungsi $L(\boldsymbol{\beta})$ yang tidak linier digunakan metode *Fisher Scoring*.

2.2 Regresi Probit

Regresi probit merupakan salah satu model regresi yang dapat menjelaskan hubungan antara variabel dependen kualitatif (kategorik) berdistribusi normal dan Bernoulli dan variabel independen kualitatif, kuantitatif, atau gabungan kualitatif dan kuantitatif.

Pemodelan regresi probit menggunakan fungsi penghubung (*link function*) distribusi normal dengan fungsi distribusi probabilitas dan fungsi distribusi kumulatif berturut-turut seperti Persamaan (6) dan (7).

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad (6)$$

untuk $-\infty < y < \infty, \sigma^2 > 0$,

$$\begin{aligned} P(Y \leq y) &= \Phi(y) \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right] du. \end{aligned} \quad (7)$$

Jika pada Persamaan (1), $\mu = 0$ dan $\sigma^2 = 1$, maka diperoleh distribusi normal standar dengan fungsi distribusi probabilitas dan fungsi distribusi kumulatif berturut-turut seperti persamaan (8) dan (9)

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right), -\infty < t < \infty \quad (8)$$

$$P(Y \leq y) = \Phi(t)$$

$$= \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du. \quad (9)$$

Model regresi probit dapat ditulis seperti persamaan sebagai berikut [20]:

$$\Phi^{-1}(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \text{ untuk } i = 1, 2, \dots, n \quad (10)$$

dengan $\mathbf{x}_i^T = [1 \ x_1 \ x_2 \ \dots \ x_k]$ adalah vektor variabel independen pada pengamatan ke- i , $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]^T$ adalah vektor parameter, $\Phi^{-1}(\pi_i)$ adalah probit link function dan $\pi_i = P(Y_i = y)$ adalah probabilitas variabel respon pada pengamatan ke- i mempunyai kategori y , untuk $y = 0, 1$ yang dinyatakan sebagai berikut:

$$P(Y_i = 1) = \Phi(\mathbf{x}_i^T \boldsymbol{\beta}), i = 1, 2, \dots, n \quad (11)$$

$$P(Y_i = 0) = 1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (12)$$

Taksiran parameter model regresi probit pada Persamaan (10) dapat diperoleh dengan metode *Maximum Likelihood Estimation* (MLE). Hasil turunan pertama fungsi *log-likelihood* terhadap parameter yang ditaksir pada model regresi probit menghasilkan fungsi yang tidak *closed form* sehingga diperlukan pendekatan numerik untuk mendapatkan estimasi dari parameter. Salah satu pendekatan numerik yang dapat digunakan adalah metode *Fisher Scoring* [21].

Setelah dilakukan estimasi parameter model regresi logistik dan regresi probit dilanjutkan dengan pengujian hipotesis parameter. Pengujian ini meliputi uji

simultan dan uji parsial. Hipotesis yang digunakan untuk uji simultan diformulasikan sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{Minimal terdapat satu } \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, k$$

Statistik uji yang digunakan pada pengujian simultan adalah statistik *likelihood ratio* yang dapat diperoleh dengan metode *Likelihood Ratio Test* (LRT) [20] sebagai berikut:

$$G^2 = -2[\ell(\hat{\omega}) - \ell(\hat{\Omega})] \quad (13)$$

Dimana $\ell(\hat{\omega}) = \ln L(\hat{\omega})$ adalah nilai maksimum fungsi *log-likelihood* dibawah H_0 dan $\ell(\hat{\Omega}) = \ln L(\hat{\Omega})$ adalah nilai maksimum *log-likelihood* di bawah populasi. Statistik uji pada Persamaan (11) secara asimtotik berdistribusi *chi-square* dengan derajat bebasnya adalah banyaknya parameter model di bawah populasi dikurangi dengan banyaknya parameter model di bawah H_0 , yaitu $v = (k + 1) - 1 = k$. Daerah kritis untuk uji simultan adalah H_0 ditolak jika $G^2 > \chi^2_{(\alpha, v)}$ atau nilai probabilitas $(P) < \alpha$, α adalah tingkat signifikansi dan $\chi^2_{(\alpha, v)}$ dapat diperoleh dari tabel distribusi *chi-square*.

Setelah dilakukan uji simultan, pengujian hipotesis parameter selanjutnya adalah uji parsial. Hipotesis untuk uji parsial diformulasikan sebagai berikut:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, k.$$

Statistik uji yang digunakan untuk uji ini adalah statistik Wald [20]:

$$Z_j = \frac{\hat{\beta}_j}{\hat{SE}(\hat{\beta}_j)} \quad (14)$$

dimana $\hat{\beta}_j$ adalah estimator parameter model dan $\hat{SE}(\hat{\beta}_j)$ adalah *standard error* dari estimator parameter model. Statistik Wald pada Persamaan (14) secara asimtotik berdistribusi normal standar. Daerah kritis untuk uji parsial adalah H_0 ditolak jika $|Z_j| > Z_{\alpha/2}$ atau nilai probabilitas $(P) < \alpha$, α adalah tingkat signifikansi dan $Z_{\alpha/2}$ dapat diperoleh dari tabel distribusi normal standar. Untuk mendapatkan model regresi terbaik digunakan kriteria nilai AIC dan devians residual.

2.3 Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) atau *Human Development Index* (HDI) menjadi ukuran standar keberhasilan pembangunan manusia sebuah negara. Konsep pembangunan manusia pada dasarnya memiliki makna yang sangat luas. Konsep ini mencakup semua dimensi dasar yang dimiliki oleh manusia. Ide dasar dari konsep pembangunan manusia cukup sederhana, yaitu menciptakan pertumbuhan positif dalam bidang ekonomi, sosial, politik, budaya, dan lingkungan, serta perubahan dalam kesejahteraan manusia. Oleh karena itu, manusia harus diposisikan sebagai kekayaan bangsa yang sesungguhnya.

Konsep pembangunan manusia diukur dengan menggunakan pendekatan tiga dimensi dasar manusia, yaitu umur panjang dan sehat, pengetahuan dan standar hidup yang layak. Dimensi umur panjang dan sehat diwakili oleh indikator harapan hidup saat lahir. Dimensi pengetahuan diwakili oleh indikator harapan lama sekolah dan rata-rata lama sekolah. Sementara itu, dimensi standar hidup layak diwakili oleh pengeluaran per kapita. Ketiga dimensi ini terangkum dalam suatu indeks komposit yang membentuk IPM. [22]

3 METODE PENELITIAN

Metode penelitian akan menguraikan tentang data metode yang digunakan.

3.1 Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Data yang diperlukan meliputi IPM kabupaten/kota di pulau Kalimantan tahun 2017 sebagai variabel respon (Y) dengan kategori $0 = \text{IPM kabupaten/kota} < \text{IPM Indonesia sebesar } 70,81$, $1 = \text{IPM kabupaten/kota} \geq \text{IPM Indonesia sebesar } 70,81$. Selanjutnya lima variabel bebas yang diduga berpengaruh terhadap IPM, yaitu persentase penduduk miskin (X_1), jumlah tenaga kesehatan (X_2), jumlah sarana kesehatan (X_3), persentase pertumbuhan ekonomi (X_4), dan tingkat pengangguran terbuka (TPT) (X_5).

Unit pengamatan yang digunakan dalam penelitian ini adalah kabupaten/kota di pulau Kalimantan pada tahun 2017 yang terdiri dari 47 kabupaten dan 9 kota. Sehingga jumlah keseluruhan dari unit pengamatan adalah sebanyak 56 kabupaten/kota.

3.2 Metode

Metode analisis data untuk pemodelan IPM kabupaten/kota di pulau Kalimantan tahun 2017 dengan regresi logistik dan regresi probit adalah sebagai berikut :

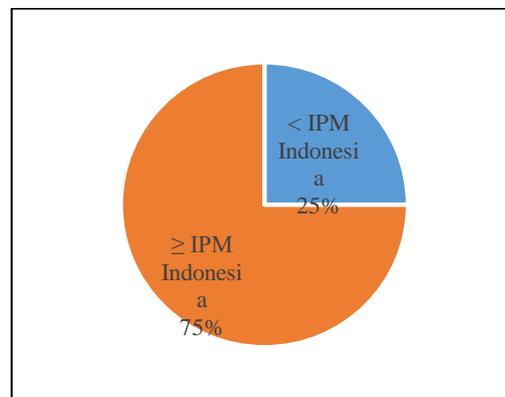
1. Melakukan analisis statistika deskriptif terhadap variabel respon dan variabel bebas.
2. Melakukan pengujian multikolinieritas pada variabel bebas.
3. Melakukan estimasi parameter model regresi logistik dan regresi probit dengan metode *Fisher Scoring*.

4. Melakukan pengujian hipotesis parameter secara serentak dan parsial berturut-turut dengan metode LRT dan uji Wald.
5. Mendapatkan variabel independen yang signifikan berpengaruh terhadap variabel dependen.
6. Melakukan pemilihan model terbaik.

Pengolahan dan analisis data untuk estimasi dan pengujian hipotesis parameter menggunakan *software* R untuk *windows* versi 3.6.1.

4 HASIL DAN PEMBAHASAN

Pemodelan IPM Kabupaten/Kota di Pulau Kalimantan tahun 2017 diawali dengan melihat karakteristik data pada variabel respon dan variabel bebas yang digunakan. Karakteristik data disajikan dalam diagram dan tabel deskripsi sebagai berikut.



Gambar 1: Deskripsi Kategori Variabel Respon

Gambar 1 menunjukkan bahwa sebanyak 75% (42) kabupaten/kota di Pulau Kalimantan memiliki nilai IPM lebih kecil dari IPM Indonesia pada tahun 2017. Sedangkan 25% (14) kabupaten/kota yang lain memiliki nilai IPM lebih besar dari IPM Indonesia.

Tabel 1 : Deskripsi Variabel Bebas

Variabel Bebas	Min	Maks	Mean	Simpangan Baku
X1	2,820	12,830	6,439	2,352
X2	58	1500	441,200	304,675
X3	13	212	74,210	45,814
X4	0,680	7,990	5,197	1,595
X5	0,640	12,440	4,740	2,305

Variabel X_1 merupakan persentase penduduk miskin yang mana Tabel 1 menunjukkan persentase penduduk miskin terendah adalah sebesar 2,820 yaitu Kota Balikpapan dan persentase penduduk miskin tertinggi adalah sebesar 12,830 yaitu Kabupaten Melawi dengan *mean* sebesar 6,439 dan simpangan baku sebesar 2,352. Variabel X_2 yang merupakan jumlah tenaga kesehatan memiliki jumlah tenaga kesehatan paling sedikit terdapat di Kabupaten Mahakan Ulu sebesar 58

orang dan jumlah tenaga kesehatan paling banyak terdapat di Kota Samarinda sebesar 1500 orang dengan *mean* sebesar 441 orang dan simpangan baku sebesar 305 orang. Variabel X_3 yang merupakan jumlah sarana kesehatan yang mana jumlah sarana kesehatan paling sedikit terdapat di Kota Bontang dan Kabupaten Bulungan sebesar 13 orang dan jumlah sarana kesehatan paling banyak terdapat di Kabupaten Kutai Kertanegara sebesar 212 orang dengan *mean* sebesar 74 orang dan simpangan baku sebesar 46 orang. Variabel X_4 merupakan persentase pertumbuhan ekonomi menunjukkan persentase pertumbuhan ekonomi terendah adalah sebesar 0,680 yaitu Kota Bontang dan persentase pertumbuhan ekonomi tertinggi adalah sebesar 7,990 yaitu Kabupaten Kotawaringin Timur dengan *mean* sebesar 5,197 dan simpangan baku sebesar 1,595. Variabel X_5 merupakan tingkat pengangguran terbuka menunjukkan tingkat pengangguran terbuka terendah terdapat di Kabupaten Sekadau yaitu sebesar 0,640 persen dan tingkat pengangguran terbuka tertinggi terdapat di Kota Bontang yaitu sebesar 12,440 dengan *mean* sebesar 4,740 persen dan simpangan baku sebesar 2,305 persen.

Pemodelan IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 akan menggunakan beberapa variabel bebas sehingga perlu dilakukan pengecekan multikolinieritas. Pengecekan multikolinieritas dilakukan dengan melihat nilai VIF untuk masing-masing variabel bebasnya.

Tabel 2: Pengecekan Multikolinieritas

Variabel Bebas	VIF
X1	1,271
X2	1,238
X3	1,118
X4	1,189
X5	1,345

Tabel 2 menampilkan nilai VIF untuk masing-masing variabel bebas. Kelima variabel bebas memiliki nilai VIF yang lebih kecil dari 10 sehingga dapat disimpulkan bahwa tidak terjadi multikolinieritas pada variabel bebasnya. Pemodelan IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 dilakukan dengan menggunakan dua model yaitu model regresi logistik dan model regresi probit.

Tabel 3: Nilai Estimasi dan Pengujian Parameter Model Regresi Logistik

Parameter	Estimasi	<i>P-value</i>	<i>Odds</i>
β_0	-7,7143	0,0354*	0,0004
β_1	-0,0358	0,8807	0,9648
β_2	0,0077	0,0150*	1,0078
β_3	0,0100	0,2500	1,0100
β_4	-0,4507	0,1848	0,6372
β_5	0,9570	0,0061*	2,6038
G^2			31,245

Ket: (*) parameter signifikan pada $\alpha = 0,05$

Tabel 3 menampilkan nilai uji G^2 sebesar 31,245 yang jika dibandingkan dengan nilai *chi-square* tabel yaitu sebesar 1,145 maka H_0 gagal ditolak karena nilai G^2 lebih besar daripada nilai *chi-square* tabel. Hal tersebut berarti secara simultan variabel bebas berpengaruh terhadap IPM. Parameter yang signifikan pada pengujian parsial adalah β_0 , β_2 dan β_5 sehingga variabel bebas yang berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 pada model regresi logistik adalah jumlah tenaga kesehatan dan tingkat pengangguran terbuka.

Tabel 4: Nilai Estimasi dan Pengujian Parameter Model Regresi Probit

Parameter	Estimasi	<i>P-value</i>
β_0	-4,2967	0,0297*
β_1	-0,0357	0,7933
β_2	0,0042	0,0090*
β_3	0,0062	0,2072
β_4	-0,2256	0,2218
β_5	0,5259	0,0038*
G^2		31,428

Ket: (*) parameter yang signifikan $\alpha = 0,05$

Hasil estimasi dan nilai pengujian parameter model regresi probit untuk IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 ditampilkan pada Tabel 4. Nilai G^2 sebesar 31,428 lebih kecil daripada nilai *chi-square* tabel sebesar 1,145 sehingga tolak H_0 yang berarti secara simultan variabel bebas berpengaruh terhadap IPM. Pengujian secara parsial, Tabel 4 menunjukkan bahwa β_0 , β_2 dan β_5 merupakan parameter yang signifikan sehingga variabel bebas yang berpengaruh adalah jumlah tenaga kesehatan dan tingkat pengangguran terbuka.

Tabel 5: Perbandingan Model Terbaik

Model	Kriteria	
	AIC	Devians Residual
Regresi Logistik	42,245	31,245
Regresi Probit	42,428	31,428

Model regresi logistik menghasilkan nilai AIC sebesar 42,245 yang mana nilai tersebut lebih kecil daripada nilai AIC yang dihasilkan oleh model regresi probit yaitu sebesar 42,428 dengan selisih sebesar 0,183. Perbandingan nilai devians residual juga memberikan hasil yang sama yaitu nilai devians residual pada model regresi logistik (31,245) lebih kecil daripada nilai devians residual model regresi probit (31,428). Berdasarkan nilai AIC dan devians residual kedua model tersebut, model regresi logistik lebih baik atau cocok digunakan untuk memodelkan IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 dengan variabel bebas yang berpengaruh adalah jumlah tenaga kesehatan dan tingkat pengangguran terbuka.

Persamaan model regresi logistik pada data IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 adalah sebagai berikut.

$$\pi(\mathbf{x}) = \frac{\exp(-7,7143 - 0,0356X_1 + 0,0077X_2 + 0,0010X_3 - 0,4507X_4 + 0,9570X_5)}{1 + \exp(-7,7143 - 0,0356X_1 + 0,0077X_2 + 0,0010X_3 - 0,4507X_4 + 0,9570X_5)} \quad (1)$$

Nilai koefisien jumlah tenaga kesehatan (X_2) sebesar 0,0077 memiliki arti jika terjadi pertambahan jumlah tenaga kesehatan sebesar 1 orang maka IPM kabupaten/kota di Pulau Kalimantan lebih dari atau sama dengan IPM Indonesia meningkat sebesar 1,0078 (Tabel 3). Nilai koefisien tingkat pengangguran terbuka (X_5) sebesar 0,9570 memiliki arti jika terjadi peningkatan pengangguran terbuka sebesar 1 persen maka IPM kabupaten/kota di Pulau Kalimantan lebih dari atau sama dengan IPM Indonesia meningkat sebesar 2,6038 (Tabel 3).

5 KESIMPULAN DAN SARAN

Regresi logistik merupakan model yang lebih baik digunakan untuk memodelkan IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 dibandingkan regresi probit. Berdasarkan uji parsial, variabel yang berpengaruh terhadap IPM kabupaten/kota di Pulau Kalimantan Tahun 2017 adalah jumlah tenaga kesehatan dan tingkat pengangguran terbuka.

Saran yang diajukan dari penelitian ini adalah estimasi parameter untuk model regresi logistik dan regresi probit dapat dilakukan dengan metode lain seperti *Weighted Least Square* dan Bayesian. Atau dengan menggunakan metode pendekatan numerik lain seperti Newton Rapshon dan Quasi Newton.

DAFTAR PUSTAKA

- [1] P. McCullagh, J.A. Nelder, *Generalized linear models, 2nd Edition*. Chapman and Hall, 1989.
- [2] D.J. Finney, *Probit analysis, 3rd Edition*. Cambridge University Press, 1971.
- [3] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, *Applied logistic regression, 3rd Edition*. John Wiley & Sons, 2013.
- [4] A. Agresti, *Categorical data analysis, 3rd Edition*. John Wiley & Sons, 2013.
- [5] D. Pregibon, Logistic regression diagnostics. *The Annals of Statistics*, 9, 705-724, 1981.
- [6] D.W. Hosmer, B. Jovanovic, S. Lemeshow, Best subsets logistic regression. *Biometrics*, 45, 1265-1270, 1989.
- [7] A.M. Bianco, E. Martinez, Robust testing in the logistic regression model. *Computational Statistics and Data Analysis*, 53, 4095-4105, 2009.
- [8] S. Hosseinian, E. Martinez, Robust binary regression. *Journal of Statistical Planning and Inference*, 141, 1497-1509, 2011.

- [9] R. Pingel, Some approximations of the logistic distribution with application to the covariance matrix of logistic regression. *Statistics & Probability Letters*, 85, 63-68, 2014.
- [10] G. Chen, H. Tsurumi, Probit and logit selection. *Communications in Statistics-Theory and Methods*, 40, 159-175, 2011.
- [11] K.B. Karlson, A. Holm, R. Breen, Comparing regression coefficients between same-sample nested models using logit and probit: a new method. *Sociological Methodology*, 42, 286-313, 2012.
- [12] M. Razzaghi, The probit link function in generalized linear models for data mining applications. *Journal of Modern Applied Statistical Methods*, 12, 164-169, 2013.
- [13] A. Abdulai, W. Huffman, The adoption and impact of soil and water conservation technology: an endogenous switching regression application. *Land Economics*, 90, 26-43, 2014.
- [14] T. Kliestik, K. Kocisova, M. Misankova, Logit and probit model used for prediction of financial health of company. *Procedia Economics and Finance*, 23, 850-855, 2015.
- [15] M. Kovacova, T. Kliestik, Logit and probit application for the prediction of bankruptcy in Slovak companies equilibrium. *Quarterly Journal of Economics and Economic Policy*, 12, 775-791, 2017.
- [16] A. Touloumis, Simulating correlated binary and multinomial responses under marginal model specification: the SimCorMultRes package. *The R Journal*, 8, 79-91, 2016.
- [17] H. Liu, J. Qin, Semiparametric probit models with univariate and bivariate current-status data. *Biometrics*, 74, 68-76, 2018.
- [18] Badan Pusat Statistik, *Indeks pembangunan manusia (metode baru)*. Direktorat Analisis dan Pengembangan Statistik, 2015.
- [19] M. H. Kutner, C. J. Nachtsheim dan J. Neter, *Applied Linear Regression Models, Fourth Edition*. McGraw-Hill/Irwin, 2004.
- [20] W. H. Greene, *Econometric analysis, fifth edition*. Prentice Hall, 2003.
- [21] M. Fathurahman, Pemodelan indeks pembangunan kesehatan masyarakat Kabupaten/Kota di Pulau Kalimantan menggunakan pendekatan regresi probit. *Jurnal Varian*, 2 (2), 47-54, 2019.
- [22] Badan Pusat Statistik, *Indeks pembangunan manusia*. Badan Pusat Statistik, 2017.