

Received: 30 April 2023

Revised: 16 June 2023

Accepted: 27 June 2023

Published: 30 June 2023

Pemodelan Regresi Logistik Ordinal *Backward* dengan Imputasi *K-Nearest Neighbour* pada Indeks Pembangunan Manusia di Indonesia Tahun 2021

Muftih Alwi Aliu^{1, a)}, Anwar Fitrianto^{2, b)}, Erfiani^{3, c)}, Indahwati^{4, d)}, Khusnia N.K.^{5, e)}

¹*Prodi Statistika dan Sains Data, IPB University.*

Email: ^{a)}muftihalwialiu@apps.ipb.ac.id, ^{b)}anwarstat@gmail.com, ^{c)}erfiani@apps.ipb.ac.id,
^{d)}indah.stk@gmail.com, ^{e)}khusniank@gmail.com

Abstract

The human development index (HDI) is one of the important things to note in Indonesia today. The growth of HDI in Indonesia in 2021 is not evenly distributed in all regencies/cities and has high disparities. This study aims to find out the description of HDI data, get the best model to determine the factors that significantly affect the HDI of regencies/cities in Indonesia in 2021 and identify the classification accuracy results of the best model. The independent variables used in this study are average years of schooling, open unemployment rate, population growth rate, population density, percentage of poor people and sex ratio. The independent variables in this study contained missing values, so they were handled using k-nearest neighbour (KNN) imputation and continued modelling using ordinal logistic regression using the backward elimination technique to obtain significant factors. The results showed that the proportion of the low HDI category was 4.28%, the medium HDI category was 48.64%, and the high HDI category was 47.08%. Based on logistic regression modeling using backward elimination which has the smallest AIC value of 293.387, a model with independent variables of average years of schooling (X1), population density (X4), percentage of poor people (X5) and sex ratio (X6) is a variable that significantly affects the HDI of regencies/cities in Indonesia in 2021. The accuracy value of the classification accuracy of training data and test data from the ordinal logistic regression model of HDI of regencies/cities in Indonesia in 2021 is 83.46% and 86.61%, respectively, which means that the model is good for prediction.

Keywords: Indonesia development index, backward elimination, k-nearest neighbor, ordinal logistic regression.

Abstrak

Indeks pembangunan manusia (IPM) merupakan salah satu hal penting yang perlu diperhatikan di Indonesia saat ini. Pertumbuhan IPM di Indonesia tahun 2021 tidak merata di semua Kabupaten/Kota dan memiliki kesenjangan yang tinggi. Tujuan penelitian ini mengetahui gambaran data IPM, mendapatkan model regresi logistik ordinal untuk mengetahui faktor-faktor yang signifikan memengaruhi IPM Kabupaten/Kota di Indonesia tahun 2021 dan mengidentifikasi hasil ketepatan klasifikasi dari model tersebut. Peubah bebas yang digunakan

pada penelitian ini adalah rata-rata lama sekolah, tingkat pengangguran terbuka, laju pertumbuhan penduduk, kepadatan penduduk, persentase penduduk miskin dan *sex ratio*. Peubah bebas pada penelitian ini mengandung *missing value* sehingga ditangani menggunakan imputasi *k-nearest neighbour* (KNN) dan dilanjutkan pemodelan menggunakan regresi logistik ordinal dengan menggunakan teknik *backward elimination* untuk mendapatkan faktor-faktor yang signifikan. Hasil penelitian menunjukkan proporsi kategori IPM rendah 4,28%, kategori IPM sedang 48,64% dan kategori IPM tinggi 47,08%. Berdasarkan pemodelan regresi logistik menggunakan *backward elimination* yang memiliki nilai AIC terkecil 293,387 menghasilkan model dengan peubah bebas rata-rata lama sekolah (X_1), kepadatan penduduk (X_4), presentase penduduk miskin (X_5) dan *sex ratio* (X_6) merupakan peubah yang signifikan memengaruhi IPM Kabupaten/Kota di Indonesia tahun 2021. Nilai akurasi dari ketepatan klasifikasi data latih dan data uji dari model regresi logistik ordinal IPM Kabupaten/Kota di Indonesia tahun 2021 secara berturut-turut sebesar 83,46% dan 86,61% yang artinya model tersebut baik untuk dilakukan prediksi.

Kata-kata kunci: Indeks pembangunan Indonesia, *backward elimination*, *k-nearest neighbour*, regresi logistik ordinal.

PENDAHULUAN

Pembangunan dan sumber daya manusia yang berkualitas tinggi merupakan hal yang penting bagi negara yang sedang berkembang karena keduanya menjadi tolok ukur kesuksesan suatu negara (Zarkasi et al., 2021). Kesuksesan suatu negara dapat dilihat berdasarkan kondisi kesejahteraan manusia. Kondisi tersebut merupakan hasil dari upaya pemerintah dalam bidang pembangunan dan sumber daya manusia agar dapat menjamin kondisi tersebut. Salah satu indikator yang dapat menjamin kesejahteraan manusia lewat pembangunan adalah indeks pembangunan manusia (Budiman & Cahyani, 2022).

Indeks pembangunan manusia merupakan suatu indeks komposit yang digunakan untuk mengukur capaian pembangunan manusia sebagai dasar kualitas hidup manusia. Tolok ukur yang dijadikan dasar IPM adalah indeks harapan hidup, indeks pendidikan dan indeks standar hidup layak (BPS, 2021). Badan Pusat Statistika mengkategorikan nilai IPM menjadi empat kategori yakni sangat tinggi, tinggi, sedang, dan rendah. IPM merupakan salah satu hal penting yang perlu diperhatikan di Indonesia saat ini. Pertumbuhan IPM di Indonesia tahun 2021 tidak merata di semua Kabupaten/Kota dan memiliki kesenjangan yang tinggi. Nilai IPM tertinggi di Indonesia tahun 2021 adalah 87,18 dan terendah adalah 32,84.

Berdasarkan data tersebut penelitian ini akan menelusuri faktor-faktor penyebab meningkat dan menurunnya IPM Kabupaten/Kota di Indonesia tahun 2021. Faktor-faktor di setiap sektor pembentuk IPM akan cenderung memiliki hubungan yang kuat satu sama lainnya, karena antar faktor-faktor tersebut saling memengaruhi (Putra & Ratnasari, 2015), sehingga memerlukan analisis lebih lanjut terkait hal tersebut. Penelitian ini menggunakan metode regresi logistik ordinal untuk menganalisis kasus IPM. Metode ini dipilih karena peubah bebasnya terdiri lebih dari dua kategori dan memiliki tingkatan. Regresi logistik ordinal adalah teknik regresi yang memodelkan pengaruh peubah bebas terhadap peubah respon, dimana peubah respon bersifat polikotomus dengan skala pengukuran ordinal (Zakaryah & Zain, 2015).

Penelitian sebelumnya terkait pemodelan dengan regresi logistik ordinal pernah dilakukan oleh (Rahmadeni & Yunita, 2019) menggunakan IPM provinsi Riau tahun 2014-2017 dan penelitian (Budiman & Cahyani, 2022) menggunakan IPM Jawa Timur tahun 2022. Pada penelitian sebelumnya, data yang digunakan hanya mencakup provinsi sehingga pada penelitian ini peneliti menggunakan data IPM dengan menambahkan seluruh wilayah Kabupaten/kota di Indonesia pada tahun 2021 dengan peubah bebasnya yaitu rata-rata lama sekolah, tingkat pengangguran terbuka, laju pertumbuhan penduduk, kepadatan penduduk, persentase penduduk miskin dan *Sex Ratio*.

Data yang digunakan pada penelitian ini terdapat *missing value* pada data peubah bebasnya, sehingga perlu dilakukan penanganan sebelum masuk ke pemodelan. Menurut (Khan & Hoque, 2020)

metode yang efektif untuk menangani *missing value* yaitu metode *k-nearest neighbour* (KNN). KNN adalah metode imputasi *missing value* berdasarkan informasi dari observasi terdekat yang mempunyai kemiripan dengan *missing value* (Susanti et al., 2018). Merujuk dari (Khan & Hoque, 2020), untuk menangani *missing value* pada data peubah bebas peneliti menggunakan metode KNN. Kemudian untuk mendapatkan model dengan peubah bebas yang signifikan memengaruhi peubah respon IPM Kabupaten/Kota di Indonesia tahun 2021, pemodelan regresi logistik ordinal dilakukan dengan menggunakan metode *backward elimination*.

Metode *Backward elimination* adalah teknik regresi untuk menyeleksi peubah bebas terhadap peubah respon yang signifikan secara mundur bertahap sehingga mendapatkan model yang terbaik dan meningkatkan akurasi klasifikasi (Rizal & Lutfi, 2020). Tujuan dari penelitian ini adalah (1) mengetahui gambaran data IPM Kabupaten/Kota di Indonesia tahun 2021, (2) mengaplikasikan model regresi logistik ordinal *backward elimination* dengan imputasi KNN untuk mendapatkan faktor-faktor yang signifikan memengaruhi IPM Kabupaten/Kota di Indonesia tahun 2021, (3) mengidentifikasi hasil ketepatan klasifikasi dari model.

METODOLOGI

Bahan dan Data

Penelitian ini menggunakan data sekunder yang di ambil dari *website* Badan Pusat Statistika (BPS) tahun 2021. Data terdiri dari 514 Kabupaten/Kota di Indonesia yang terdiri dari peubah respon yang disimbolkan dengan Y dan peubah bebas yang disimbolkan dengan X. Peubah respon yang digunakan yaitu data indeks pembangunan manusia yang dikategorikan rendah jika nilai IPM < 60, kategori sedang jika nilai $60 \leq \text{IPM} < 70$, dan kategori tinggi (tinggi dan sangat tinggi) jika nilai $\text{IPM} \geq 70$. Peubah bebas yang digunakan yaitu rata-rata lama sekolah, tingkat pengangguran terbuka, laju pertumbuhan penduduk, kepadatan penduduk, presentasi penduduk miskin dan *sex ratio*. Analisis dilakukan dengan bantuan *software* R studio. TABEL 1 merupakan literatur yang menjelaskan hubungan beberapa peubah dengan tingkat IPM.

TABEL 1. Peubah penelitian dan referensi

No	Peubah Penelitian	Referensi	Skala Pengukuran
1.	Kategori Nilai IPM (Y)		Ordinal
2.	Rata-rata Lama Sekolah (%) (X ₁)	(Rahmadeni & Yunita, 2019) (Mahya & Widowati, 2021)	Rasio
3.	Tingkat Pengangguran Terbuka (%) (X ₂)	(Noviatamara et al., 2019)	Rasio
4.	Laju Pertumbuhan Penduduk (%) (X ₃)	(Budiman & Cahyani, 2022) (Fibriyani & Afifah, 2018)	Rasio
5.	Kepadatan Penduduk (X ₄)	(Fibriyani & Afifah, 2018)	Rasio
6.	Presentase Penduduk Miskin (%) (X ₅)	(Fathurahman et al., 2019) (Faizia et al., 2019)	Rasio
7.	<i>Sex Ratio</i> (%) (X ₆)	(Fibriyani & Afifah, 2018)	Rasio

Metode Penelitian

Metode dan tahapan analisis yang akan digunakan pada penelitian ini adalah sebagai berikut:

1. Melakukan analisis deskriptif dan eksplorasi data awal.
2. Menangani *missing value* pada peubah bebas menggunakan metode KNN dengan k=5.

Metode KNN memprediksi *missing value* berdasarkan nilai k observasi terdekat. Penanganan *missing value* dengan *k-nearest neighbour* sebagai berikut:

- a. Mengatur jumlah k , k banyaknya pengamatan terdekat dari data yang hilang.
- b. Menghitung jarak dengan menggunakan jarak *Euclidean*. Jarak dihitung antara pengamatan yang mengandung *missing value* dan tidak mengandung *missing value*. Dimana x_{ai} dengan $i=1,2,\dots,n$ adalah peubah ke- i yang mengandung *missing value* dalam setiap pengamatan dan x_{bi} dengan $i=1,2,\dots,n$ adalah peubah ke- i yang tidak mengandung *missing value* pada setiap observasi, maka jarak (d) bisa dihitung dengan mengikuti persamaan, yaitu:

$$d = \sqrt{\sum_{i=1}^n (x_{ai} - x_{bi})^2} \tag{1}$$

- c. Mengurutkan jarak berdasarkan nilai terkecil dan menentukan k terdekat dengan observasi.
- d. Memasukkan *missing value*, Jika diketahui korelasi antara k tetangga dan *missing value* adalah r_k dan $\epsilon = 10^{-6}$. Jadi, nilai data yang hilang dari k pengamatan terdekat (Liao et al., 2014):

$$w_k = \left(\frac{r_k^2}{(1 - r_k^2 + \epsilon)} \right)^2 \tag{2}$$

3. Membagi data menjadi data latih dan data uji, dimana data latih sebesar 75% dan data uji sebesar 25%.
4. Mendeteksi multikolinearitas pada peubah bebas menggunakan nilai korelasi antar peubah.
5. Melakukan pemodelan dengan tahapan sebagai berikut:
 - Mengestimasi parameter regresi logistik ordinal menggunakan MLE

Model yang digunakan untuk regresi logistik ordinal adalah model logit. Model logit adalah penerapan model GLMs dengan fungsi hubungannya adalah *cumulative logit models*. Persamaan peluang pada kategori peubah respon yang berskala ordinal sejumlah m , dimana $r = 1 < \dots < m$ yang memiliki sifat $\pi_1 + \dots + \pi_m = 1$. Fungsi peluang kumulatif peubah respon Y adalah $P(Y \leq r | X_j) = \pi_j(x)$. Didefinisikan sebagai berikut (Smith et al., 2020):

$$\pi_j(x) = P(Y \leq r | X_j) = \frac{\exp\left(\beta_{0r} + \sum_{j=1}^p \beta_j x_{ij}\right)}{1 + \exp\left(\beta_{0r} + \sum_{j=1}^p \beta_j x_{ij}\right)} \tag{3}$$

Kemudian persamaan (1) di sambung ke fungsi linier dengan menggunakan fungsi hubung logit $Logit[\pi_r(x)] = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right]$ berikut:

$$g(x) = Logit(\pi_r(x)) = \beta_{0r} + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_{ij} \tag{4}$$

Dengan metode kemungkinan maksimum (*Maximum Likelihood Estimation*), akan didapatkan nilai β_j yang memaksimalkan fungsi $l(\beta) = \prod_{i=1}^n [\pi_0(x)^{y_{0i}} \pi_1(x)^{y_{1i}} \dots \pi_m(x)^{y_{mi}}]$. Peluang setiap kategori peubah respon dihitung berdasarkan selisih nilai peluang kumulatif masing-masing fungsi logit, yaitu $\pi_1(x) = P(Y \leq 1)$, $\pi_2(x) = P(Y \leq 2) - P(Y \leq 1)$ hingga $P(Y \leq m) = 1 - P(Y \leq m - 1)$.

- Melakukan pengujian parameter secara simultan menggunakan uji LRT

Pengujian parameter secara simultan bertujuan untuk melihat pengaruh peubah secara bersama terhadap peubah respon dengan menggunakan uji *Likelihood Ratio Test* (G^2). Hipotesisnya $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ vs $H_1 : \text{minimal ada satu } \beta_r \neq 0, r = 1, 2, \dots, p$ (terdapat pengaruh peubah bebas terhadap peubah respon). Statistik ujinya sebagai berikut:

$$G^2 = -2 \ln \left[\frac{\left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1} \dots \left(\frac{n_m}{n}\right)^{n_m}}{\prod_{i=1}^n [\pi_0(x)^{y_{0i}} \pi_1(x)^{y_{1i}} \dots \pi_m(x)^{y_{mi}}]} \right] \tag{5}$$

Keputusan tolak H_0 jika $G^2 > \chi^2_{(\alpha, db)}$ atau $P\text{-value} < \alpha$, dimana db merupakan banyaknya parameter.

- Melakukan pengujian parameter secara parsial menggunakan uji *Wald*

Pengujian parameter secara parsial bertujuan untuk melihat pengaruh masing-masing peubah bebas terhadap peubah respon dengan menggunakan uji *Wald*. Hipotesisnya $H_0 : \beta_r = 0, r = 1, 2, \dots, p$ vs $H_1 : \beta_r \neq 0, r = 1, 2, \dots, p$ (terdapat pengaruh peubah bebas β_r terhadap peubah respon) Statistik ujinya sebagai berikut:

$$W = \frac{\hat{\beta}_r}{SE(\hat{\beta})} \tag{6}$$

Keputusan tolak H_0 jika $|W| > Z_{\alpha/2}$ atau $P\text{-value} < \alpha$

6. Mereduksi peubah-peubah bebas yang tidak signifikan terhadap peubah responnya menggunakan *backward elimination*.
7. Mendapatkan model yang di diperoleh dari langkah (6) dengan nilai AIC terkecil dan melakukan interpretasi pada model menggunakan nilai *odds ratio*.
8. Melakukan pengujian kesesuaian model dengan menggunakan uji Hosmer Lemeshow.

Data yang telah dimodelkan dengan regresi logistik ordinal kemudian diukur kesesuaian datanya dengan model. Uji kesesuaian yang digunakan dalam analisis ini adalah uji Hosmer Lemeshow. Uji ini memiliki hipotesis H_0 dimana model yang digunakan mengikuti data, dan H_1 dimana model yang digunakan tidak mengikuti data. Hasil uji Hosmer Lemeshow akan menolak H_0 jika $P\text{-value}$ kurang dari α (Díaz-Pérez et al., 2019).

9. Mengevaluasi model menggunakan nilai akurasi pada ketepatan hasil klasifikasi data latih dan data uji

Evaluasi model menggunakan ketepatan hasil klasifikasi dari data latih dan data uji. Hasil klasifikasi dari pemodelan dapat diukur akurasinya dengan melihat nilai *confusion matrix*. Nilai akurasi dari hasil klasifikasi model regresi logistik ordinal dapat dilihat pada TABEL 2 (Yuniarty et al., 2022).

TABEL 2. Tabel confusion matrix IPM Kabupaten/Kota tahun 2021

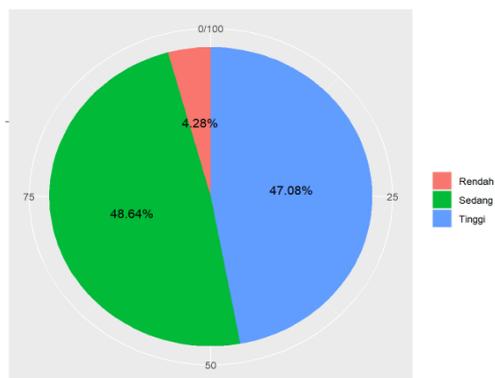
Prediksi	Aktual		
	Rendah	Sedang	Tinggi
Rendah	a	b	c
Sedang	d	e	f
Tinggi	g	h	i

Dari *confusion matrix* pada TABEL 2. nilai akurasi dapat dihitung $Accuracy = \frac{a + e + i}{n}$, dimana n menyatakan jumlah observasi.

HASIL DAN PEMBAHASAN

Analisis Deskriptif dan Eksplorasi Data Awal

Analisis IPM Kabupaten/Kota di Indonesia tahun 2021 menggunakan regresi logistik ordinal dimana data diberi label 1 (rendah) untuk $IPM < 60$, label 2 (sedang) untuk $60 \leq IPM < 70$, dan label 3 (tinggi) untuk $IPM \geq 70$ yang disajikan pada (GAMBAR 1).



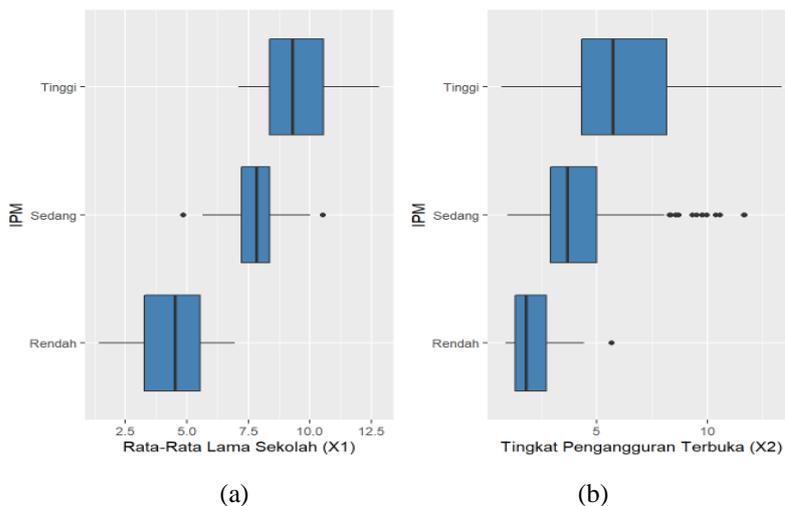
GAMBAR 1. Proporsi IPM Kabupaten/Kota di Indonesia tahun 2021

Kategori rendah berjumlah 22 amatan dengan proporsi 4,28%, kategori sedang berjumlah 250 amatan dengan proporsi 48,64% dan kategori tinggi berjumlah 242 amatan dengan proporsi 47,08%. Selanjutnya melihat statistik deskriptif dari peubah bebas yang digunakan.

TABEL 3. Statistika deskriptif peubah bebas

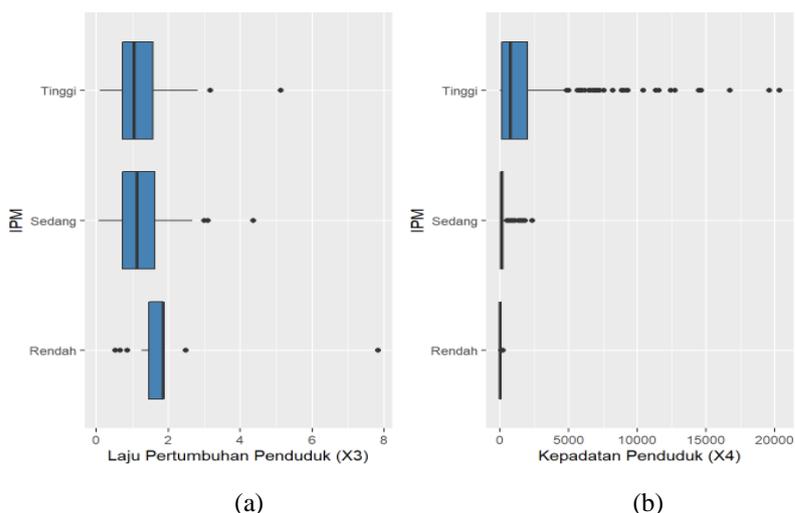
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Min	1,42	0,70	0,06	1	2,38	9,9
Max	12,83	13,37	7,84	20360	41,66	138,9
Mean	8,44	5,14	1,23	1147,70	12,27	103,1
Median	8,31	4,66	1,12	167	10,46	102,6
NA's	-	10	-	24	-	13

Terlihat bahwa peubah tingkat pengangguran terbuka (X₂), kepadatan penduduk (X₄) dan *sex ratio* (X₆) memiliki *missing value* secara berturut-turut sebanyak 10 amatan, 24 amatan dan 13 amatan sehingga penanganan *missing value* perlu dilakukan sebelum melakukan pemodelan.



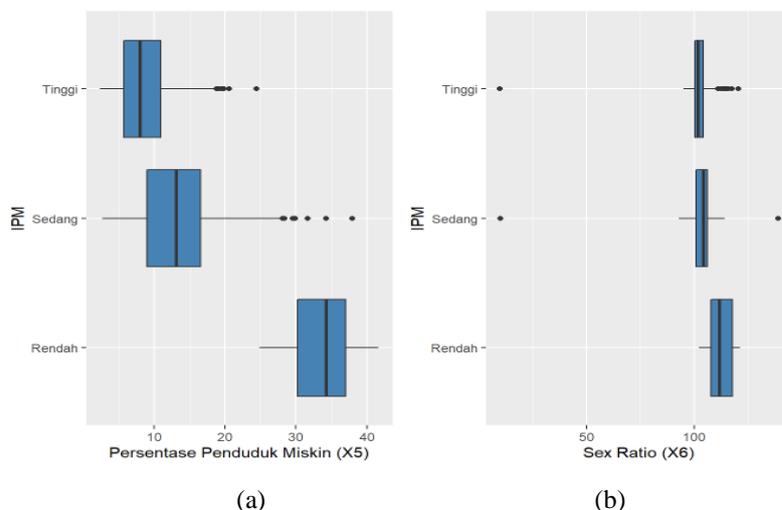
GAMBAR 2. Hubungan (a) Rata-rata lama sekolah (X_1) dengan IPM, (b) TPT (X_2) dengan IPM

Berdasarkan (GAMBAR 2), bagian a dan b menjelaskan kecenderungan rata-rata lama sekolah dan tingkat pengangguran terbuka memiliki sebaran nilai IPM terbanyak yaitu pada kategori tinggi dibandingkan kategori sedang dan rendah.



GAMBAR 3. Hubungan (a) Laju pertumbuhan penduduk (X_3) dengan IPM, (b) Kepadatan penduduk (X_4) dengan IPM

Berdasarkan (GAMBAR 3), bagian a menjelaskan kecenderungan laju pertumbuhan penduduk memiliki sebaran nilai IPM terbanyak yaitu pada kategori rendah dibandingkan kategori tinggi dan sedang. Bagian b menjelaskan kecenderungan kepadatan penduduk memiliki sebaran nilai IPM terbanyak yaitu pada kategori tinggi dibandingkan kategori sedang dan rendah.

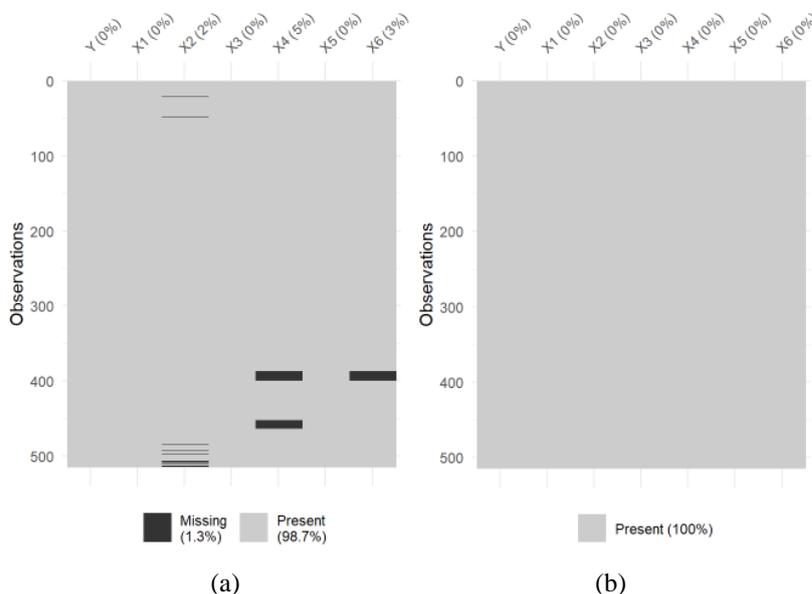


GAMBAR 4. Hubungan (a) Presentase penduduk miskin (X_5) dengan IPM, (b) *sex ratio* (X_6) dengan IPM

Berdasarkan (GAMBAR 4), bagian a dan bagian b menjelaskan kecenderungan persentase penduduk miskin dan *sex ratio* memiliki sebaran nilai IPM terbanyak yaitu pada kategori rendah dibandingkan kategori tinggi dan sedang

Penanganan *Missing Value*

Data yang hilang pada peubah X_2 , X_4 , dan X_6 dalam penelitian ini diduga dengan menggunakan metode k-nearest neighbor (KNN) dengan $k = 5$, artinya data yang hilang diduga berdasarkan lima data terdekat.



GAMBAR 5. (a) Sebelum *missing value* ditangani, (b) sesudah *missing value* ditangani metode KNN

Berdasarkan (GAMBAR 5) terdapat 1,3% *missing value* pada data kemudian diimput menggunakan metode KNN. Setelah dilakukan penanganan *missing value* kemudian dilakukan pembagian data latih dan uji.

Data Latih dan Data Uji

Pada tahapan ini data dibagi menjadi 75% data latih yang digunakan untuk pemodelan dan 25% data uji yang digunakan untuk melakukan klasifikasi. Hasil yang di dapat yaitu jumlah data latih sebesar 382 Kabupaten/Kota dan data uji sebesar 127 Kabupaten/Kota.

Pendeteksian Multikolinearitas

Pendeteksian multikolinearitas pada peubah bebas perlu dilakukan sebelum masuk ke tahap pemodelan regresi logistik ordinal dengan melihat nilai korelasi antar peubah bebas. Nilai korelasi antar peubah bebas disajikan pada TABEL 4.

TABEL 4. Nilai korelasi antar peubah bebas

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
X ₁		0,480	0,056	0,448	0,585	0,192
X ₂			0,067	0,511	0,407	0,127
X ₃				0,178	0,173	0,256
X ₄					0,273	0,147
X ₅						0,199
X ₆						

TABEL 4 menunjukkan nilai korelasi antar peubah bebas tidak lebih dari 0,8 sehingga tidak terdeteksi adanya multikolinearitas pada peubah bebas yang digunakan.

Pemodelan Regresi Logistik Ordinal

Pemodelan regresi logistik ordinal menggunakan data latih dan pendugaan model menggunakan metode *Maximum Likelihood*. Hasil pendugaan parameter model disajikan pada TABEL 5.

TABEL 5. Hasil pendugaan parameter model regresi logistik ordinal

Peubah	Koefisien $\hat{\beta}_j$	Standard Error
Peubah bebas		
X ₁	1,884	0,232
X ₂	0,026	0,084
X ₃	-0,274	0,237
X ₄	0,002	0,000
X ₅	-0,209	0,034
X ₆	-0,039	0,015
Intercept		
Rendah Sedang	1,151	0,005
Sedang Tinggi	10,014	1,012

Pengujian Parameter

Langkah awal yaitu pengujian secara simultan yang bertujuan untuk melihat pengaruh peubah bebas secara bersama terhadap peubah respon dengan menggunakan uji *Likelihood Ratio Test* (G^2). berdasarkan hasil yang diperoleh nilai $G^2 = 372,279 > \chi^2_{0,05;6} = 12,5916$ maka keputusannya adalah menolak H_0 yang artinya minimal ada satu peubah bebas yang memiliki pengaruh terhadap peubah respon.

Selanjutnya dilakukan uji secara parsial yang bertujuan untuk melihat pengaruh masing-masing peubah bebas terhadap peubah respon dengan menggunakan uji *Wald*. Hasil pengujian secara parsial disajikan pada TABEL 6.

TABEL 6. Hasil uji parameter secara parsial

Peubah	Koefisian $\hat{\beta}_j$	Standard Error	Nilai Wald	P-value
Peubah bebas				
X ₁	1,884	0,232	8,122	0,000
X ₂	0,026	0,084	0,308	0,758
X ₃	-0,274	0,237	-1,155	0,248
X ₄	0,002	0,000	4,748	0,000
X ₅	-0,209	0,034	-6,212	0,000
X ₆	-0,039	0,015	-2,638	0,008
Intercept				
Rendah Sedang	1,151	0,005	252,985	0,000
Sedang Tinggi	10,014	1,012	9,893	0,000

TABEL 6 menunjukkan bahwa peubah rata-rata lama sekolah (X₁), kepadatan penduduk (X₄), persentase penduduk miskin (X₅) dan Sex Ratio (X₆) nilai *p-value* < α sehingga H₀ ditolak yang artinya keempat peubah tersebut berpengaruh secara signifikan terhadap IPM Kabupaten/Kota di Indonesia tahun 2021. Sedangkan peubah tingkat pengangguran terbuka (X₂) dan laju pertumbuhan penduduk (X₃) nilai *p-value* > α sehingga gagal menolak H₀ yang artinya kedua peubah tersebut tidak berpengaruh secara signifikan terhadap IPM Kabupaten/Kota di Indonesia tahun 2021. Karena terdapat peubah yang tidak signifikan pada model maka akan dilanjutkan mereduksi peubah yang tidak signifikan tersebut dengan menggunakan metode *backward elimination*.

Backward Elimination

Backward elimination adalah teknik untuk meregresikan peubah bebas terhadap peubah respon menggunakan langkah mundur dengan mengeluarkan peubah bebas yang tidak signifikan atau nilai *p-value* > α secara bertahap. Langkah pertama memasukkan semua peubah bebas ke dalam model, kemudian dilanjutkan dengan pengujian parameter model, selanjutnya berdasarkan nilai *p-value* terbesar dilakukan eliminasi. Hasil *backward elimination* disajikan pada TABEL 7.

TABEL 7. Hasil *backward elimination*

Step	Step 1					
Peubah masuk	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Peubah keluar						
P-value	0,000	0,758	0,248	0,000	0,000	0,008
AIC	296,052					
Step	Step 2					
Peubah masuk	X ₁		X ₃	X ₄	X ₅	X ₆
Peubah keluar		X ₂				
P-value	0,000		0,258	0,000	0,000	0,009
AIC	294,146					
Step	Step 3					
Peubah masuk	X ₁			X ₄	X ₅	X ₆
Peubah keluar		X ₂	X ₃			
P-value	0,000			0,000	0,000	0,001
AIC	293,387					

Tahap pertama memasukkan semua peubah bebas dan menghasilkan nilai *p-value* terbesar yang tidak signifikan pada peubah X_2 sebesar 0,758 sehingga tahap selanjutnya tingkat pegangguran terbuka (X_2) akan dikeluarkan dari model. Tahap kedua tanpa memasukkan peubah X_2 pada model dan menghasilkan nilai *p-value* terbesar yang tidak signifikan pada peubah X_3 sebesar 0,258 sehingga tahap selanjutnya laju pertumbuhan penduduk (X_3) akan dikeluarkan dari model. Tahap ketiga tanpa memasukkan peubah X_2 dan X_3 pada model dan menghasilkan peubah X_1 , X_4 , X_5 dan X_6 merupakan peubah yang signifikan yakni nilai *p-value* lebih kecil dari 0,05 dan memperoleh nilai AIC terkecil 293,387. Tahap ketiga adalah tahap akhir dari metode *backward elimination* yang menghasilkan peubah X_1 , X_4 , X_5 dan X_6 merupakan peubah bebas yang akan digunakan untuk memodelkan IPM Kabupaten/Kota di Indonesia tahun 2021 menggunakan regresi logistik ordinal.

Model Regresi Logistik Ordinal

Berdasarkan hasil *backward elimination* dengan nilai AIC paling kecil sebesar 293,387 model dengan peubah X_1 , X_4 , X_5 dan X_6 merupakan model yang digunakan untuk memodelkan data IPM Kabupaten/Kota di Indonesia tahun 2021. Hasil estimasi dan *odds ratio* model tersebut disajikan pada TABEL 9.

TABEL 9. Nilai koefisien dan *odds ratio* dari model

Peubah	Koefisien $\hat{\beta}_j$	Odds Ratio
Peubah Bebas		
X_1	1,849	6,357
X_4	0,002	1,002
X_5	-0,274	0,811
X_6	-0,209	0,951
Intercept		
Rendah Sedang	0,131	1,139
Sedang Tinggi	8,855	7007,102

Sehingga persamaan regresi logistik ordinal disajikan pada persamaan (7) dan (8)

$$\text{Logit}[P(Y \leq 1|X_j)] = 0,131 + 1,849X_1 + 0,002X_4 - 0,002X_5 - 0,209X_6 \tag{7}$$

$$\text{Logit}[P(Y \leq 2|X_j)] = 8,855 + 1,849X_1 + 0,002X_4 - 0,002X_5 - 0,209X_6 \tag{8}$$

Dari TABEL 9 nilai *odds ratio* peubah bebas dapat diinterpretasikan sebagai berikut:

1. Peubah rata-rata lama sekolah (X_1) memiliki nilai *odds ratio* sebesar 6,357. artinya setiap kenaikan 1% rata-rata lama sekolah akan menaikkan *odds* kategori IPM Kabupaten/Kota tinggi atau sedang sebesar 6,357 dibandingkan *odds* IPM Kabupaten/Kota kategori rendah dengan asumsi peubah bebas lainnya konstan.
2. Peubah kepadatan penduduk (X_4) memiliki nilai *odds ratio* sebesar 1,002. artinya setiap kenaikan 1 jiwa/km² kepadatan penduduk akan menaikkan *odds* IPM Kabupaten/Kota kategori tinggi atau sedang sebesar 1,002 dibandingkan *odds* IPM Kabupaten/Kota kategori rendah dengan asumsi peubah bebas lainnya konstan.
3. Peubah persentase penduduk miskin (X_5) memiliki nilai *odds ratio* sebesar 0,811. artinya setiap kenaikan 1% persentase penduduk miskin akan menurunkan *odds* IPM Kabupaten/Kota kategori tinggi atau sedang sebesar 0,811 dibandingkan *odds* IPM Kabupaten/Kota kategori rendah dengan asumsi peubah bebas lainnya konstan.

4. Peubah *sex ratio* (X_6) memiliki nilai *odds ratio* sebesar 0,951. artinya setiap kenaikan 1% *sex ratio* akan menurunkan *odds* IPM Kabupaten/Kota kategori tinggi atau sedang sebesar 0,951 dibandingkan *odds* IPM Kabupaten/Kota kategori rendah dengan asumsi peubah bebas lainnya konstan.

Pengujian Kesesuaian Model

Pengujian kesesuaian model atau *Goodness of fit* bertujuan untuk mengetahui model yang digunakan sesuai atau tidak dengan data yang diamati. Hasil pengujian kesesuaian model disajikan pada TABEL 10.

TABEL 10. Hasil uji kesesuaian model

Model	χ^2	P-value
Logit Ordinal	14,521	0,6299

Berdasarkan TABEL 10 nilai *p-value* lebih besar dari α maka gagal menolak H_0 yang artinya model yang didapat sesuai dengan data yang diamati.

Evaluasi Model

Model regresi logistik ordinal yang dibentuk pada persamaan (7) dan (8) dievaluasi dengan menguji akurasi model untuk data latih dan data uji menggunakan *confusion matrix*. *Confusion matrix* data latih dan data uji ditunjukkan pada TABEL 10 dan TABEL 11.

TABEL 11. *Confusion matrix* regresi logistik ordinal untuk data latih

Prediksi	Aktual		
	Rendah	Sedang	Tinggi
Rendah	14	1	0
Sedang	3	165	38
Tinggi	0	22	144

TABEL 12. *Confusion matrix* regresi logistik ordinal untuk data uji

Prediksi	Aktual		
	Rendah	Sedang	Tinggi
Rendah	4	0	0
Sedang	1	57	11
Tinggi	0	5	49

Berdasarkan hasil *confusion matrix* diperoleh nilai akurasi data latih sebesar 83,46% dan nilai akurasi data uji sebesar 86,61% sehingga model regresi logistik ordinal yang diperoleh dikategorikan baik untuk dilakukan prediksi.

KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan diperoleh proporsi kategori IPM rendah 4,28%, kategori IPM sedang 48,64% dan kategori IPM tinggi 47,08%. Pemodelan regresi logistik ordinal menggunakan teknik *backward elimination* dengan imputasi *missing value* KNN menghasilkan peubah bebas rata-rata lama sekolah (X_1), kepadatan penduduk (X_4), presentase penduduk miskin (X_5) dan *sex ratio* (X_6) merupakan faktor-faktor yang signifikan memengaruhi IPM Kabupaten/Kota di Indonesia tahun 2021 hal ini dibuktikan dengan nilai AIC yang terkecil diperoleh sebesar 293,387. Nilai akurasi dari

ketepatan klasifikasi data latih dan data uji IPM Kabupaten/Kota di Indonesia tahun 2021 secara berturut-turut sebesar 83,46% dan 86,61% yang artinya tersebut baik untuk dilakukan prediksi. Untuk penelitian selanjutnya dapat menambahkan metode seleksi peubah lainnya seperti *forward selection* dan menggunakan algoritma genetika.

REFERENSI

- BPS. (2021). *Indeks Pembangunan Manusia*. Jakarta: Badan Pusat Statistik.
- Budiman, M. A., & Cahyani, N. (2022). Pemodelan Regresi Logistik Ordinal Pada Indeks Pembangunan Manusia (IPM) Di Jawa Timur Tahun 2020. *Jurnal Statistika Dan Komputasi*, 1(2), 64–73. <https://doi.org/10.32665/statkom.v1i2.1169>
- Díaz-Pérez, M., Carreño-Ortega, Á., Salinas-Andújar, J. A., & Callejón-Ferre, Á. J. (2019). Application of logistic regression models for the marketability of cucumber cultivars. *Agronomy*, 9(1), 3–19. <https://doi.org/10.3390/agronomy9010017>
- Faizia, T., Prahutama, A., & Yasin, H. (2019). *Pemodelan Indeks Pembangunan Manusia Di Jawa Tengah Dengan Regresi Komponen Utama Robust*. 8(2), 253–271.
- Fathurahman, M., Siringoringo, M., Satriya, A. M. A., & Sari Wulan, N. (2019). Pemodelan regresi logistik dan regresi probit pada indeks pembangunan manusia kabupaten/kota di pulau Kalimantan. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, 1, 172–182.
- Fibriyani, V., & Afifah, A. (2018). Pemodelan Indeks Pembangunan Manusia Kabupaten/Kota Di Provinsi Papua Dan Papua Barat Tahun 2014. *Jurnal VARIAN*, 1(2), 30–40.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley & Sons. Inc.
- Khan, S. I., & Hoque, A. S. M. L. (2020). SICE: an improved missing data imputation technique. *Journal of Big Data*, 7(1), 1–21.
- Liao, S. G., Lin, Y., Kang, D. D., Chandra, D., Bon, J., Kaminski, N., Sciurba, F. C., & Tseng, G. C. (2014). Missing value imputation in high-dimensional phenomic data: Imputable or not, and how? *BMC Bioinformatics*, 15(1), 1–12.
- Mahya, A. J., & Widowati. (2021). Analisis Pengaruh Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Dan Pengeluaran Per Kapita Terhadap Indeks Pembangunan Manusia Di Provinsi Jawa Tengah. *Prismatika: Jurnal Pendidikan Dan Riset Matematika*, 3(2), 126–140.
- Noviatamara, A., Ardina, T., & Amalia, N. (2019). Analisis Pengaruh Pertumbuhan Ekonomi Dan Tingkat Pengangguran Terbuka Di Daerah Istimewa Yogyakarta. *Jurnal REP (Riset Ekonomi Pembangunan)*, 4(1), 53–60.
- Putra, D. M., & Ratnasari, V. (2015). Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Dengan Menggunakan Metode Regresi Logistik Ridge. *Jurnal Sains Dan Seni ITS*, 4(2), 175–180.
- Rahmadeni, & Yunita, V. (2019). Pemodelan Indeks Pembangunan Manusia di Provinsi Riau dengan Menggunakan Regresi Logistik Ordinal. *Jurnal Sains Matematika Dan Statistika*, 5(1), 120–126.
- Rizal, A. S., & Lutfi, M. (2020). Prediksi Hasil Pemilu Legislatif Menggunakan Algoritma K-Nearest Neighbor Berbasis Backward Elimination. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, 3(1), 27–42.
- Smith, T. J., Walker, D. A., & McKenna, C. M. (2020). An Exploration of Link Functions Used in Ordinal Regression. *Journal of Modern Applied Statistical Methods*, 18(1), 2–15.
- Susanti, Martha, S., & Sulistianingsih, E. (2018). K Nearest Neighbor Dalam Imputasi Missing Data. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 7(1), 9–14.
- Yuniarty, T. Y., Erfiani, Indahwati, Fitrianto, A., & Khusnia N. K. (2022). Regresi Ordinal Logit dan Probit pada Faktor Kesejahteraan Rumah Tangga Petani Tanaman Pangan di Provinsi Sulawesi Tenggara. *Jurnal Statistika Dan Aplikasinya*, 6(2), 313–325.
- Zakariyah, & Zain, I. (2015). Analisis Regresi Logistik Ordinal pada Prestasi Belajar Lulusan Mahasiswa di ITS Berbasis SKEM. *Jurnal Sains Dan Seni ITS*, 4(1), 121–126.
- Zarkasi, R. N., Sifriyani, & Prangga, S. (2021). *Identifikasi Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia di Kalimantan Menggunakan Regresi Panel*. 15(2), 277–282.