

## PEMODELAN PERSENTASE PENDUDUK MISKIN DI INDONESIA MENGGUNAKAN REGRESI PROBIT DAN REGRESI LOGISTIK

Grace Wulandari<sup>1\*</sup>, Nur Afifah Febriyanti<sup>1</sup>, Khoiril Anwar<sup>1</sup>,  
Darnah Andi Nohe<sup>1</sup>

<sup>1</sup>Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Mulawarman, Indonesia

*Corresponding author:* [darnah.98@gmail.com](mailto:darnah.98@gmail.com)

**Abstrak.** Regresi probit dan regresi logistik merupakan pendekatan alternatif untuk memodelkan hubungan antara variabel respon kategorik dan variabel bebas, di mana variabel respon berdistribusi Bernoulli atau Multinomial. Model probit dan model logistik adalah model non linier yang menggunakan bilangan biner (variabel boneka) sebagai variabel responnya. Variabel boneka yang dimaksud adalah jenis variabel diskrit yang mempunyai dua nilai yaitu 0 dan 1. Tujuan penelitian ini adalah untuk memperoleh model regresi terbaik pada pemodelan persentase penduduk miskin dan mengetahui faktor-faktor yang memengaruhi persentase penduduk miskin di Indonesia tahun 2020. Hasil penelitian menunjukkan bahwa model terbaik pada pemodelan persentase penduduk miskin di Indonesia tahun 2020 adalah model regresi probit dengan nilai AIC sebesar 25,137 dan ketepatan klasifikasi sebesar 91,18%. Berdasarkan model terbaik diperoleh faktor-faktor yang memengaruhi persentase penduduk miskin di Indonesia tahun 2020 yaitu tingkat pengangguran terbuka, proyek investasi dalam negeri, indeks pembangunan manusia (IPM), angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun, dan tingkat setengah pengangguran.

**Kata Kunci:** *Data kategorik, persentase penduduk miskin, regresi logistik, regresi probit.*

## **1 PENDAHULUAN**

Analisis regresi adalah suatu metode yang digunakan untuk mengetahui apakah terdapat pengaruh variabel bebas terhadap variabel terikat. Variabel yang mempengaruhi sering disebut variabel bebas atau variabel independen, sedangkan variabel yang dipengaruhi sering disebut dengan variabel terikat atau variabel dependen. Analisis regresi dapat dilakukan pada data kuantitatif dan kualitatif [1]. Regresi dibagi dalam dua persamaan matematis yang bentuknya dapat linier atau non-linier. Regresi linier membentuk hubungan antara variabel bebas terhadap variabel terikat yang bentuk fungsinya linier. Regresi linier hanya dapat digunakan pada data skala interval dan rasio. Sedangkan, regresi non linier tidak membentuk hubungan yang linier antara variabel bebas terhadap variabel terikat dan biasanya regresi non linier dapat digunakan pada data skala nominal dan ordinal [2].

Dalam menganalisis data kualitatif skala data nominal maupun ordinal dengan fungsi yang terbentuk bersifat non linier maka dapat digunakan metode regresi probit dan regresi logistik. Regresi probit dan regresi logistik merupakan pendekatan alternatif untuk memodelkan hubungan antara variabel respon kategorik dan variabel bebas, di mana variabel respon berdistribusi Bernoulli atau multinomial [3]. Model probit dan model logit merupakan model yang non linier yang menggunakan bilangan biner (variabel boneka) sebagai variabel responnya. Variabel boneka yang dimaksud adalah jenis variabel diskrit yang mempunyai dua nilai yaitu 0 dan 1 [4].

Pembangunan ekonomi merupakan salah satu tujuan utama bagi negara sedang berkembang, termasuk Indonesia. Pembangunan tidak hanya berkaitan dengan pertumbuhan tetapi juga pada peningkatan kesejahteraan, keamanan serta kualitas sumber daya termasuk sumber daya manusia. Pembangunan ekonomi maupun pembangunan pada bidang-bidang lainnya selalu melibatkan sumber daya manusia sebagai salah satu pelaku pembangunan.

Kemiskinan merupakan salah satu persoalan mendasar yang menjadi pusat perhatian pemerintah di negara manapun. Kemiskinan adalah kondisi di mana seseorang tidak dapat menikmati segala macam pilihan dan kesempatan dalam pemenuhan kebutuhan dasarnya, seperti tidak dapat memenuhi kesehatan, standar hidup layak, kebebasan, harga diri dan rasa dihormati seperti orang lain [5]. Kemiskinan merupakan masalah yang dihadapi oleh seluruh negara, terutama di negara berkembang seperti Indonesia. Hal ini dikarenakan kemiskinan bersifat multidimensional, artinya karena kebutuhan manusia itu bermacam-macam, maka kemiskinan pun memiliki aspek primer yang berupa miskin aset, organisasi sosial politik pengetahuan, dan keterampilan serta aspek sekunder yang berupa miskin akan jaringan sosial, sumber-sumber keuangan dan informasi. Kemiskinan di Indonesia dapat direpresentasikan melalui persentase penduduk miskin. Persentase kemiskinan di Indonesia merupakan persentase agregat dari 34 provinsi yang ada di Indonesia. Persentase penduduk miskin di 34 provinsi di Indonesia masih tidak merata dan sebagian besar masih dalam kategori tinggi.

Terdapat beberapa penelitian terdahulu yang mengkaji regresi probit salah satunya yaitu penelitian mengenai Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Pulau Kalimantan Menggunakan Pendekatan

Regresi Probit yang menghasilkan kesimpulan bahwa faktor-faktor yang berpengaruh signifikan terhadap IPKM kabupaten/kota di Pulau Kalimantan tahun 2013 berdasarkan pemodelan dengan regresi probit adalah IPM dan TPAK [6]. Kemudian, penelitian terdahulu yang mengkaji regresi logistik salah satunya ialah penelitian mengenai Analisis Regresi Logistik untuk Menentukan Faktor-Faktor yang mempengaruhi IPK Mahasiswa FMIPA Universitas Sam Ratulangi Manado diperoleh kesimpulan bahwa variabel program studi dan tempat tinggal berpengaruh pada IPK mahasiswa di FMIPA Unsrat Manado [7].

Oleh karena itu berdasarkan uraian permasalahan di atas maka peneliti tertarik melakukan penelitian untuk mengetahui gambaran umum dari persentase penduduk miskin di Indonesia tahun 2020 dan memperoleh model regresi terbaik pada pemodelan persentase penduduk miskin serta mengetahui faktor-faktor yang mempengaruhi persentase penduduk miskin di Indonesia tahun 2020.

## 2 TINJAUAN PUSTAKA

### 2.1 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 [8].

Pemodelan regresi probit menggunakan fungsi penghubung (*link function*) distribusi normal dengan fungsi distribusi probabilitas dan fungsi distribusi kumulatif berturut-turut pada persamaan (1) dan (2).

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], -\infty < y < \infty, \sigma^2 > 0 \quad (1)$$

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

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 pada persamaan (3) dan (4).

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right), -\infty < t < \infty \quad (3)$$

$$P(Y \leq y) = \Phi(t) \\ = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du \quad (4)$$

Model regresi probit dapat ditulis seperti pada persamaan berikut :

$$\Phi^{-1}(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \text{ untuk } i = 1, 2, \dots, n \quad (5)$$

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:

$$P(Y_i = 1) = \Phi(\mathbf{x}_i^T \boldsymbol{\beta}), \quad i = 1, 2, \dots, n \quad (6)$$

$$P(Y_i = 0) = 1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (7)$$

Taksiran parameter model regresi probit pada Persamaan (5) 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* [6].

## 2.2 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 dapat mengikuti distribusi Bernoulli dengan fungsi probabilitas [9].

$$P(Y = y) = \pi^y, y = 0,1 \quad (8)$$

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

Model regresi logistik dengan  $k$  variabel bebas dapat dituliskan sebagai berikut :

$$\pi(x) = \frac{\exp(\beta^T x)}{1 + \exp(\beta^T x)} \quad (9)$$

dengan  $\beta^T = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]$  adalah vektor parameter dan  $x = [x_0 \ x_1 \ x_2 \ \dots \ x_k]^T$  adalah vektor variabel bebas dengan  $x_0 = 1$ . Untuk mendapatkan model regresi logistik yang linier dalam parameter dilakukan transformasi logit terhadap persamaan

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta^T x \quad (10)$$

Variabel bebas pada Persamaan (10) 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 [10].

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 berikut.

$$l(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (11)$$

dalam bentuk  $\log l(\beta)$  menjadi dan dinyatakan dalam bentuk  $L(\beta)$  dengan persamaan sebagai berikut

$$L(\beta) = \sum_{j=0}^p (\sum_{i=1}^n y_i x_{ij}) \beta_j - \sum_{i=1}^n \log(1 + \exp(\sum_{j=0}^p \beta_j x_{ij})) \quad (12)$$

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

## 2.3 Pengujian Signifikansi Parameter

Setelah dilakukan estimasi parameter model regresi probit lalu dilanjutkan dengan pengujian parameter. Pengujian ini meliputi uji simultan dan uji parsial. Hipotesis yang digunakan untuk uji simultan diformulasikan sebagai berikut:

$H_0 : \beta_1 = \beta_2 = \dots = 0$   
 $H_1 : \text{minimal terdapat satu } \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, k$

Statistik uji yang digunakan pada uji simultan adalah statistik *Likelihood Ratio* yang diperoleh dengan metode *Likelihood Ratio Test* (LRT) sebagai berikut:

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

di mana  $l(\hat{\omega}) = \ln L(\hat{\omega})$  adalah nilai maksimum *log-likelihood* di bawah  $H_0$  dan  $l(\hat{\Omega}) = \ln L(\hat{\Omega})$  adalah nilai maksimum *log-likelihood* di bawah populasi. Statistik uji pada persamaan (6) 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 ( $P$ )  $< \alpha$ ,  $\alpha$  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

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

di mana  $\hat{\beta}_j$  adalah estimator parameter model dan  $SE(\hat{\beta}_j)$  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 ( $P$ )  $< \alpha$ ,  $\alpha$  adalah tingkat signifikansi dan  $Z_{\alpha/2}$  dapat diperoleh dari tabel distribusi normal standar [8].

#### 2.4 Pengujian Kesesuaian Model

Uji kecocokan model digunakan untuk mengevaluasi kecocokan model dengan data, nilai observasi yang diperoleh sama atau mendekati dengan yang diharapkan dalam model. Adapun model yang digunakan harus memenuhi *Goodness of Fit* (GoF). Suatu model dikategorikan memenuhi *GoF* jika terdapat kesesuaian antara data yang dimasukkan dalam model dengan data yang diamati. Dalam regresi logistik, metode yang digunakan untuk menguji kelayakan model dapat diukur dengan nilai *chi-square* dengan uji *Hosmer and Lemeshow*. Pengujian ini dapat dilakukan dengan melihat nilai *GoF test* yang diukur dengan nilai *Chi-Square* pada tingkat signifikan 5%. Adapun hipotesis yang diuji adalah:

$H_0 = \text{Model yang dihipotesiskan fit dengan data;}$   
 $H_1 = \text{Model yang dihipotesiskan tidak fit dengan data.}$

Statistik uji yang digunakan yaitu :

$$\chi^2_{HL} = \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g(1 - \frac{E_g}{n_g})} \tag{15}$$

di mana  $O_g$  menandakan kejadian yang diamati,  $E_g$  menandakan kejadian yang diharapkan dan  $n_g$  menandakan jumlah pengamatan untuk kelompok ke- $g$ , dan  $G$  adalah jumlah kelompok. Statistik uji mengikuti distribusi *Chi-Square* dengan derajat kebebasan ( $G-2$ ) [11].

## 2.5 Multikolinieritas

Multikolinieritas merupakan keadaan dimana terjadinya hubungan (korelasi) di antara variabel bebas dan yang diharapkan adalah tidak adanya hubungan antar variabel bebas atau *non multikolinieritas* agar hasil pengujian tidak bias. Untuk mendeteksi adanya multikolinieritas adalah dengan menggunakan nilai *Variance Inflation Factor* (VIF). Jika VIF lebih kecil dari 10, maka dalam model tidak terdapat multikolinieritas [12].

$$VIF = \frac{1}{R^2_k} \quad (16)$$

Keterangan :

$\frac{R^2}{k}$  = koefisien determinasi ( $R^2$ ) berganda ketika  $X_k$  diregresikan dengan variabel-variabel  $X$  lainnya.

## 2.6 Akaike Information Criterion (AIC)

Metode untuk memilih model terbaik dapat menggunakan metode AIC. Metode ini didasarkan pada metode *Maximum Likelihood Estimation* (MLE). Untuk menghitung nilai AIC digunakan rumus sebagai berikut:

$$AIC = e^{\frac{2k}{n} \sum_{i=1}^n \hat{u}_i^2} \quad (17)$$

dimana  $k$  adalah jumlah parameter yang diestimasi dalam model regresi,  $n$  adalah jumlah observasi,  $u$  adalah sisa [13].

## 2.7 Pengukuran Ketepatan Klasifikasi

Evaluasi fungsi klasifikasi dilakukan dengan cara membuat tabulasi antara *actual group* dan *predicted group* yang diperoleh dari fungsi klasifikasi dari analisis diskriminan kuadratik dan regresi logistik. Selanjutnya dihitung proporsi pengamatan yang salah diklasifikasikan. Diharapkan proporsi pengamatan yang salah diklasifikasikan tersebut bias sekecil mungkin. Untuk menghitung nilai APER dapat dilakukan melalui tabel klasifikasi sebagai berikut:

Actual Membership	Predicted Membership		Total
	$\hat{y} = 1$	$\hat{y} = 0$	
$y = 1$	$n_{11}$	$n_{12}$	A
$y = 0$	$n_{21}$	$n_{22}$	B
Total	C	D	E

Keterangan:

$n_{11}$  : jumlah obyek dari  $y = 1$  tepat diklasifikasikan sebagai  $\hat{y} = 1$

$n_{12}$ : jumlah obyek dari  $y = 1$  tepat diklasifikasikan sebagai  $\hat{y} = 0$

$n_{21}$ : jumlah obyek dari  $y = 0$  tepat diklasifikasikan sebagai  $\hat{y} = 1$

$n_{22}$  : jumlah obyek dari  $y = 0$  tepat diklasifikasikan sebagai  $\hat{y} = 0$

$$APER = \frac{n_{12} + n_{21}}{n} \quad (18)$$

Kemudian, nilai ketepatan klasifikasi dapat dihitung dengan  $1 - APER$  [14].

## 2.8 Kemiskinan

Kemiskinan merupakan masalah sosial yang terus ada di kehidupan masyarakat. Masalah kemiskinan sangatlah lama, dan dalam waktu yang panjang, sama seperti halnya dengan usia manusia itu sendiri, dan unsur pokok permasalahannya adalah menyangkut berbagai macam bentuk atau karakter kehidupan manusia. Dengan kata lain bahwa kemiskinan ini merupakan masalah kehidupan yang sifatnya global atau mendunia, artinya masalah kemiskinan sudah menjadi perhatian dunia, dan masalah tersebut ada di semua negara, walaupun dampak dari kemiskinan sangatlah berbeda-beda [15].

Kemiskinan adalah ketidakmampuan dari sisi ekonomi, materi dan fisik untuk mencukupi kebutuhan dasar makanan dan bukan makanan yang di ukur dengan pengeluaran. Ukuran kemiskinan yaitu menggunakan Garis kemiskinan. Yang terdiri dari garis kemiskinan makanan (GKM), dan garis kemiskinan non makanan (GKNM). Garis kemiskinan makanan adalah nilai pengeluaran yang di hasilkan dari nilai kebutuhan minimum makanan yang di hitung dalam 2.100 kkalori perkapita per hari, sedangkan garis kemiskinan non makanan di hitung dari kebutuhan minimum untuk sandang, pendidikan, dan kesehatan dan kebutuhan dasar lainnya [16].

## 3 DATA

### 3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari *website* resmi Badan Pusat Statistik Indonesia <https://www.bps.go.id/> yang dikumpulkan berdasarkan teknik *purposive sampling*. Data yang diambil yaitu data persentase penduduk miskin sebagai variabel terikat ( $Y$ ) dengan kategori 0 = persentase penduduk miskin provinsi < rata-rata persentase penduduk miskin Indonesia sebesar 10,80% maka dikatakan tidak miskin, 1 = persentase penduduk miskin provinsi  $\geq$  rata-rata persentase penduduk miskin di Indonesia sebesar 10,80% maka dikatakan miskin.

### 3.2 Variabel Penelitian

Variabel yang digunakan pada penelitian kali ini terdiri dari variabel terikat persentase penduduk miskin dan enam variabel bebas yaitu tingkat pengangguran terbuka (TPT) ( $X_1$ ), persentase tenaga kerja formal ( $X_2$ ), proyek investasi dalam negeri ( $X_3$ ), indeks pembangunan manusia (IPM) ( $X_4$ ), angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun ( $X_5$ ), dan tingkat setengah pengangguran (TSP) ( $X_6$ ). Adapun unit pengamatan yang digunakan dalam penelitian ini adalah provinsi di Indonesia pada tahun 2020 yang terdiri dari 34 provinsi.

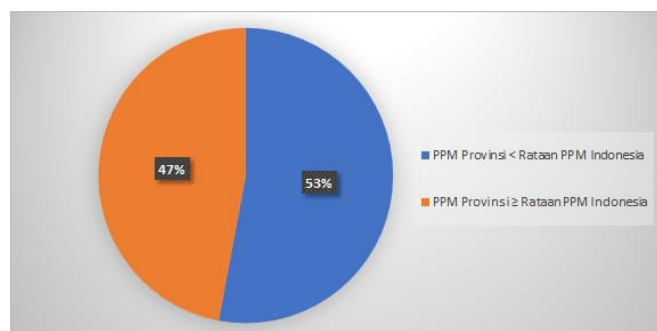
### 3.3 Teknik Analisis Data

Adapun teknik analisis data yang digunakan dalam penelitian ini dengan menggunakan bantuan software R 4.0.3 adalah sebagai berikut:

- Melakukan analisis statistika deskriptif pada variabel penelitian.
- Melakukan uji multikolinieritas pada semua variabel bebas berdasarkan persamaan (16).
- Melakukan estimasi parameter model regresi probit dan regresi logistik dengan metode *maximum likelihood* dan *fisher scoring*.
- Melakukan pengujian hipotesis parameter secara simultan dan parsial berturut-turut pada regresi probit dan regresi logistik menggunakan metode LRT dan uji Wald berdasarkan persamaan (13) dan (14).
- Mendapatkan variabel bebas yang signifikan berpengaruh terhadap variabel terikat pada regresi probit dan regresi logistik.
- Menghitung nilai AIC regresi probit dan regresi logistik menggunakan persamaan (17) lalu membandingkan keduanya.
- Mengitung ketepatan klasifikasi berdasarkan nilai APER menggunakan persamaan (18) lalu membandingkan keduanya.
- Melakukan pemilihan model terbaik berdasarkan nilai AIC dan ketepatan klasifikasi.
- Memperoleh kesimpulan berdasarkan hasil dan pembahasan penelitian.

## 4 HASIL DAN PEMBAHASAN

Data penelitian yang akan dianalisis menggunakan regresi probit diawali dengan mendeskripsikan variabel penelitian. Deskripsi dari variabel terikat dan variabel bebas disajikan pada Gambar 1 dan Tabel 1.



**Gambar 1.** Deskripsi variabel terikat

Berdasarkan Gambar 1 dapat diketahui bahwa pada provinsi di Indonesia tahun 2020 persentase penduduk miskin provinsi yang kurang dari rata-rata persentase



penduduk miskin Indonesia adalah sebanyak 53% atau sebanyak 18 provinsi. Sedangkan, persentase penduduk miskin provinsi yang lebih dari atau sama dengan rata-rata persentase penduduk miskin Indonesia adalah sebanyak 47% atau sebanyak 16 provinsi.

**Tabel 1.** Deskripsi Variabel Bebas

Variabel	Min	Maks	Mean	Std
$X_1$	3,32	10,95	6,03	2,01
$X_2$	20,08	64,59	38,99	9,62
$X_3$	113	17.667	2.841,85	4.112,83
$X_4$	60,44	80,77	71,08	3,90
$X_5$	17,47	51,81	27,36	6,58
$X_6$	5,75	16,83	10,70	2,66

Berdasarkan Tabel 1, dapat diketahui bahwa pada variabel  $X_1$  yaitu TPT menunjukkan bahwa TPT paling rendah adalah sebesar 3,32% yang terletak di provinsi Sulawesi Barat dan TPT paling tinggi adalah sebesar 10,95% yang terletak di provinsi DKI Jakarta dengan rata-rata TPT sebesar 6,03% dan simpangan bakunya sebesar 2,01%. Variabel  $X_2$  yaitu persentase tenaga kerja formal menunjukkan bahwa persentase tenaga kerja paling rendah adalah sebesar 20,08% yang terletak di provinsi Papua dan yang paling tinggi adalah sebesar 64,59% yang terletak di provinsi Kepulauan Riau dengan rata-rata sebesar 38,99% dan simpangan bakunya sebesar 9,62%. Variabel  $X_3$  yaitu jumlah proyek investasi dalam negeri menunjukkan bahwa jumlah proyek investasi paling sedikit adalah sebesar 113 unit yang terletak di provinsi Sulawesi Barat dan yang paling banyak adalah sebesar 17.667 unit yang terletak di provinsi DKI Jakarta dengan rata-rata sebesar 2.841,85 unit dan simpangan bakunya sebesar 4.112,83 unit. Variabel  $X_4$  yaitu IPM menunjukkan IPM paling rendah adalah sebesar 60,44 yang terletak di provinsi Papua dan yang paling tinggi adalah sebesar 80,77 yang terletak di provinsi DKI Jakarta dengan rata-rata sebesar 71,08 dan simpangan bakunya sebesar 3,9. Variabel  $X_5$  yaitu angka partisipasi murni dan kasar sekolah usia 19 – 24 (APS) menunjukkan APS paling rendah adalah sebesar 17,47 yang terletak di provinsi Kepulauan Bangka Belitung dan yang paling tinggi adalah sebesar 51,81 yang terletak di provinsi DI Yogyakarta dengan rata-rata sebesar 27,86 dan simpangan bakunya sebesar 6,58. Variabel  $X_5$  yaitu tingkat setengah pengangguran (TPS) menunjukkan TPS paling rendah adalah sebesar 5,75% yang terletak di provinsi Kepulauan Riau dan yang paling tinggi adalah sebesar 16,83% yang terletak di provinsi Nusa Tenggara Barat dengan rata-rata sebesar 10,7% dan simpangan bakunya sebesar 2,66%.

Sebelum melakukan estimasi parameter pada regresi probit terlebih dahulu dilakukan pemeriksaan multikolinieritas antara variabel bebas agar hasil estimasi tidak bersifat bias jika tidak ada korelasi antar variabel bebasnya.

**Tabel 2.** Nilai VIF Multikolinieritas

Variabel Bebas	VIF
$X_1$	9,251
$X_2$	1,246
$X_3$	8,715
$X_4$	4,525
$X_5$	7,698
$X_6$	9,887

Berdasarkan Tabel 2, dapat diketahui bahwa keenam variabel bebas memiliki nilai VIF yang lebih kecil dari 10 sehingga dapat disimpulkan bahwa tidak terjadi multikolinieritas pada variabel bebasnya. Sehingga, analisis regresi probit dapat dilanjutkan menggunakan variabel-variabel terikat ini

Tahapan awal dalam memodelkan persentase penduduk miskin di Indonesia tahun 2020 dengan pendekatan regresi probit dan regresi logistik adalah menentukan *estimator* parameter model regresi probit dan regresi logistik dengan metode MLE dan *Fisher Scoring*. Hasil estimasi dan nilai statistik uji dari regresi probit yang diperoleh melalui *software R* disajikan pada Tabel 3.

**Tabel 3.** Estimasi Parameter dan Nilai Statistik Uji Parsial Model Probit

Parameter	Estimasi	SE	$Z_{hit}$	<i>P-Value</i>
$\beta_0$	43,215	21,83	1,979	0,0478*
$\beta_1$	-3,261	1,50	-2,164	0,0305*
$\beta_2$	-0,050	0,14	-0,356	0,7218
$\beta_3$	0,001	0,0007	2,249	0,0245*
$\beta_4$	-0,739	0,349	-2,115	0,0344*
$\beta_5$	0,395	0,165	2,394	0,0167*
$\beta_6$	1,344	0,667	2,014	0,0440*

\*) Signifikan pada  $\alpha = 0,05$ .

Setelah diperoleh estimasi parameter model regresi probit pada Tabel 3, selanjutnya dapat dilakukan uji simultan dan parsial terhadap parameter model regresi probit.

**Tabel 4.** Uji Simultan Model Probit

<i>Likelihood Ratio Test</i>	
$G^2$	35,88
<i>P-Value</i>	$2,909 \times 10^{-6}$

Berdasarkan statistik uji pada persamaan (13) dan Tabel 4 dapat diketahui bahwa nilai  $G^2 = 35,88 > \chi^2_{(0,05;6)} = 12,592$  dan  $P\text{-value} = 2,909 \times 10^{-6} < \alpha = 0,05$  maka diputuskan tolak  $H_0$ . Sehingga dapat disimpulkan bahwa secara simultan parameter intersep dan parameter dari variabel tingkat pengangguran terbuka (TPT), persentase tenaga kerja formal, proyek investasi dalam negeri, indeks pembangunan manusia (IPM), angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun, dan tingkat setengah pengangguran (TSP) berpengaruh pada persentase penduduk miskin di Indonesia tahun 2020.

Berdasarkan Statistik uji pada persamaan (14) dan Tabel 3 dapat diketahui bahwa parameter yang signifikan berpengaruh terhadap persentase penduduk miskin di Indonesia tahun 2020 adalah parameter intersep dan parameter dari

variabel tingkat pengangguran terbuka (TPT), proyek investasi dalam negeri, indeks pembangunan manusia (IPM), angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun, dan tingkat setengah pengangguran (TSP) karena berdasarkan uji Wald variabel-variabel tersebut memiliki nilai  $Z_{hit} > Z_{(0,025)} = 1,96$  dan  $P\text{-value} < \alpha = 0,05$  maka diputuskan tolak  $H_0$ .

**Tabel 5.** Kesesuaian Model Probit

<i>Hosmer and Lemeshow Goodness of Fit (GOF) Test</i>	
$\chi^2$	2,084
<i>P-Value</i>	0,9783

Berdasarkan Statistik uji pada persamaan (15) dan Tabel 8 dapat diketahui bahwa nilai  $\chi^2 = 2,084 < \chi^2_{(0,05;8)} = 15,507$  dan  $P\text{-value} = 0,9783 > \alpha = 0,05$  maka diputuskan gagal tolak  $H_0$  sehingga dapat disimpulkan bahwa tidak terdapat perbedaan hasil pengamatan dengan hasil dugaan.

Adapun hasil estimasi dan nilai statistik uji model regresi logistik yang diperoleh melalui perhitungan *software R* disajikan pada Tabel 6.

**Tabel 6.** Estimasi Parameter dan Nilai Statistik Uji Parsial Model Logistik

Parameter	Estimasi	SE	$Z_{hit}$	<i>P-Value</i>
$\beta_0$	76,016	41,94	1,813	0,0699
$\beta_1$	-5,683	2,88	-1,972	0,0487*
$\beta_2$	-0,106	0,264	-0,402	0,6879
$\beta_3$	0,002	0,0014	1,998	0,0457*
$\beta_4$	-1,280	0,679	-1,883	0,0597
$\beta_5$	0,679	0,320	2,119	0,0341*
$\beta_6$	2,325	1,243	1,870	0,0614

\*) Signifikan pada  $\alpha = 0,05$ .

Setelah diperoleh *estimator* parameter model regresi probit pada Tabel 6. Selanjutnya dapat dilakukan uji simultan dan parsial terhadap parameter model regresi logistik.

**Tabel 7.** Uji Simultan Model Logistik

<i>Likelihood Ratio Test</i>	
$G^2$	35,806
<i>P-Value</i>	$3,006 \times 10^{-6}$

Berdasarkan statistik uji pada persamaan (13) dan Tabel 7, dapat diketahui bahwa nilai  $G^2 = 35,806 > \chi^2_{(0,05;6)} = 12,592$  dan  $P\text{-value} = 3,006 \times 10^{-6} < \alpha = 0,05$  maka diputuskan tolak  $H_0$ . Sehingga dapat disimpulkan bahwa secara simultan parameter intersep dan parameter dari variabel tingkat pengangguran terbuka (TPT), persentase tenaga kerja formal, proyek investasi dalam negeri, indeks pembangunan manusia (IPM), angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun, dan tingkat setengah pengangguran (TSP) berpengaruh pada persentase penduduk miskin di Indonesia tahun 2020.

Berdasarkan Statistik uji pada persamaan (14) dan Tabel 6, dapat diketahui bahwa parameter yang signifikan berpengaruh terhadap persentase penduduk miskin di Indonesia tahun 2020 adalah parameter intersep dan parameter dari

variabel tingkat pengangguran terbuka (TPT), proyek investasi dalam negeri, serta angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun karena berdasarkan uji Wald variabel-variabel tersebut memiliki nilai  $Z_{hit} > Z_{(0,025)} = 1,96$  dan  $P\text{-value} < \alpha = 0,05$  maka diputuskan tolak  $H_0$ .

**Tabel 8.** Kesesuaian Model Logistik

<i>Hosmer and Lemeshow Goodness of Fit (GOF) Test</i>	
$\chi^2$	5,585
<i>P-Value</i>	0,6937

Berdasarkan Statistik uji pada persamaan (15) dan Tabel 8, dapat diketahui bahwa nilai  $\chi^2 = 5,585 < \chi^2_{(0,05;8)} = 15,507$  dan  $P\text{-value} = 0,6937 > \alpha = 0,05$  maka diputuskan gagal tolak  $H_0$  sehingga dapat disimpulkan bahwa tidak terdapat perbedaan hasil pengamatan dengan hasil dugaan.

Selanjutnya, akan dibandingkan nilai AIC dari model regresi probit dan regresi logistik. Berikut adalah tabel perbandingan nilai AIC kedua model:

**Tabel 9.** Perbandingan Nilai AIC

Model	Nilai AIC
Regresi Probit	25,137
Regresi Logistik	25,21

Berdasarkan Tabel 9, dapat dilihat bahwa model regresi probit menghasilkan nilai AIC sebesar 25,137 di mana nilai tersebut lebih kecil daripada nilai AIC model regresi logistik yaitu sebesar 25,21 dengan selisih nilai AIC model regresi probit dan regresi logistik adalah sebesar 0,073. Sehingga berdasarkan nilai AIC tersebut dapat disimpulkan bahwa model regresi probit lebih baik atau sesuai untuk memodelkan persentase penduduk miskin di Indonesia tahun 2020.

Kemudian, akan dilihat ketepatan klasifikasi dari kedua model. Model dengan nilai ketepatan klasifikasi terbesar akan menjadi model terbaik. Berikut adalah tabel klasifikasi model probit dan model logistik.

**Tabel 10.** Klasifikasi Model Regresi Probit

Aktual	Prediksi		Total
	Miskin	Tidak Miskin	
Miskin	13	3	16
Tidak Miskin	0	18	18
Total	13	21	34

Setelah diketahui ketepatan klasifikasi untuk tiap-tiap kelompok maka ketepatan hasil klasifikasi secara total dapat diketahui dengan menghitung nilai APER. Berdasarkan Persamaan (18) maka diperoleh nilai APER sebagai berikut:

$$APER = \frac{3 + 0}{34} = 0,0882$$

$$\text{Ketepatan Klasifikasi} = 1 - 0,0882 = 0,9118$$

Berdasarkan nilai APER di atas dapat diketahui bahwa proporsi salah klasifikasi sebesar 8,82% sehingga ketepatan klasifikasi adalah 91,18%.

**Tabel 11.** Klasifikasi Model Regresi Logistik

Aktual	Prediksi		Total
	Miskin	Tidak Miskin	
Miskin	13	3	16
Tidak Miskin	1	17	18
Total	14	20	34

Setelah diketahui ketepatan klasifikasi untuk tiap-tiap kelompok maka ketepatan hasil klasifikasi secara total dapat diketahui dengan menghitung nilai APER. Berdasarkan Persamaan (18) maka diperoleh nilai APER sebagai berikut :

$$APER = \frac{3 + 1}{34} = 0,1176$$

$$\text{Ketepatan Klasifikasi} = 1 - 0,1176 = 0,8824$$

Berdasarkan nilai APER di atas dapat diketahui bahwa proporsi salah klasifikasi sebesar 11,76% sehingga ketepatan klasifikasi adalah 88,24%.

**Tabel 12.** Perbandingan Ketepatan Klasifikasi

	Model Probit	Model Logistik
APER	8,82%	11,76%
Ketepatan Klasifikasi	91,18%	88,24%.

Berdasarkan nilai ketepatan klasifikasi, dapat dilihat bahwa model probit memiliki nilai ketepatan yang lebih tinggi. Sehingga, model probit merupakan model terbaik untuk memodelkan persentase penduduk miskin di Indonesia tahun 2020.

Setelah diperoleh perbandingan nilai AIC dan perbandingan ketepatan klasifikasi dari model probit dan model logistik, maka pada tahap ini dapat dilakukan pemilihan model terbaik.

**Tabel 13.** Perbandingan Model Terbaik

Model	Kriteria	
	Nilai AIC	Nilai Ketepatan Klasifikasi
Regresi Probit	25,137	91,18%
Regresi Logistik	25,21	88,24%

Berdasarkan Tabel 13, dapat diketahui bahwa model regresi probit memiliki nilai AIC yang lebih kecil dibandingkan model regresi logistik dan nilai ketepatan klasifikasi model regresi probit lebih besar dibandingkan model regresi logistik. Sehingga dapat disimpulkan bahwa model terbaik untuk memodelkan persentase penduduk miskin di Indonesia tahun 2020 adalah model regresi probit.

Adapun model regresi probit pada data persentase penduduk miskin di Indonesia tahun 2020 adalah sebagai berikut :

$$\Phi^{-1}(\pi_i) = -3,315X_1 + 0,00164X_3 - 0,762X_4 + 0,399X_5 + 1,387X_6$$

## 5 KESIMPULAN

Berdasarkan statistika deskriptif maka gambaran umum dari persentase penduduk miskin di Indonesia tahun 2020 adalah persentase penduduk miskin provinsi yang kurang dari rata-rata persentase penduduk miskin Indonesia adalah sebanyak 53% atau sebanyak 18 provinsi, Sedangkan, persentase penduduk miskin provinsi yang lebih dari atau sama dengan rata-rata persentase penduduk miskin Indonesia adalah sebanyak 47% atau sebanyak 16 provinsi.

Berdasarkan hasil analisis model regresi probit dan regresi logistik diperoleh model terbaik pada pemodelan persentase penduduk miskin di Indonesia adalah model regresi probit dengan faktor-faktor yang mempengaruhi persentase penduduk miskin adalah tingkat pengangguran terbuka (TPT), proyek investasi dalam negeri, indeks pembangunan manusia (IPM), angka partisipasi murni dan kasar sekolah usia 19 – 24 tahun, dan tingkat setengah pengangguran (TSP).

Saran yang diajukan dari penelitian ini adalah dalam memilih model terbaik dapat digunakan ukuran keakuratan lainnya seperti *Mean percentage Error* (MAPE) atau *Mean Square Error* (MSE).

## DAFTAR PUSTAKA

- [1] Sudjana. (2005). *Metode Statistika*. Bandung: PT. Tarsito.
- [2] Supratno. (2000). *Statistika: Teori dan Aplikasi*. Erlangga: Jakarta.
- [3] Agresti. (2013). *Categorical data analysis, 3rd Edition*. New York: John Wiley & Sons.
- [4] Yong. (2003). *Penaksiran Maksimum Likelihood bagi Model Probit dan Model Probit Bivariat*. Diakses tanggal 28 Februari 2017. <http://home.unpar.ac.id/~integral/Volume%208/integral%201%20%201/penaksiran%20maksimum%20likelihood.pdf>.
- [5] Khandker, Shahidur. (2009). *Pedoman Tentang Kemiskinan dan Ketimpangan*. Jakarta: Salemba Empat.
- [6] Fathurahman, M. (2019). Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Pulau Kalimantan Menggunakan Pendekatan Regresi Probit, *Jurnal Varian*, 2 (2), 47-54.
- [7] Yumira. (2017). Analisis Regresi Logistik untuk Menentukan Faktor-Faktor yang mempengaruhi IPK Mahasiswa FMIPA Universitas Sam Ratulangi Manado, *Jurnal JdC*, 6 (2), 56 - 62.
- [8] Greene, W, H. (2003). *Econometric analysis fifth edition*. New York: Prentice Hall.
- [9] Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X., (2013). *Applied logistic regression 3rd Edition*. New York: John Wiley & Sons.
- [10] Kutner, M.H., C.J. Nachtsheim., dan J. Neter. (2004). *Applied Linear Regression Models 4th Editon*. New York: McGraw-Hill Companies, Inc.
- [11] Kusmana. (2014). Model regresi logistik biner stratifikasi pada partisipasi ekonomi perempuan di provinsi Jawa Timur, *Jurnal sains dan seni pomits* Vol. 3 No. 1.
- [12] Zain, Sumarno. (1995). *Ekonometrika Dasar*. Jakarta: Erlangga.
- [13] Widarjono, A. (2007), *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*, Ekonisia Yogyakarta: Fakultas Ekonomi Universitas Islam Indonesia.

- [14] Johnson, R. A., dan D. W Winchern. (2002). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- [15] Nurwati, Nunung. (2008). Kemiskinan Model Pengukuran, Permasalahan dan Alternatif Kebijakan. *Jurnal Ilmu Ekonomi*, Vol. 1, No. 2, Hal. 1 – 8.
- [16] Badan Pusat Statistik. (2021). *Statistik Indonesia 2021*. BPS: Indonesia.