

PERBANDINGAN EMPAT FUNGSI GENERALIZED
LINEAR MODEL UNTUK MENGANALISIS
FAKTOR-FAKTOR YANG MEMPENGARUHI INDEKS
PEMBANGUNAN MANUSIA (IPM)

(COMPARING FOUR GENERALIZED LINEAR MODELING
LINK FUNCTIONS OF ANALYZING FACTORS
AFFECTING THE HUMAN DEVELOPMENT INDEX
(HDI))

ROSSA FITRIA HALIM, RENANTA DZAKIYA NAFALANA, ADITYO
WAHYU SAPUTRO, MOHAMMAD AGUS KHOLILURROHMAN

Abstract. Development is a process of change that is planned to improve various aspects of people's lives. The development process carried out in Indonesia covers many aspects, one of which is human development. Indicators of the success of a region in the process of developing the quality of human life can be measured using the Human Development Index (HDI). HDI is a number that represents the condition of the population in terms of development, education, health, income, and various other aspects. Human development is important to do in order to achieve the prosperity of a region's population. Each region in Indonesia continues to make various efforts to improve human development, and Java is no exception. To find out what factors affect the Human Development Index on the Java Island, a comparison of Four Mathematical Function Link Models on Generalized Linear Models (GLM) is carried out to analyze the factors that affect the Human Development Index. The Four Mathematical Link Function Models are Logit, Probit, Cauchit, and Complementary Log-log (Clog-log) models. The ordinal Probit regression model is the best model to analyze the factors affecting HDI in Java Island in 2023, with classification accuracy of 86.316%.

Keywords: Human Development Index (HDI), Mathematical Link Function Model, Generalized Linear Models, Classification.

Abstrak. Pembangunan merupakan proses perubahan yang direncanakan untuk memperbaiki berbagai aspek kehidupan masyarakat. Proses pembangunan yang dilakukan di Indonesia mencakup banyak aspek, salah satunya pembangunan manusia. Indikator keberhasilan suatu wilayah dalam proses pembangunan kualitas hidup manusia dapat diukur menggunakan Indeks Pembangunan Manusia (IPM). IPM merupakan angka yang mewakili kondisi dari penduduk terhadap aspek pembangunan, pendidikan, kesehatan, pendapatan, dan berbagai aspek lainnya. Pembangunan manusia penting untuk dilakukan agar mencapai kemakmuran penduduk suatu wilayah. Tiap wilayah di Indonesia terus melakukan berbagai upaya untuk meningkatkan pembangunan manusia, tidak terkecuali pulau Jawa. Guna mengetahui faktor apa saja yang mempengaruhi Indeks Pembangunan Manusia di pulau Jawa maka dilakukan perbandingan empat model link fungsi matematika pada *Generalized Linear Models* (GLM) untuk menganalisis faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia. Adapun empat model link fungsi matematika tersebut adalah model Logit, Probit, Cauchit, dan *Complementary Log-log* (Clog-log). Diperoleh hasil bahwa model regresi Probit ordinal merupakan model terbaik untuk menganalisis faktor-faktor yang mempengaruhi IPM di Pulau Jawa tahun 2023, dengan ketepatan klasifikasi yaitu akurasi 86,316%.

Kata-kata kunci: Indeks Pembangunan Manusia (IPM), Model Link Fungsi Matematika, *Generalized Linear Models*, Klasifikasi.

1. PENDAHULUAN

Indikator keberhasilan suatu wilayah dalam proses pembangunan kualitas hidup manusia dapat diukur menggunakan Indeks Pembangunan Manusia (IPM). IPM merupakan angka yang mewakili kondisi dari penduduk terhadap aspek pembangunan, pendidikan, kesehatan, pendapatan, dan berbagai aspek lainnya. Banyak faktor yang mempengaruhi nilai IPM [4]. Menurut BPS (2023) terdapat faktor-faktor yang mempengaruhi IPM antara lain: usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, persentase penduduk miskin, tingkat pengangguran terbuka, persentase rumah tangga dengan akses air minum layak minum.

IPM pertama kali diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 melalui publikasi *Human Development Report* 1990. UNDP menetapkan tiga dimensi pembentuk IPM. Ketiga dimensi ini merupakan pendekatan yang dipilih dalam penggambaran kualitas hidup manusia dan tidak mengalami perubahan hingga saat ini. Adapun ketiga dimensi tersebut mencakup: Umur panjang dan hidup sehat, pengetahuan, dan standart hidup layak [6]. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya, untuk mengukur dimensi pengetahuan digunakan gabungan indikator angka melek huruf dan rata-rata lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator PNB (Pendapatan Nasional Bruto) per kapita.

Sejak tahun 2014 angka IPM di Indonesia disajikan secara periodik setiap tahun pada tingkat nasional, provinsi, dan kabupaten/kota. Penyajian IPM secara periodik menurut daerah memungkinkan setiap provinsi dan kabupaten/kota mengetahui peta

pembangunan manusia di daerahnya, baik capaian, kecepatan, posisi, maupun disparitas antardaerah. Menurut BPS (2020), capaian pembangunan manusia di suatu wilayah pada waktu tertentu dapat dikelompokkan ke dalam empat kelompok berdasarkan status capaiannya. Adapun pengelompokkan tersebut sebagai berikut [11]: (i) Sangat tinggi : $IPM \geq 80$; (ii) Tinggi: $70 \leq IPM < 80$; (iii) Sedang : $60 \leq IPM < 70$; (iv) Rendah : $IPM < 60$.

Tiap wilayah di Indonesia terus melakukan berbagai upaya untuk meningkatkan pembangunan manusianya, tidak terkecuali pulau Jawa. Dengan jumlah penduduk yang besar, yaitu sekitar 151,59 juta menjadikan pulau Jawa memiliki modal yang bagus dalam proses pembangunannya. Jumlah penduduk yang besar jika tidak dikelola dengan baik akan menimbulkan masalah di masa yang akan datang. Pembangunan manusia penting untuk dilakukan agar mencapai kemakmuran penduduk suatu wilayah. Oleh karena itu, penelitian ini bertujuan untuk mengetahui faktor apa saja yang paling berpengaruh dari faktor-faktor yang telah disebutkan dalam peningkatan nilai IPM di pulau Jawa. Hal ini dapat menjadi pertimbangan bagi pemerintah dan masyarakat agar bisa lebih fokus untuk meningkatkan faktor tersebut guna menaikkan angka IPM.

2. TINJAUAN PUSTAKA

2.1. Indeks Pembangunan Manusia (IPM).

Pembangunan manusia merupakan suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia yang terdiri dari tiga komponen dasar, yaitu untuk berumur panjang dan sehat, untuk memiliki ilmu pengetahuan dan yang ketiga untuk mempunyai akses terhadap sumber daya yang dibutuhkan sehingga dapat menjalani kehidupan yang layak HDR [9]. Indeks Pembangunan Manusia (IPM) diartikan sebagai suatu indeks komposit yang digunakan untuk mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup manusia. Ketiga komponen dasar yang digunakan sebagai ukuran kualitas hidup tersebut diukur dengan menggunakan suatu indeks untuk masing-masing komponen, yaitu indeks harapan hidup, indeks pendidikan dan indeks standar hidup layak BPS [5].

Ketiga dimensi dasar pembangun IPM memiliki pengertian yang sangat luas, hal tersebut dikarenakan masing-masing dimensi memiliki keterkaitan dengan banyak faktor yang mempengaruhi. Pada pengukuran dimensi berumur panjang dan sehat, digunakan Angka Harapan Hidup sebagai ukurannya. Angka Harapan Hidup (AHH) merupakan rata-rata perkiraan jumlah tahun yang dapat ditempuh oleh seseorang selama hidup atau dapat dikatakan sebagai rata-rata perkiraan usia seseorang BPS [4]. Nilai AHH dihitung berdasarkan dua komponen, yaitu Anak Lahir Hidup (ALH) dan Anak Masih Hidup (AMH).

Dimensi pengetahuan penduduk suatu wilayah, diukur berdasarkan indeks pendidikan yang didapatkan berdasarkan dua indikator yaitu Rata-rata Lama Sekolah dan Angka Melek Huruf. Rata-rata lama sekolah merupakan jumlah tahun yang digunakan oleh penduduk usia 15 tahun ke atas dalam menjalani pendidikan formal. Sedangkan

angka melek huruf merupakan prosentase penduduk usia 15 tahun ke atas yang dapat membaca dan menulis huruf latin atau huruf lainnya (BPS, 2008).

Indeks standar hidup layak adalah indeks untuk mengukur dimensi kehidupan yang layak. Standar hidup layak menggambarkan tingkat kesejahteraan yang dinikmati oleh penduduk sebagai dampak semakin membaiknya kondisi ekonomi suatu wilayah. United Nations Development Programme (UNDP) mengukur standar hidup layak menggunakan Produk Domestik Regional Bruto (PDRB) riil yang disesuaikan, sedangkan Badan Pusat Statistik (BPS) menggunakan rata-rata pengeluaran per kapita riil yang disesuaikan. Perhitungan daya beli dilakukan berdasarkan 27 komoditas kebutuhan pokok masyarakat.

2.2. Regresi Logistik.

Regresi Logistik merupakan analisis yang digunakan untuk mencari hubungan variabel respon yang bersifat dikotomis (berskala nominal atau ordinal dengan dua kategori) atau polikotomis (memiliki skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel prediktor. Sedangkan variabel prediktor bersifat kontinu atau kategorik Agresti [1]. Selain itu, Regresi Logistik adalah metode analisis statistik yang digunakan untuk menggambarkan hubungan antara variabel independen dan variabel dependen yang mempunyai dua atau lebih kategori dengan variabel independen berskala kategorik maupun interval Hosmer and Lemeshow [10]. Regresi Logistik dilakukan untuk menjelaskan hubungan antar variabel dependen maupun independen yang memiliki karakteristik yang berbeda dengan regresi biasa. Jika regresi biasa menggunakan variabel dependen bersifat kontinu, maka pada Regresi Logistik variabel dependen bersifat dikotomis maupun polikotomis.

2.2.1. *Model Logit.* Model Logit merupakan salah satu model Regresi Logistik yang mempunyai persamaan

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.1)$$

dengan p merupakan banyaknya variabel independen.

Model Logit merupakan salah satu model regresi *non linear* yang menghasilkan persamaan dimana variabel dependen bersifat kategorikal. Penggunaan model Logit seringkali digunakan dalam klasifikasi Gujarati [8]. Pada model Regresi Logistik persamaan yang digunakan adalah

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}. \quad (2.2)$$

Untuk lebih mudah melakukan estimasi parameter, maka persamaan 2.2 dilakukan transformasi Logit sehingga akan didapatkan

$$\pi(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.3)$$

dengan p merupakan banyaknya variabel independen.

2.2.2. *Model Probit.* Model Probit ditemukan pertama kali oleh Chester Bliss pada tahun 1930. Probit merupakan singkatan dari probability unit dimana model ini menggunakan fungsi normal kumulatif atau normal baku Gujarati [8].

Persamaan model Probit dapat ditulis dengan

$$P(y = 1|x) = \Phi(\beta'x) \quad (2.4)$$

atau dapat ditulis dengan

$$P(y = 1|x) = \int_{-\infty}^{\beta'x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x^2)\right) dx. \quad (2.5)$$

Secara umum model Probit dapat dinyatakan dalam bentuk

$$P_i = F(X_i) = F(\beta_0 + \beta_1x_1 + \dots + \beta_px_p) \quad (2.6)$$

dengan F merupakan fungsi peluang kumulatif dan X_i merupakan variabel independen. Untuk memperoleh suatu dugaan dari nilai peluang Probit (X_i) maka dapat digunakan invers dari fungsi normal kumulatif sehingga diperoleh persamaan

$$X_i = F^{-1}(P_i) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p. \quad (2.7)$$

2.2.3. *Model Cauchit.* Model Cauchit merupakan salah satu model dalam Regresi Logistik. Model Cauchit digunakan jika peubah laten mempunyai nilai yang ekstrim Agresti [2]. Model Cauchit mempunyai persamaan

$$Y = \tan\left[\pi\left(P - \frac{1}{2}\right)\right]. \quad (2.8)$$

Fungsi distribusi kumulatif (CDF) dari distribusi Cauchy yang kurvanya sigmoid seperti kurva Probit dan Logistik adalah sebagai berikut:

$$P = \frac{1}{\pi} \arctan(Y) + \frac{1}{2}. \quad (2.9)$$

2.2.4. *Model Complementary Log-log (Clog-log).* Model Complementary Log-log atau yang lebih dikenal dengan Clog-log merupakan salah satu model yang dinilai tepat untuk digunakan ketika data pada variabel respon termasuk ke dalam kategori jarang (*rare events*) Agresti [2]. Transformasi Complementary Log-log dinilai lebih tepat untuk digunakan karena kebanyakan observasi pada variabel respon akan berada pada kelompok kategori tertentu atau dapat dikatakan bahwa pada variabel respon tidak berdistribusi simetris. Model Clog-log mempunyai persamaan

$$Y = \ln(-\ln(1-p)) \quad (2.10)$$

yang merupakan kebalikan dari CDF dari distribusi nilai ekstrim (atau Log-Weibull), dengan CDF

$$F(Y) = 1 - e^{-e^Y}. \quad (2.11)$$

2.3. Uji Signifikansi Parameter Secara Keseluruhan (Uji Rasio Likelihood).

Uji Rasio Likelihood adalah pengujian yang dilakukan untuk memeriksa signifikansi parameter β terhadap variabel respon secara keseluruhan Hosmer and Lemeshow [10]. Pengujian signifikansi parameter β secara serentak menggunakan Likelihood Ratio Test. Uji ini membandingkan model lengkap (model dengan variabel independen) terhadap model yang dengan hanya menggunakan konstanta (model tanpa variabel independen). Hipotesis pengujiannya adalah sebagai berikut:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = 0, j = 1, 2, \dots, p$ (Variabel prediktor tidak berpengaruh signifikan terhadap variabel respon).

$H_1 : \exists \beta_j \neq 0, j = 1, 2, \dots, p$ (Paling sedikit terdapat satu variabel prediktor tidak berpengaruh signifikan terhadap variabel respon).

Statistik uji yang digunakan adalah Likelihood Ratio Test dengan rumus

$$G = -2 \ln \left[\frac{\text{likelihood (model B)}}{\text{likelihood (model A)}} \right] \quad (2.12)$$

dengan *model B* merupakan model dengan hanya konstanta, sedangkan *model A* merupakan model dengan variabel independen, dengan tingkat kepercayaan α dan derajat bebas (*df*) yang merupakan banyaknya variabel prediktor. Pengambilan keputusan didasarkan dengan menolak H_0 apabila nilai $G > \chi_{\alpha, df}^2$ dan *p-value* $< \alpha$. Dengan menolak H_0 maka terdapat hubungan yang signifikan antara variabel respon dengan variabel prediktor.

2.4. Uji Signifikansi Parameter Parsial (Uji Wald).

Uji Wald merupakan pengujian yang dilakukan untuk mengetahui signifikansi parameter terhadap variabel respon Hosmer and Lemeshow [10]. Pengujian signifikansi parameter β secara parsial menggunakan uji *Wald*. Hipotesis pengujiannya sebagai berikut:

$H_0 : \beta_j = 0, j = 1, 2, \dots, p$ (Variabel prediktor ke-*j* tidak berpengaruh signifikan terhadap variabel respon).

$H_1 : \beta_j \neq 0, j = 1, 2, \dots, p$ (Variabel prediktor ke-*j* berpengaruh signifikan terhadap variabel respon).

Statistik uji yang digunakan adalah uji *Wald* dengan rumus

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.13)$$

dengan $SE(\hat{\beta}_j)$ merupakan $\sqrt{\text{var}(\hat{\beta}_j)}$ dan $\hat{\beta}_j$ merupakan nilai koefisien dengan variabel independen ke-*j*.

Pada tingkat kepercayaan α dan derajat bebas (*df*) yang merupakan banyaknya variabel prediktor, H_0 ditolak jika nilai $W > Z_{\alpha/2}$. Dengan menolak H_0 maka parameter β berpengaruh signifikan terhadap variabel respon secara parsial.

2.5. Uji Kesesuaian Model (Uji *Goodness of Fit*).

Uji kesesuaian model atau uji *Goodness of Fit* merupakan pengujian yang dilakukan untuk melihat apakah model yang dihasilkan sudah layak. Pengujian ini menggunakan uji *Pearson Chi-square* Hosmer and Lemeshow [10]. Uji kesesuaian model memiliki hipotesis sebagai berikut:

$H_0 : \pi_k = 0, k = 1, 2, \dots, g$ (Model sesuai atau tidak ada perbedaan antara hasil observasi dengan hasil prediksi).

$H_1 : \exists \pi_k \neq 0, k = 1, 2, \dots, g$ (Minimal terdapat satu model yang tidak sesuai atau ada perbedaan antara hasil observasi dengan hasil prediksi).

Statistik uji yang digunakan dengan melihat nilai *Pearson Chi-square*. Pada tingkat kepercayaan α dan derajat bebas (*df*) yang merupakan banyaknya variabel prediktor, maka pengambilan keputusan dilakukan dengan cara menolak H_0 apabila nilai $\hat{C} > \chi_{\alpha, df}^2$. Dengan menolak H_0 maka model sesuai atau tidak terdapat perbedaan antara hasil observasi dengan hasil prediksi.

2.6. Odds Ratio.

Regresi logistik juga menghasilkan rasio peluang (*odds ratio*) terkait dengan nilai pada variabel dependen. Peluang (*odds*) dari suatu kejadian diartikan sebagai probabilitas hasil yang muncul yang dibagi dengan probabilitas suatu kejadian tidak terjadi Agresti [1]. Secara umum rasio peluang (*odds ratio*) merupakan kumpulan peluang yang dibagi oleh peluang lainnya. Rasio peluang bagi prediktor diartikan sebagai jumlah relatif dimana peluang hasil meningkat (rasio peluang > 1) atau turun (rasio peluang < 1) ketika nilai variabel prediktor meningkat sebesar 1 unit.

Jika peluang kejadian yang terjadi dari setiap grup adalah p (grup pertama) dan q (grup kedua) maka *odds ratio* dirumuskan sebagai

$$OR = \frac{\frac{p}{1-p}}{\frac{q}{1-q}} = \frac{p(1-q)}{q(1-p)}. \quad (2.14)$$

2.7. Ketepatan Klasifikasi (*Confusion Matrix*).

Confusion matrix merupakan matriks yang menampilkan prediksi klasifikasi dan klasifikasi yang aktual. *Confusion matrix* berukuran $L \times L$, dimana L adalah jumlah label klasifikasi yang berbeda [7].

Aktual	Prediksi	
	Positif	Negatif
Positif	a	b
Negatif	c	d

TABEL 1. *Confusion Matrix* untuk $L = 2$

Nilai akurasi didapatkan dari:

$$\text{Akurasi} = \frac{a + d}{a + b + c + d} \quad (2.15)$$

Ketepatan klasifikasi model digunakan apakah data diklasifikasikan dengan benar atau tidak Agresti [2]. Akurasi merupakan perhitungan untuk mengetahui seberapa bagus model memprediksi data dengan benar. Pada tabel *confusion matrix* dapat juga diketahui hasil *error rate* dari algoritma klasifikasi yang digunakan.

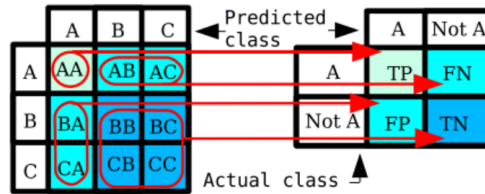


FIGURE 1. *Extended Confusion Matrix 3x3*

Gambar 1 menunjukkan perubahan dari *extended confusion matrix* berukuran 3x3 menjadi berukuran 2x2, dengan kelas ‘A’ sebagai kelas positif dan kelas ‘Not A’ sebagai kelas negatif.

Ketepatan prediksi dapat diketahui tingkat ketelitiannya dengan menghitung hasil presisi, sensitivitas (*recall*), dan spesifisitas. Presisi merupakan perhitungan untuk mengetahui tingkat ketelitian atau ketepatan dalam pengklasifikasian. Nilai presisi, sensitivitas (*recall*), dan spesifisitas dapat dihitung sebagai berikut [12]:

$$\text{Presisi} = \frac{TP}{TP + FP} \tag{2.16}$$

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \tag{2.17}$$

$$\text{Spesifisitas} = \frac{TN}{TN + FP} \tag{2.18}$$

Keterangan :

TP : jumlah observasi positif yang tepat diklasifikasikan sebagai kelas positif

FP : jumlah observasi Positif yang diklasifikasikan sebagai kelas negatif

FN : jumlah observasi negatif yang diklasifikasikan sebagai kelas positif

TN : jumlah observasi negatif yang tepat diklasifikasikan sebagai kelas negatif

3. METODOLOGI PENELITIAN

Pada bagian ini akan dijelaskan secara lebih rinci masing-masing tahapan yang dilakukan untuk penelitian ini.

3.1. Identifikasi Masalah.

Pada tahap ini dilakukan penggalian dan analisa studi kasus yang akan menjadi pokok permasalahan serta metodologi yang akan digunakan dalam menyelesaikan permasalahan tersebut. Pada penelitian ini topik yang menjadi pokok permasalahan adalah

tentang perbandingan analisis Regresi Logistik model Logit, Probit, Cauchit, dan Clog-log untuk menentukan faktor yang mempengaruhi indeks pembangunan manusia (IPM) di Pulau Jawa.

3.2. Studi Literatur.

Studi literatur dilakukan melalui berbagai sumber diantaranya buku, jurnal, artikel dan beberapa sumber terkait lainnya. Tujuan studi literatur dilakukan adalah untuk mendapatkan informasi-informasi yang *reliable*, terkini dan dapat dipertanggung jawabkan. Dari tahapan ini, didapatkan pemahaman konsep dan *knowledge* dari beberapa penelitian sebelumnya, dan menentukan metode yang paling tepat untuk diterapkan dalam penelitian ini. Adapun studi literatur yang digunakan adalah konsep Regresi Logistik, model Logit, Probit, Cauchit, dan Clog-log serta metode evaluasi model.

3.3. Persiapan Data.

Data yang akan digunakan dalam penelitian penelitian ini diperoleh dari Badan Pusat Statistik Provinsi Banten, DKI Jakarta, Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi D.I Yogyakarta, dan Provinsi Jawa Timur. Data yang dilibatkan adalah data indeks pembangunan nasional di setiap kabupaten di Pulau Jawa. Variabel dependen yang digunakan adalah Usia Harapan Hidup (X_1), Harapan Lama Sekolah (X_2), Rata-rata lama sekolah (X_3), Persentase penduduk miskin (X_4), Tingkat Pengangguran Terbuka (X_5), dan Persentase rumah tangga dengan akses air minum layak (X_6). Capaian pembangunan manusia di suatu wilayah pada waktu tertentu dapat dikelompokkan ke dalam tiga kelompok berdasarkan status capaiannya. Adapun pengelompokkan tersebut sebagai berikut: (i) tinggi : $IPM \geq 80$; (ii) sedang: $70 \leq IPM < 80$; (iii) Rendah : $IPM < 70$.

3.4. Pengujian Signifikansi Parameter.

Sebelum membuat permodelan regresi dilakukan pengujian signifikansi parameter. Pengujian ini dilakukan untuk mengetahui variabel independen yang signifikan terhadap indeks pembangunan manusia (IPM). Pengujian ini dilakukan dengan 2 cara yaitu uji signifikansi parameter secara keseluruhan dan uji signifikansi parameter secara parsial.

3.4.1. *Uji Signifikansi Parameter Secara Keseluruhan (Uji Rasio Likelihood)*. Uji ini dilakukan untuk memeriksa kemaknaan koefisien secara keseluruhan (keberartian model). Pengujian ini dilihat berdasarkan nilai uji statistik G. Uji ini membandingkan model lengkap beserta variabel independen dengan model yang hanya menggunakan konstanta saja tanpa variabel independen.

3.4.2. *Uji Signifikansi Parameter Parsial (Uji Wald)*. Uji ini dilakukan untuk mengetahui signifikansi parameter terhadap variabel dependen. Pengujian ini dilihat berdasarkan nilai uji Wald dimana menguji keberartian koefisien β secara parsial dengan membandingkan dugaan β dengan penduga pada standar erornya.

3.5. Pemilihan Model Terbaik.

Pada tahap ini dilakukan pemilihan model terbaik yang dihasilkan antara model Logit, Probit, Cauchit, dan Clog-log dengan melihat nilai akurasi model, AIC, BIC, dan nilai Pseudo R^2 .

3.6. Pengujian Kesesuaian Model (Uji *Goodness of Fit*).

Uji ini dilakukan untuk mengetahui apakah terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi serta melihat kelayakan model yang dipilih.

3.7. Menghitung Ketepatan Klasifikasi.

Pada tahap ini dihitung ketepatan klasifikasi untuk mengetahui hasil observasi yang secara tepat diklasifikasikan dengan hasil prediksi. Perhitungan ketepatan klasifikasi model ditunjukkan dengan nilai akurasi, presisi, sensitivitas (*recall*), spesifisitas, dan tingkat kesalahan (*error rate*). Semakin kecil tingkat kesalahan, maka semakin sedikit eror yang dihasilkan sehingga menunjukkan model tersebut sudah baik.

4. ANALISIS DAN PEMBAHASAN

4.1. Gambaran Umum Karakteristik IPM dan Variabel-variabel yang Mempengaruhi IPM di Pulau Jawa.

Karakteristik data Indeks Pembangunan Manusia (IPM) dan variabel-variabel yang mempengaruhi IPM di Pulau Jawa menggunakan analisis statistik deskriptif disajikan pada Tabel 2 dan Tabel 3.

IPM (Y)	Frekuensi	Persentase (%)
Rendah	19	15,97%
Sedang	77	64,70%
Tinggi	23	19,33%

TABEL 2. Statistika Deskriptif IPM di Pulau Jawa Tahun 2023

Variabel	Rata-rata	Varian	Standar Deviasi	Minimum	Maksimum
X_1	73,70	4,30	2,07	67,60	77,93
X_2	13,29	1,06	1,03	11,80	17,62
X_3	8,70	2,64	1,62	5,07	12,11
X_4	9,39	14,45	3,80	2,38	21,76
X_5	5,83	4,17	2,04	1,52	12,16
X_6	94,75	32,14	5,67	75,64	100,00

TABEL 3. Statistik Deskriptif Variabel-variabel yang Mempengaruhi IPM di Pulau Jawa Tahun 2023

Variabel respon yang akan dijelaskan adalah IPM. Berdasarkan Tabel 2, data IPM yang digunakan dalam penelitian ini sebanyak 119 Kabupaten/Kota, terdiri dari 19 Kabupaten/Kota berstatus IPM tingkat rendah, 77 Kabupaten/Kota berstatus IPM tingkat sedang, dan 23 Kabupaten/Kota berstatus IPM tingkat tinggi. Terdapat enam variabel yang mempengaruhi IPM di Pulau Jawa. Berdasarkan Tabel 3 dapat disimpulkan menjadi beberapa poin, yaitu:

- Variabel Usia Harapan Hidup (X_1) memiliki nilai terendah sebesar 67,60 dan nilai tertinggi sebesar 77,93 dengan nilai rata-ratanya sebesar 73,70 dan standar deviasi (tingkat sebaran datanya) sebesar 2,07.
- Variabel Harapan Lama Sekolah (X_2) memiliki nilai terendah sebesar 11,80 dan nilai tertinggi sebesar 17,62 dengan nilai rata-ratanya sebesar 13,29 dan standar deviasi (tingkat sebaran datanya) sebesar 1,03.
- Variabel Rata-rata Lama Sekolah (X_3) memiliki nilai terendah sebesar 5,07 dan nilai tertinggi sebesar 12,11 dengan nilai rata-ratanya sebesar 8,70 dan standar deviasi (tingkat sebaran datanya) sebesar 1,62.
- Variabel Persentase penduduk miskin (X_4) memiliki nilai terendah sebesar 2,38 persen dan nilai tertinggi sebesar 21,76 persen dengan nilai rata-ratanya sebesar 9,39 dan standar deviasi (tingkat sebaran datanya) sebesar 3,80.
- Variabel Tingkat pengangguran terbuka (X_5) memiliki nilai terendah sebesar 1,52 dan nilai tertinggi sebesar 12,16 dengan nilai rata-ratanya sebesar 5,83 dan standar deviasi (tingkat sebaran datanya) sebesar 2,04.
- Variabel Persentase rumah tangga dengan akses air minum layak (X_6) memiliki nilai terendah sebesar 75,64 persen dan nilai tertinggi sebesar 100 persen dengan nilai rata-ratanya sebesar 94,75 dan standar deviasi (tingkat sebaran datanya) sebesar 5,67.

Hasil analisis deskriptif pada masing-masing variabel faktor yang mempengaruhi tinggi atau rendahnya IPM disajikan dalam Gambar 2.

4.2. Pemodelan Data dengan Empat Link Fungsi Matematika pada *Generalized Linear Models*.

Estimasi parameter digunakan untuk membuat pemodelan pengaruh variabel-variabel prediktor terhadap variabel respon. Berikut ini estimasi parameter menggunakan empat model link fungsi matematika pada *Generalized Linear Models*.

Berdasarkan estimasi parameter yang dijelaskan pada Tabel 4, maka model awal regresi ordinal yang terbentuk adalah sebagai berikut:

(1) Logit

$$g_1(x) = 71,008 + 0,159X_1 + 1,830X_2 + 3,105X_3 + 0,087X_4 - 0,304X_5 + 0,165X_6$$

$$g_2(x) = 84,858 + 0,159X_1 + 1,830X_2 + 3,105X_3 + 0,087X_4 - 0,304X_5 + 0,165X_6$$

Perbandingan Empat Fungsi GLM untuk Menganalisis Faktor-faktor yang Mempengaruhi IPM¹⁷

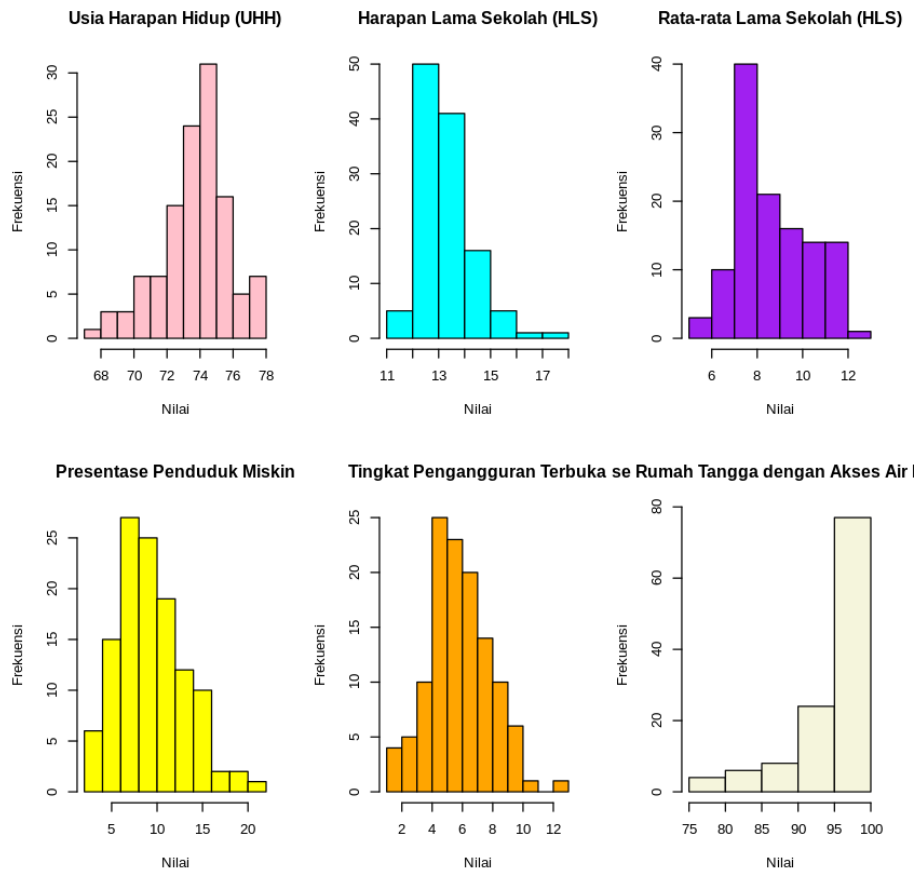


FIGURE 2. Grafik Variabel Prediktor

Koefisien	Logit	Probit	Cauchit	Clog-log
Konstanta (1)	71,008	42,687	457,973	48,084
Konstanta (2)	84,858	50,741	519,992	58,744
X_1	0,159	0,106	2,477	0,056
X_2	1,830	1,072	10,754	1,223
X_3	3,105	1,810	9,282	2,571
X_4	0,087	0,058	-0,640	0,109
X_5	-0,304	-0,179	-1,077	-0,234
X_6	0,165	0,098	0,954	0,120

TABEL 4. Estimasi Parameter Empat Model Link Fungsi Matematika pada *Generalized Linear Models*

(2) Probit

$$g_1(x) = 42,687 + 0,106X_1 + 1,072X_2 + 1,810X_3 + 0,058X_4 - 0,179X_5 + 0,098X_6$$

$$g_2(x) = 50,741 + 0,106X_1 + 1,072X_2 + 1,810X_3 + 0,058X_4 - 0,179X_5 + 0,098X_6$$

(3) Cauchit

$$g_1(x) = 457,973 + 2,477X_1 + 10,754X_2 + 9,282X_3 - 0,640X_4 - 1,077X_5 + 0,954X_6$$

$$g_2(x) = 519,992 + 2,477X_1 + 10,754X_2 + 9,282X_3 - 0,640X_4 - 1,077X_5 + 0,954X_6$$

(4) Complementary Log-log (Clog-log)

$$g_1(x) = 48,084 + 0,056X_1 + 1,223X_2 + 2,571X_3 + 0,109X_4 - 0,234X_5 + 0,120X_6$$

$$g_2(x) = 58,744 + 0,056X_1 + 1,223X_2 + 2,571X_3 + 0,109X_4 - 0,234X_5 + 0,120X_6$$

4.3. Uji Signifikansi Parameter Secara Keseluruhan (Uji Rasio Likelihood).

Uji Rasio Likelihood digunakan untuk mengetahui apakah variabel-variabel prediktor secara bersama-sama berpengaruh terhadap model IPM di Pulau Jawa. Pengujian ini dilakukan dengan melihat nilai statistik G yang dibandingkan dengan nilai *Chi-Square* tabel atau dengan melihat nilai *p-value*. Berdasarkan Tabel 4 dilakukan pengujian hipotesis sebagai berikut:

Model		G	Nilai Tabel Chi-Square	Keputusan
Logit	Awal	156,850	12,592	Tolak H_0
	Akhir	151,947	7,815	Tolak H_0
Probit	Awal	157,667	12,952	Tolak H_0
	Akhir	152,108	7,815	Tolak H_0
Cauchit	Awal	152,753	12,952	Tolak H_0
	Akhir	152,753	12,952	Tolak H_0
Clog-log	Awal	157,152	12,952	Tolak H_0
	Akhir	151,900	7,815	Tolak H_0

TABEL 5. Uji Rasio Likelihood Empat Model Link Fungsi Matematika pada *Generalized Linear Models*

(1) Hipotesis

$H_0 : \beta_j = 0, j = 1, 2, 3, 4, 5, 6$ (Variabel prediktor tidak berpengaruh signifikan terhadap variabel respon)

$H_1 : \exists \beta_j \neq 0, j = 1, 2, 3, 4, 5, 6$ (Minimal terdapat satu variabel prediktor tidak berpengaruh signifikan terhadap variabel respon)

(2) Taraf Signifikansi

$\alpha = 5\%$

(3) Statistik Uji

$$G = -2 \ln \left[\frac{\text{likelihood (model B)}}{\text{likelihood (model A)}} \right]$$

dengan

Model B : fungsi likelihood tanpa variabel bebas.

Model A : fungsi likelihood dengan variabel bebas.

(4) Daerah Kritis

Tolak H_0 jika nilai $p\text{-value} < \alpha$ atau jika nilai $G > \chi^2_{\alpha,p}$

(5) Keputusan

Tolak H_0 karena nilai $G > \chi^2_{\alpha,p}$

(6) Kesimpulan

Berdasarkan pada Tabel 5 dapat dilihat untuk model awal dan model akhir dari masing-masing model, uji Rasio Likelihood bahwa tolak H_0 karena nilai $G > \chi^2_{(0,05;p)}$ sehingga dapat disimpulkan bahwa paling sedikit ada satu variabel prediktor memberi pengaruh signifikan terhadap variabel respon.

4.4. Uji Signifikansi Parameter Parsial (Uji Wald).

Uji Wald digunakan untuk mengetahui apakah masing-masing variabel respon mempengaruhi model IPM di Pulau Jawa. Pengujian ini dilakukan dengan melihat nilai statistik uji Wald yang dibandingkan dengan nilai t tabel atau dengan melihat nilai $p\text{-value}$. Output uji Wald ditampilkan pada Tabel 6.

Koefisien		Logit		Probit		Cauchit		Clog-log	
		Awal	Akhir	Awal	Akhir	Awal	Akhir	Awal	Akhir
X_1	Wald	0,760		0,911		11,145	11,145	0,413	
	$p\text{-value}$	0,447		0,362		0,000	0,000	0,679	
X_2	Wald	2,836	3,482	2,942	3,641	12,357	12,357	2,803	3,698
	$p\text{-value}$	0,005	0,000	0,003	0,000	0,000	0,000	0,005	0,000
X_3	Wald	3,693	4,224	3,827	4,406	10,608	10,608	3,934	4,392
	$p\text{-value}$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
X_4	Wald	0,527		0,616		-2,289	-2,289	0,938	
	$p\text{-value}$	0,598		0,538		0,022	0,022	0,348	
X_5	Wald	-1,568		-1,664		-3,333	-3,333	-1,518	
	$p\text{-value}$	0,117		0,096		0,001	0,001	0,129	
X_6	Wald	2,377	2,587	2,425	2,566	12,185	12,185	2,474	2,418
	$p\text{-value}$	0,017	0,010	0,015	0,010	0,000	0,000	0,013	0,016

TABEL 6. Uji Wald Empat Model Link Fungsi Matematika pada *Generalized Linear Models*

Berdasarkan output pada Tabel 6 dilakukan pengujian hipotesis sebagai berikut:

(1) Hipotesis

$H_0 : \beta_j = 0, j = 1, 2, 3, 4, 5, 6$ (Variabel prediktor ke- j tidak berpengaruh signifikan terhadap variabel respon)

$H_1 : \beta_j \neq 0, j = 1, 2, 3, 4, 5, 6$ (Variabel prediktor ke- j berpengaruh signifikan terhadap variabel respon)

(2) Taraf Signifikansi

$\alpha = 5\%$

(3) Statistik Uji

$$W_k = \left[\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right]^2$$

dengan $\hat{\beta}_k$ merupakan penaksir parameter β_k dan standar eror $\hat{\beta}_k$ diperoleh dari $SE(\hat{\beta}_k) = \sqrt{\hat{v}ar(\hat{\beta}_k)}$

(4) Daerah Kritis

Tolak H_0 jika nilai $p\text{-value} < \alpha$ atau jika nilai $W^2 > \chi_{\alpha,1}^2$

(5) Kesimpulan

Berdasarkan uji Wald pada empat Model Link Fungsi Matematika pada *Generalized Linear Models* yang dijelaskan pada Tabel 6 didapatkan bahwa variabel yang signifikan terhadap model Logit, Probit, dan Clog-log, yaitu Harapan Lama Sekolah (X_2), Rata-rata lama sekolah (X_3), dan Persentase rumah tangga dengan akses air minum layak (X_6). Sedangkan variabel yang signifikan terhadap model Cauchit yaitu Usia Harapan Hidup (X_1), Harapan Lama Sekolah (X_2), Rata-rata lama sekolah (X_3), Persentase penduduk miskin (X_4), Tingkat Pengangguran Terbuka (X_5), dan Persentase rumah tangga dengan akses air minum layak (X_6).

4.5. Perbandingan Empat Model Link Fungsi Matematika pada *Generalized Linear Models*.

Model	Ketepatan Klasifikasi	AIC	BIC	Pseudo R^2
Regresi Logit Ordinal	86,32%	70,417	83,186	0,661
Regresi Probit Ordinal	86,32%	70,256	83,025	0,662
Regresi Cauchit Ordinal	92,63%	77,131	97,562	0,659
Regresi Clog-log Ordinal	86,32%	70,464	83,233	0,661

TABEL 7. Perbandingan Empat Model Fungsi Link pada Regresi Ordinal

Berdasarkan Tabel 7 diperoleh ketepatan klasifikasi yang dihasilkan oleh regresi Logit ordinal memiliki hasil yang sama besar dengan regresi Probit ordinal dan regresi Complementary Log-log ordinal. Nilai AIC dan BIC pada regresi Probit ordinal memiliki hasil paling kecil dibandingkan ketiga model lainnya. Sedangkan nilai Pseudo R^2 pada regresi Probit ordinal memiliki hasil paling besar dibandingkan ketiga model lainnya. Sehingga dapat disimpulkan bahwa model regresi Probit ordinal merupakan model terbaik untuk menganalisis faktor-faktor yang mempengaruhi IPM di Pulau Jawa tahun 2023.

4.6. Faktor-faktor yang Mempengaruhi IPM di Pulau Jawa Tahun 2023.

Setelah melakukan perbandingan Empat Model Link Fungsi Matematika pada *Generalized Linear Models*, didapatkan model regresi Probit ordinal merupakan model terbaik untuk menganalisis faktor-faktor yang mempengaruhi IPM di Pulau Jawa tahun 2023. Oleh karena itu, untuk analisis lebih lanjut terkait variabel-variabel yang mempengaruhi IPM akan menggunakan model regresi Probit ordinal sebagai berikut:

$$P(Y_i \leq 1|X_i) = \phi(34,160 + 1,201X_2 + 1,534X_3 + 0,089X_6)$$

$$P(Y_i \leq 2|X_i) = \phi(41,640 + 1,201X_2 + 1,534X_3 + 0,089X_6)$$

Bentuk persamaan di atas merupakan bentuk persamaan non linear, maka perlu ditransformasikan ke dalam bentuk Probit.

$$g_1(x) = 34,160 + 1,201X_2 + 1,534X_3 + 0,089X_6$$

$$g_2(x) = 41,640 + 1,201X_2 + 1,534X_3 + 0,089X_6$$

Berdasarkan persamaan di atas dapat disimpulkan faktor-faktor yang mempengaruhi IPM di Pulau Jawa tahun 2023 adalah Harapan Lama Sekolah, Rata-rata lama sekolah, dan Persentase rumah tangga dengan akses air minum layak. Nilai koefisien regresi untuk variabel X_2 sebesar 1,201 artinya jika variabel Harapan Lama Sekolah mengalami kenaikan sebesar 1, maka variabel IPM akan mengalami kenaikan sebesar 1,201. Nilai koefisien regresi untuk variabel X_3 sebesar 1,534 artinya jika variabel Rata-rata Lama Sekolah mengalami kenaikan sebesar 1, maka variabel IPM akan mengalami kenaikan sebesar 1,534. Nilai koefisien regresi untuk variabel X_6 sebesar 0,089 artinya jika variabel persentase rumah tangga dengan akses air minum layak mengalami kenaikan sebesar 1, maka variabel IPM akan mengalami kenaikan sebesar 0,089.

4.7. Resiko Faktor-faktor yang Mempengaruhi IPM.

Odds Ratio adalah ukuran yang digunakan untuk mengevaluasi hubungan antara dua kejadian atau faktor dalam statistik. Berdasarkan uji signifikan parameter secara keseluruhan maupun uji signifikan parameter parsial diketahui bahwa variabel prediktor yang mempengaruhi IPM secara signifikan adalah Harapan Lama Sekolah (X_2), Rata-rata lama sekolah (X_3), dan Persentase rumah tangga dengan akses air minum layak (X_6) dengan nilai $Exp(\beta)$ yang dihasilkan pada Tabel 7.

Variabel	$Exp(\beta)$
X_2	3,325
X_3	4,638
X_6	1,093

TABEL 8. Nilai *Odds Ratio*

Berdasarkan nilai *odds ratio* pada Tabel 8, maka model regresi ordinal dapat diinterpretasikan sebagai berikut:

- Setiap kenaikan satu satuan Harapan Lama Sekolah (X_2) akan memberikan pengaruh sebesar 3,325 kali lebih besar untuk meningkatkan nilai IPM sebesar satu.
- Setiap kenaikan satu satuan Rata-rata Lama Sekolah (X_3) akan memberikan pengaruh sebesar 4,638 kali lebih besar untuk meningkatkan nilai IPM sebesar satu.
- Setiap kenaikan satu satuan Persentase rumah tangga dengan akses air minum layak (X_6) akan memberikan pengaruh sebesar 1,093 kali lebih besar untuk meningkatkan nilai IPM sebesar satu.

4.8. Uji Kesesuaian Model (Uji *Goodness of Fit*).

Uji kesesuaian model dilakukan untuk mengevaluasi model regresi Probit ordinal yang telah terbentuk.

D	Derajat Bebas	Nilai Tabel Chi-Square	<i>p-value</i>
60,256	90	113,145	0,993

TABEL 9. Uji Kesesuaian Model Regresi Probit Ordinal

Berdasarkan output pada Tabel 9 dilakukan pengujian hipotesis sebagai berikut:

- (1) Hipotesis
 $H_0 : \pi_k = 0, k = 1, 2, \dots, g$ (Model sesuai atau tidak ada perbedaan antara hasil observasi dengan hasil prediksi)
 $H_1 : \exists \pi_k \neq 0, k = 1, 2, \dots, g$ (Minimal terdapat satu model yang tidak sesuai atau ada perbedaan antara hasil observasi dengan hasil prediksi)
- (2) Taraf Signifikansi
 $\alpha = 5\%$
- (3) Statistik Uji

$$D = -2 \sum_{i=1}^n \sum_{g=1}^G \left[y_{ig} \ln \left(\frac{\hat{\pi}_{ig}}{y_{ig}} \right) \right]$$

- (4) Daerah Kritis
Tolak H_0 jika nilai *p-value* $< \alpha$ atau jika nilai $D > \chi_{df, \alpha}^2$
- (5) Keputusan
Gagal tolak H_0 karena Deviance = 60,256 $< \chi_{(0,05;90)}^2 = 113,145$ atau nilai *p-value* = 0,993 $> \alpha = 0,05$
- (6) Kesimpulan
Berdasarkan taraf signifikansi 5%, gagal tolak H_0 sehingga dapat disimpulkan bahwa model sesuai atau tidak ada perbedaan antara hasil observasi dengan hasil prediksi.

Prediksi	Observasi			Jumlah
	Rendah	Sedang	Tinggi	
Rendah	12	3	0	15
Sedang	6	53	2	61
Tinggi	0	2	17	19
Jumlah	18	58	19	95

TABEL 10. Confusion Matrix Model Regresi Probit Ordinal

4.9. Ketepatan Klasifikasi (*Confusion Matrix*).

Berdasarkan Tabel 10 dapat dihitung ketepatan klasifikasi yaitu akurasi dan tingkat kesalahan (*error rate*) sebagai berikut:

$$\text{Akurasi} = \frac{\text{jumlah prediksi benar}}{\text{jumlah prediksi}} = \frac{12 + 53 + 17}{95} = \frac{82}{95} = 0,86316$$

$$\text{Error rate} = \frac{\text{jumlah prediksi salah}}{\text{jumlah prediksi}} = \frac{3 + 0 + 6 + 2 + 0 + 2}{95} = \frac{13}{95} = 0,13684$$

Untuk ketepatan klasifikasi presisi, sensitivitas (*recall*), dan spesifisitas dihitung untuk masing-masing kelas sehingga diperoleh:

(1) IPM Rendah

$$\text{Presisi}_{\text{rendah}} = \frac{TP}{TP + FP} = \frac{12}{12 + (6 + 0)} = \frac{12}{18} = 0,6667$$

$$\text{Sensitivitas}_{\text{rendah}} = \frac{TP}{TP + FN} = \frac{12}{12 + (3 + 0)} = \frac{12}{15} = 0,8$$

$$\text{Spesifisitas}_{\text{rendah}} = \frac{TN}{TN + FP} = \frac{53 + 2 + 2 + 17}{(53 + 2 + 2 + 17) + (6 + 0)} = \frac{74}{80} = 0,925$$

(2) IPM Sedang

$$\text{Presisi}_{\text{sedang}} = \frac{TP}{TP + FP} = \frac{53}{53 + (3 + 2)} = \frac{53}{58} = 0,9138$$

$$\text{Sensitivitas}_{\text{sedang}} = \frac{TP}{TP + FN} = \frac{53}{53 + (6 + 2)} = \frac{53}{61} = 0,8689$$

$$\text{Spesifisitas}_{\text{sedang}} = \frac{TN}{TN + FP} = \frac{12 + 0 + 0 + 17}{(12 + 0 + 0 + 17) + (3 + 2)} = \frac{29}{34} = 0,8529$$

(3) IPM Tinggi

$$\text{Presisi}_{\text{tinggi}} = \frac{TP}{TP + FP} = \frac{17}{17 + (0 + 2)} = \frac{17}{19} = 0,8947$$

$$\text{Sensitivitas}_{\text{tinggi}} = \frac{TP}{TP + FN} = \frac{17}{17 + (0 + 2)} = \frac{17}{19} = 0,8947$$

$$\text{Spesifisitas}_{\text{tinggi}} = \frac{TN}{TN + FP} = \frac{12 + 3 + 6 + 53}{(12 + 3 + 6 + 53) + (0 + 2)} = \frac{74}{76} = 0,9737$$

Sehingga didapatkan ketepatan klasifikasi untuk IPM kategori rendah yaitu presisi 66,67%, sensitivitas (*recall*) 80%, dan spesifisitas 92,5%. Untuk IPM kategori sedang yaitu presisi 91,38%, sensitivitas (*recall*) 86,89%, dan spesifisitas 85,29%. Untuk IPM kategori tinggi yaitu presisi 89,47%, sensitivitas (*recall*) 89,47%, dan spesifisitas 97,37%. Serta didapatkan ketepatan klasifikasi untuk model regresi Probit ordinal yaitu akurasi 86,316% dan tingkat kesalahan (*error rate*) 13,684%.

4.10. Perbandingan Data *Training* dan Data *Testing*.

Pembagian data menjadi data *training* (data pelatihan) dan data *testing* (data pengujian) dilakukan untuk mengujur kinerja dari suatu model. Data *training* digunakan untuk melatih model, dimana pola dari data ini untuk membuat prediksi atau keputusan. Sedangkan data *testing* digunakan untuk menguji kinerja model yang telah dilatih. Penelitian kali ini menggunakan perbandingan 80-20, dimana 80% data untuk pelatihan dan 20% untuk pengujian.

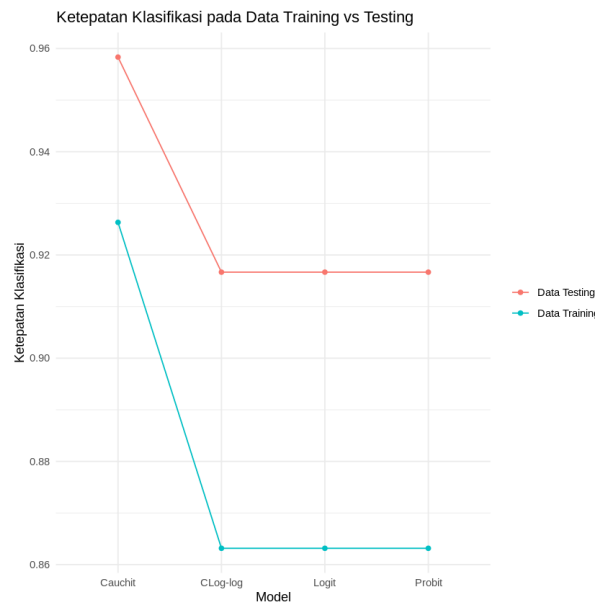


FIGURE 3. Perbandingan Data *Training* vs Data *Testing* pada Ketepatan Klasifikasi

Berdasarkan Gambar 3, diketahui bahwa ketepatan klasifikasi keempat model menggunakan data *testing* lebih baik dari data *training*. Berdasarkan perbandingan keempat model link fungsi matematika pada GLM, model regresi Probit ordinal merupakan model terbaik untuk menganalisis faktor-faktor yang mempengaruhi IPM di Pulau Jawa tahun 2023. Berikut perbandingan data *training* dan data *testing* pada model regresi Probit ordinal yang dirangkuman dalam Tabel 11.

Data	Ketepatan Klasifikasi	AIC	BIC	Pseudo R^2
<i>Training</i>	86,32%	70,256	83,025	0,662
<i>Testing</i>	91,67%	16,000	25,425	0,999

TABEL 11. Perbandingan Data *Training* dan Data *Testing* pada Model Regresi Probit Ordinal

5. PENUTUP

Kesimpulan. Berdasarkan hasil dan pembahasan perbandingan empat model link fungsi matematika pada *Generalized Linear Models* (GLM) untuk menganalisis faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) di Pulau Jawa tahun 2023, diperoleh kesimpulan sebagai berikut:

- (1) Model regresi Probit ordinal merupakan model terbaik untuk menganalisis faktor-faktor yang mempengaruhi IPM di Pulau Jawa.
- (2) Indeks Pembangunan Manusia (IPM) di Pulau Jawa dipengaruhi oleh Harapan Lama Sekolah (HLS), Rata-rata lama sekolah (RLS), dan Persentase rumah tangga dengan akses air minum layak.
- (3) Ketepatan klasifikasi untuk IPM kategori rendah yaitu presisi 66,67%, sensitivitas (*recall*) 80%, dan spesifisitas 92,5%. Untuk IPM kategori sedang yaitu presisi 91,38%, sensitivitas (*recall*) 86,89%, dan spesifisitas 85,29%. Untuk IPM kategori tinggi yaitu presisi 89,47%, sensitivitas (*recall*) 89,47%, dan spesifisitas 97,37%.

References

- [1] Agresti, A, *Categorical Data Analysis*. New York: John Wiley and Sons, 1990.
- [2] Agresti, A, *Categorical Data Analysis*. New York: John Wiley and Sons, 2002.
- [3] Arafat, L. Rindayati, W. Sahara, *Faktor-faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Kalimantan Tengah*. Bogor: Jurnal Ekonomi dan Kebijakan Pembangunan, hlm 140-158, 2018.
- [4] BPS, *Indeks Pembangunan Manusia*. Jakarta : Badan Pusat Statistik, 2008.
- [5] BPS, *Konsep Indeks Pembangunan Manusia*. Jakarta : Badan Pusat Statistik, 2014.
- [6] BPS Analisis Kualitas Pembangunan Manusia Provinsi Jawa Tengah 2023. Semarang : Badan Pusat Statistik, 2023.
- [7] Felkin, M. *Comparing Classification Results between N-ary and Binary Problems*. Quality Measures in Data Mining, hlm 277-301, 2007.
- [8] Gujarati, D.N, *Ekonometrika Dasar : Edisi Keenam*. Jakarta: Erlangga, 2003.
- [9] HDR, H, *Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. New York: United Nations Development Programme (UNDP), 2014.
- [10] Hosmer and Lemeshow, *Applied Logistic Regression*. New York: John Wiley and Sons, 2000.
- [11] Rohmawati, N. M, *Analisis Faktor-faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Timur (pada 38 Kabupaten/Kota)*. Malang: FEB UB, 2021.
- [12] Witten, I. H., & Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques (2nd ed.)*. 2005.
- [13] Haller, G., *Chaos Near Resonance*, in : *Applied Mathematical Sciences*, vol 138, Springer, New York, 1999.

- [14] Nayfeh, S. A. dan Nayfeh, A. H., Nonlinear Interaction between Two Widely Spaced Modes-external Excitation, *Int. J. Bifurcat. Chaos* **3** (1993), 417-427.

ROSSA FITRIA HALIM: Universitas Gadjah Mada.
E-mails: rossafitriahalim@mail.ugm.ac.id

RENANTA DZAKIYA NAFALANA: Universitas Gadjah Mada.
E-mails: renantadzakiyanafalana@mail.ugm.ac.id

ADITYO WAHYU SAPUTRO: Universitas Gadjah Mada.
E-mails: adityowahyusaputro@mail.ugm.ac.id

MOHAMMAD AGUS KHOLILURROHMAN: Universitas Gadjah Mada.
E-mails: mohammadaguskhohilurrohman@mail.ugm.ac.id