



Kampus
Merdeka
INDONESIA JAYA



Sidang Skripsi

Program Studi Matematika
Universitas Pendidikan Indonesia
2023



ANALISIS REGRESI LOGISTIK BINER DENGAN METODE GROUP LASSO DALAM DATA BERDIMENSI TINGGI

**(Studi Kasus : Analisis Faktor-Faktor Capaian Indeks
Pembangunan Manusia Kota/Kabupaten di Jawa Barat)**

Penyusun

**Rizal Padhilah
NIM 1900826**

Pembimbing

**Drs. Nar Herrhyanto, M.Pd.
Dr. Lukman, S.Si., M.Si.**



Overview

Topik Bahasan Presentasi

1

Pendahuluan

2

Kajian Pustaka

3

Metodologi
Penelitian

4

Hasil dan
Pembahasan

5

Kesimpulan
dan Saran



Kampus
Merdeka
INDONESIA JAYA


The Education University

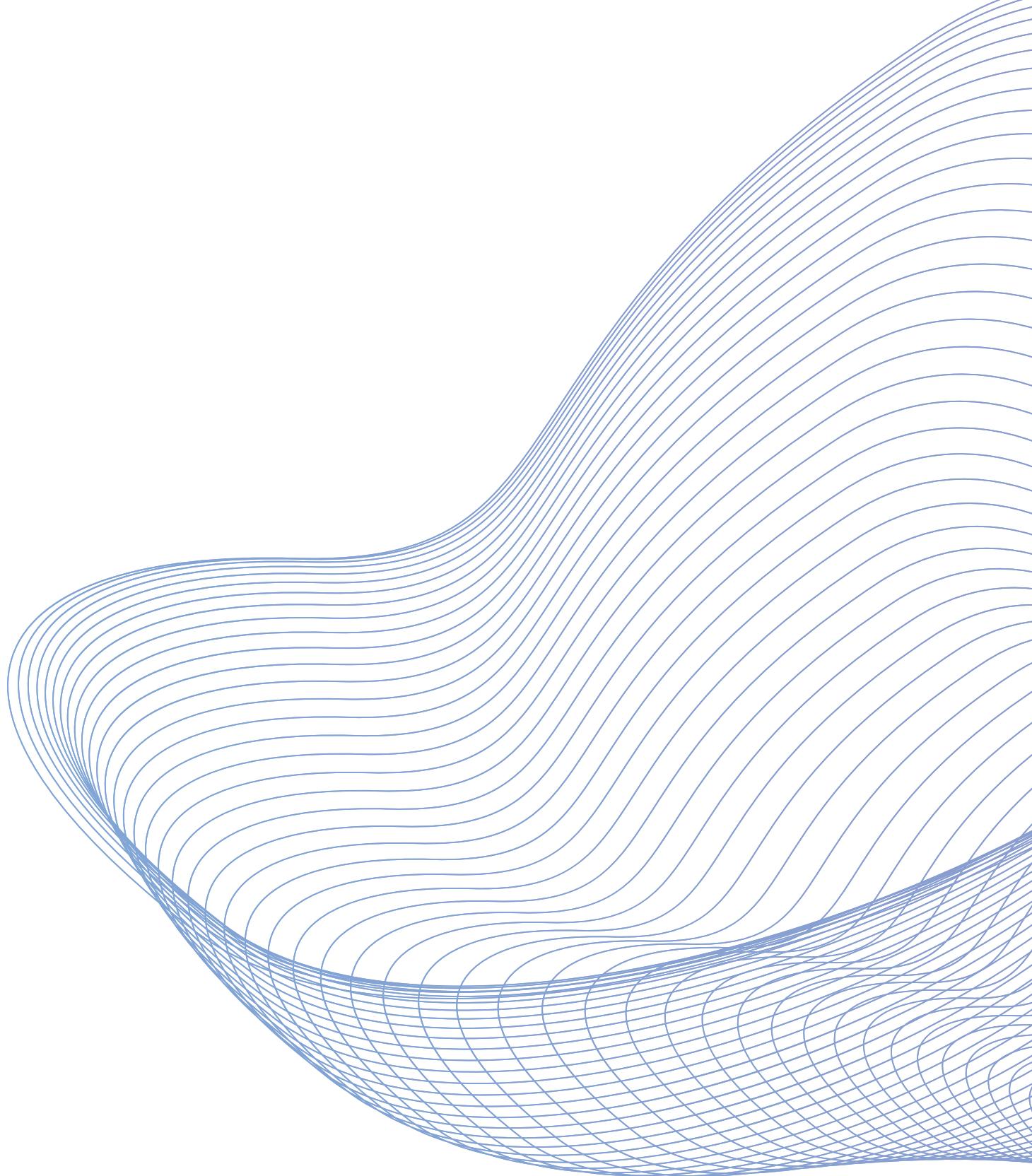
PENDAHULUAN

Latar Belakang

Rumusan Masalah

Tujuan

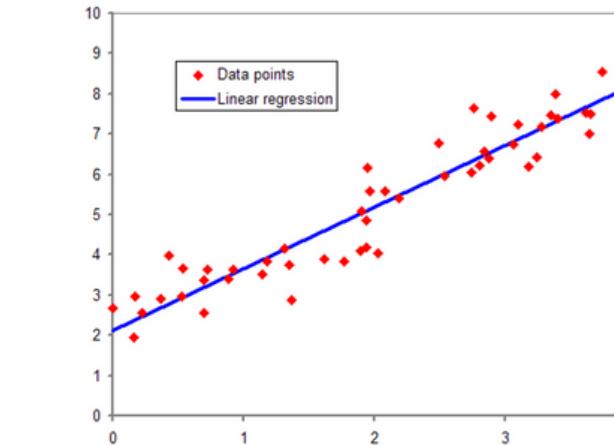
Manfaat



Latar Belakang



Perkembangan data saat ini sudah meningkat pesat dalam kehidupan manusia

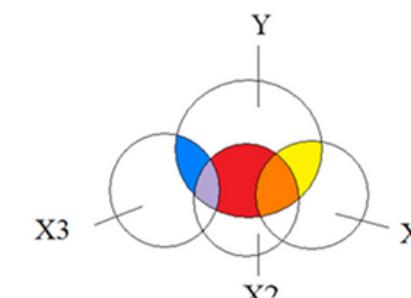


Analisis regresi menjadi salah satu teknik yang sering dipakai (Bazdaric dkk., 2021)

Regresi logistik digunakan untuk variabel terikat kategorikal (Hosmer dkk., 2013).

Tidak Dapat Diterapkan

Nilai koefisien regresi umumnya ditaksir dengan metode **Ordinary Least Square (OLS)**



Permasalahan multikolinearitas (Zhao dkk., 2020)



Varians parameter yang ditaksir terlalu besar dan mengurangi akurasi penaksiran (Vatcheva dkk., 2016)

	P_1	P_2	P_3	P_4	P_5	...	P_m
n_1							
n_2							
n_3							
..							
n_k							

Data berdimensi tinggi merupakan data dengan karakteristik jumlah variabel bebas lebih banyak dibandingkan amatan (Narisetty, 2020)

Tidak Dapat Diterapkan

OLS tidak dapat diimplementasikan pada data yang bersifat singular seperti data berdimensi tinggi (Setiawan & Sutikno, 2010)

Latar Belakang

LASSO

LASSO digunakan untuk menyeleksi variabel dan mengecilkan beberapa koefisien menuju nol (Hastie dkk., 2009).



Data berdimensi tinggi sangat memungkinkan untuk membentuk suatu kelompok variabel



Group LASSO

Group LASSO digunakan dalam pemilihan variabel pada data variabel bebas yang membentuk suatu kelompok (Chen & Xiang, 2017).

Group LASSO lebih baik dari LASSO

Pemilihan variabel
(Chen dkk. 2020)

Data berkorelasi tinggi
(Yunus dkk., 2017)



Latar Belakang

IPM dimanfaatkan dalam melihat tingkatan pembangunan suatu wilayah (Sapaat dkk., 2020)

10 kota/kabupaten
berada di atas capaian
IPM Jawa Barat



IPM 72,09
Jawa Barat

Tahun 2020
Badan Pusat Statistik (2021)

17 kota/kabupaten
berada di bawah capaian
IPM Jawa Barat

Adanya ketimpangan capaian IPM antar kota/kabupaten yang menjadi permasalahan

Umumnya untuk menganalisis faktor-faktor yang mempengaruhi terhadap capaian IPM digunakan *analisis regresi*

Penelitian Terdahulu
Putra dan Ratnasari (2015)
Sari dan Purhadi (2021)
Khairunnisa dkk. (2022)

- Regresi logistik dapat memprediksi pola hubungan variabel bebas terhadap variabel terikat kategorik.
- Regresi logistik dapat digunakan dalam klasifikasi (Edgar & Manz, 2017)

- Variabel yang digunakan dan diduga berpengaruh dalam penelitian-penelitian sebelumnya masih sangat sedikit

Penambahan banyak variabel bebas yang besar

↓
Merubah
Karakteristik Data

Data Berdimensi Tinggi

Perlu Pendekatan Lain dalam Analisis

Sunandi (2021) pada kasus analisis faktor yang berpengaruh terhadap IPM Provinsi Bengkulu tahun 2019.



Latar Belakang



**Data capaian IPM Kota/Kabupaten
di Jawa Barat tahun 2020**

Variabel Terikat Kategorik

Data Berdimensi Tinggi

27 Amatan & 40 Variabel Bebas



Rumusan Masalah

1

Bagaimana model regresi logistik biner dengan Group LASSO untuk mengetahui faktor-faktor yang mempengaruhi capaian IPM Kota dan Kabupaten di Jawa Barat tahun 2020?

2

Faktor-faktor manakah yang mempengaruhi capaian IPM Kota dan Kabupaten di Jawa Barat tahun 2020?



Tujuan

01

Memperoleh model regresi logistik biner menggunakan metode *Group LASSO* pada data capaian IPM Kota dan Kabupaten di Jawa Barat tahun 2020.

02

Memperoleh faktor-faktor yang mempengaruhi IPM Kota dan Kabupaten di Jawa Barat tahun 2020.



Manfaat Penelitian

01

Memberi wawasan baru berkaitan dengan analisis regresi logistik biner menggunakan metode *Group LASSO*

teoritis

02

Acuan untuk meningkatkan pembangunan dalam periode ke depan.

praktis



Kampus
Merdeka
INDONESIA JAYA

UPI
The Education University

KAJIAN PUSTAKA

Data Berdimensi Tinggi

Regressi Logistik

Asumsi

Multikolinearitas

Regressi Logistik Biner

Penaksir Regressi Logistik Biner

Interpretasi Parameter

LASSO

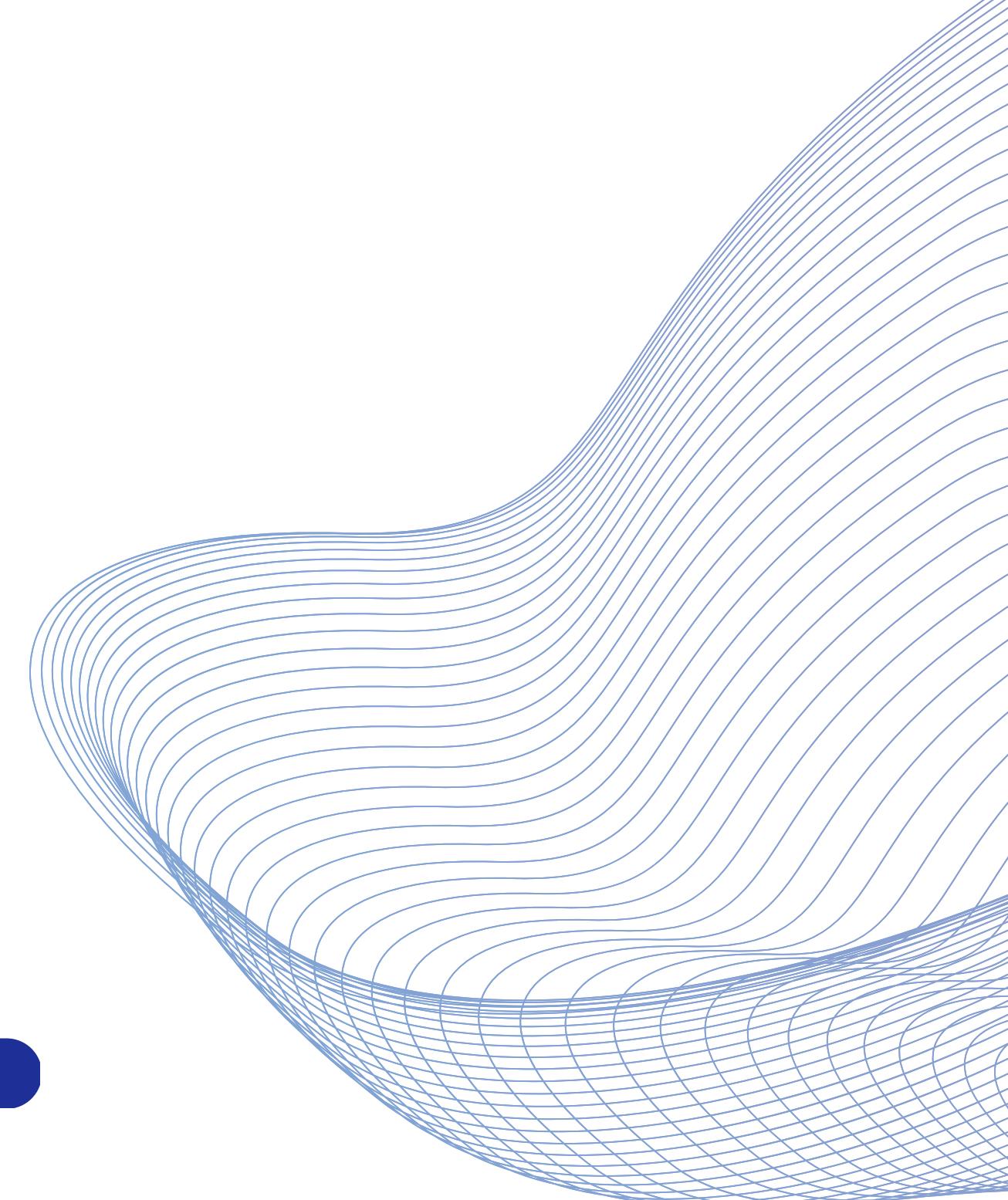
Group LASSO

Penaksir Regressi Logistik Biner Group LASSO

Algoritma Group Descent

Validasi Silang

Indeks Pembangunan Manusia





Data Berdimensi Tinggi

Data berdimensi tinggi merupakan data dengan karakteristik jumlah variabel bebas yang lebih tinggi dibandingkan amatan (**Narisetty, 2020**)

Masalah Multikolinearitas

Masalah *Overfitting*

Masalah Penaksiran

Bielza, dkk. (2011)



Regresi Logistik

Regresi logistik adalah salah satu jenis regresi yang digunakan untuk mengetahui keterkaitan antara variabel terikat kategorikal dengan satu atau lebih variabel bebas kategorikal atau kontinu.
(Hosmer dkk, 2013)

Asumsi Regresi Logistik

Bentuk distribusi variabel terikatnya tidak harus berdistribusi normal melainkan berada dalam keluarga distribusi eksponensial.

Tidak terdapat asumsi homogenitas varians

Tidak memerlukan hubungan yang linier antara variabel terikat dengan variabel bebas

Tidak terdapat gejala multikolinearitas untuk variabel bebas

Sisaan (error) tidak diasumsikan berdistribusi normal

Skala data variabel bebas dalam regresi logistik tidak mengharuskan dalam data interval.

(Garson, 2014)



Regresi Logistik Biner

Regresi logistik yang digunakan ketika variabel terikatnya memiliki **dua kategori** saja (Garson, 2014). Variabel Y pada regresi logistik biner mengikuti **distribusi Bernoulli** seperti berikut.

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

Penaksir Parameter Regresi Logistik Biner

Fungsi penaksir parameter regresi logistik biner dengan **Maximum Likelihood Method** sebagai berikut.

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

Untuk memudahkan perhitungan penaksiran parameter dilakukan dengan cara memaksimumkan fungsi **log-likelihood** (Herrhyanto, 2003)

$$l(\beta) = \sum_{i=1}^n y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]$$

Persamaan Model Regresi Logistik Biner

$$\pi(X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_p X_p)}$$

Bentuk Logit: $g(X) = \beta_0 + \sum_{i=1}^p \beta_i X_i$

Interpretasi Parameter Regresi Logistik Biner

Interpretasi dilakukan dengan melihat nilai **Odds Ratio** yang dinyatakan sebagai berikut.

$$OR = \frac{odds_1}{odds_2} = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = \exp(\beta_i)$$

(Hosmer dkk, 2013)



Uji Multikolinearitas

Korelasi Antarvariabel Bebas

Apabila terdapat korelasi antara dua variabel bebas yang nilainya lebih besar dari 0,8, maka hal ini menandakan adanya permasalahan multikolinearitas.

(Senaviratna & Cooray, 2019)

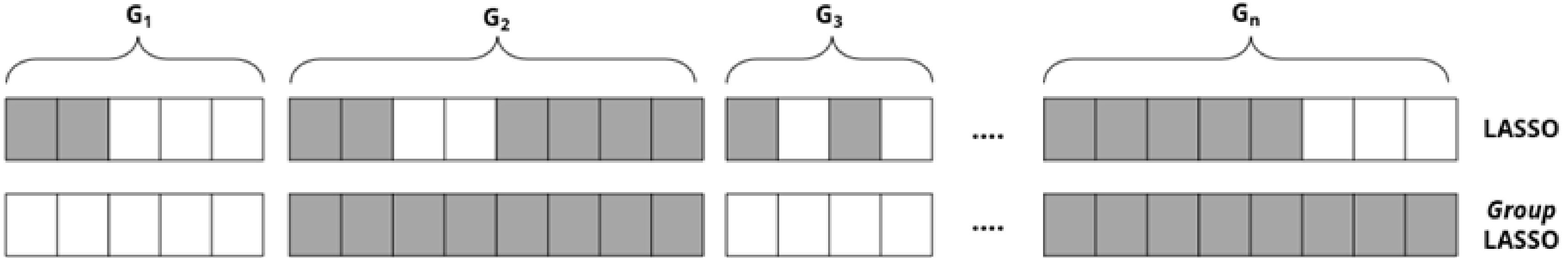
Varians Inflation Factor

Persamaan perhitungan VIF

$$VIF = \frac{1}{1 - R_i^2}$$

koefisien
determinasi

Jika nilai VIF lebih besar dari 10, maka menandakan adanya masalah multikolinearitas.
(Ferré, 2009)



LASSO

Least Absolute Shrinkage and Selection Operator

Teknik pemilihan variabel dengan melakukan penyusutan koefisien dari variabel bebas mendekati nol (**Hastie dkk., 2009**)

Bentuk penaksir parameter LASSO sebagai berikut.

$$\hat{\beta} = \arg \min \frac{1}{2} \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \sum_{k=1}^p x_{ik} \hat{\beta}_k \right)^2 + \lambda \sum_{k=1}^p |\hat{\beta}_k|$$

Group LASSO

Group LASSO merupakan pengembangan metode LASSO (**Yuan & Lin, 2006**)

Group LASSO digunakan untuk mempermudah penyeleksian variabel yang membentuk sebuah kelompok variabel (**Liu & Zhang, 2009**)

Bentuk penaksir parameter Group LASSO sebagai berikut.

$$\hat{\beta} = \arg \min \frac{1}{2} \left\| Y - \sum_{j=1}^k X_j \beta_j \right\|^2 + \lambda \sum_{j=1}^k \|\beta_j\|$$



Penaksir Parameter Regresi Logistik Biner Group LASSO

Penaksir parameter regresi logistik biner Group LASSO sebagai berikut (**Breheny & Huang, 2013**).

$$\hat{\beta} = \arg \min \left(-\frac{1}{n} \left(\sum_{i=1}^n y_i \ln[\pi(x_i)] + (1-y_i) \ln[1-\pi(x_i)] \right) + \lambda \sum_{j=1}^k \|\beta_j\| \right)$$

Algoritma Group Descent

- Tujuan dari algoritma ini adalah untuk membentuk nilai parameter baru
- Dikembangkan pada pemrograman R dalam **paket grpreg**

(Breheny & Huang, 2013)

Validasi Silang Lipat k

- Validasi silang lipat k yang sering digunakan, yaitu dengan $k=5$ atau $k=10$ karena dapat menghasilkan ragam yang rendah

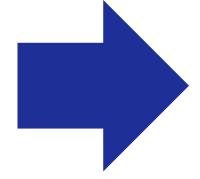
(James dkk., 2013)



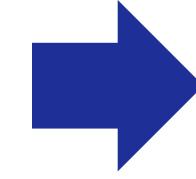
Indeks Pembangunan Manusia

Indikator penting untuk mengukur capaian pembangunan kualitas hidup manusia

IPM diukur dengan tiga indikator, yaitu **dimensi pendidikan, dimensi kesehatan dan dimensi ekonomi.**



<p>Khoiriah dkk. (2017) Sektor lingkungan seperti luas hutan dan perubahan lahan juga berpengaruh terhadap capaian IPM</p>	<p>Latuconsina (2017) Terdapat beberapa aspek yang mempengaruhi capaian IPM, seperti aspek pendidikan, aspek kesehatan, dan aspek kependudukan.</p>
<p>Hayati (2019) Fokusan pemerintah pada sektor pendidikan, kesehatan, dan perlindungan sosial berpengaruh terhadap IPM</p>	<p>Zakaria (2018) Jumlah penduduk, tingkat pengangguran, tingkat kemiskinan dan belanja modal berpengaruh terhadap capaian IPM</p>



Faktor-Faktor yang Mempengaruhi IPM

- Aspek Pendidikan
- Aspek Kesehatan
- Aspek Ekonomi
- Aspek Lingkungan
- Aspek Kependudukan
- Aspek Kesejahteraan Sosial



Kampus
Merdeka
INDONESIA JAYA

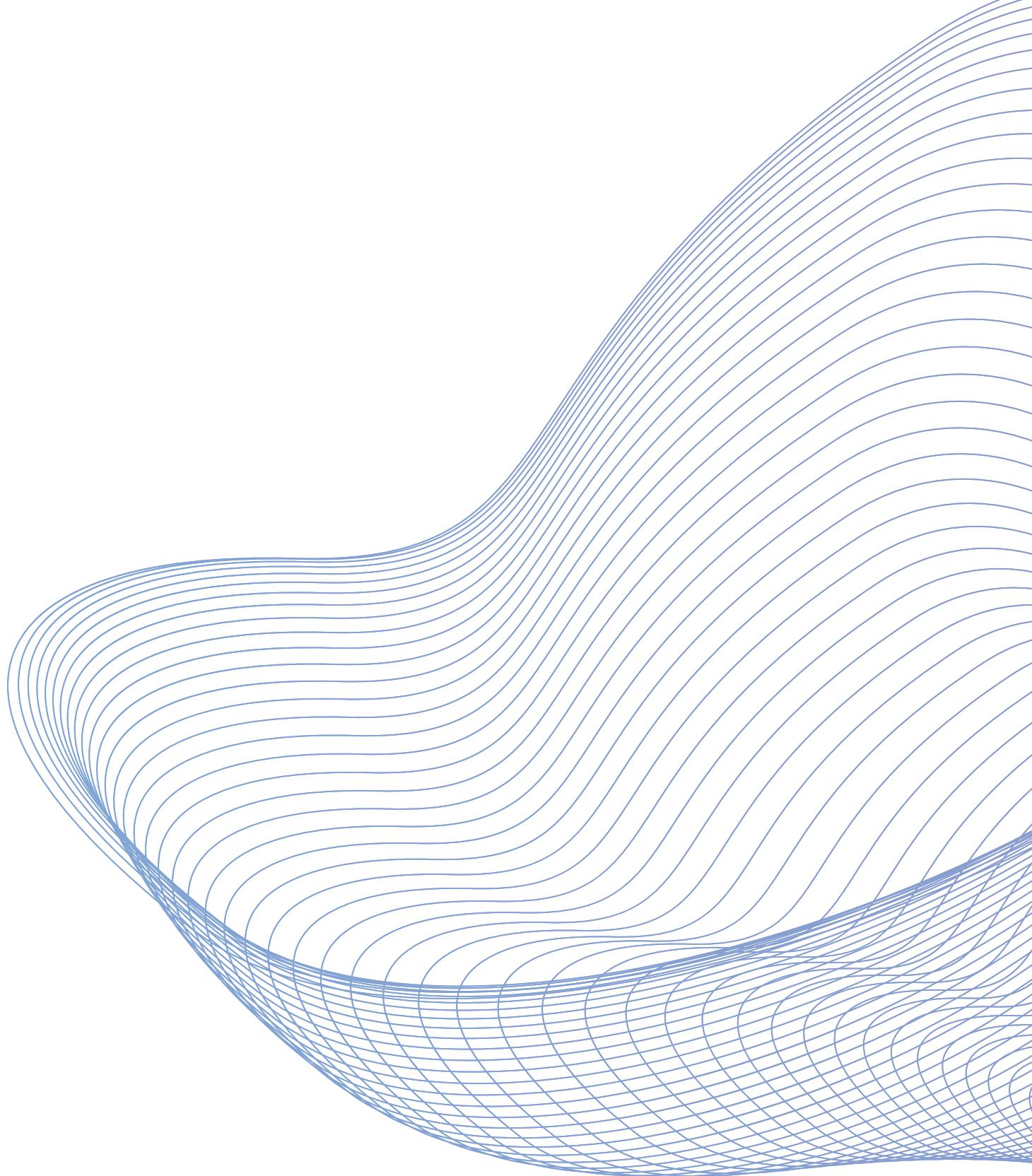
UPI
The Education University

METODOLOGI PENELITIAN

Sumber Data

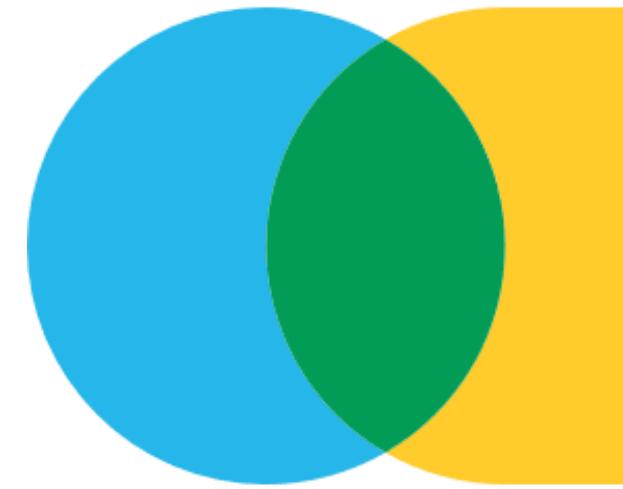
Variabel

Langkah-Langkah





Sumber Data Penelitian



OPEN
DATA
JABAR



BPS JABAR

Data yang digunakan merupakan data sekunder yang berasal dari website Badan Pusat Statistik Jawa Barat dan *Open Data* Jabar

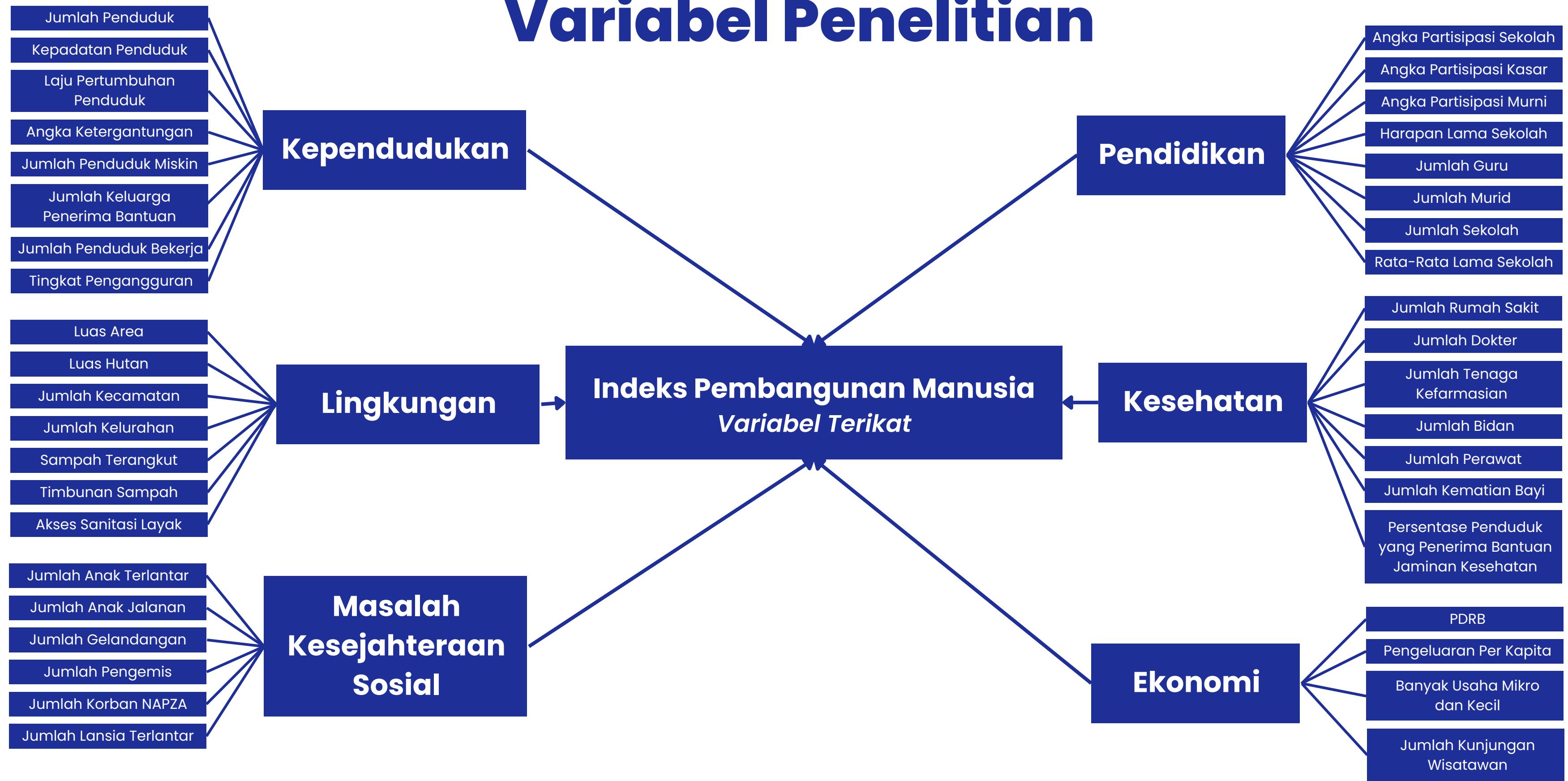
Bentuk Data Penelitian

Kota/Kabupaten	Y	X_1	X_2	X_3	...	X_{40}
1	y_1	$x_{1.1}$	$x_{1.2}$	$x_{1.3}$...	$x_{1.40}$
2	y_2	$x_{2.1}$	$x_{2.2}$	$x_{2.3}$...	$x_{2.40}$
3	y_3	$x_{3.1}$	$x_{3.2}$	$x_{3.3}$...	$x_{3.40}$
4	y_4	$x_{4.1}$	$x_{4.2}$	$x_{4.3}$...	$x_{4.40}$
...
27	y_{27}	$x_{27.1}$	$x_{27.2}$	$x_{27.3}$...	$x_{27.40}$

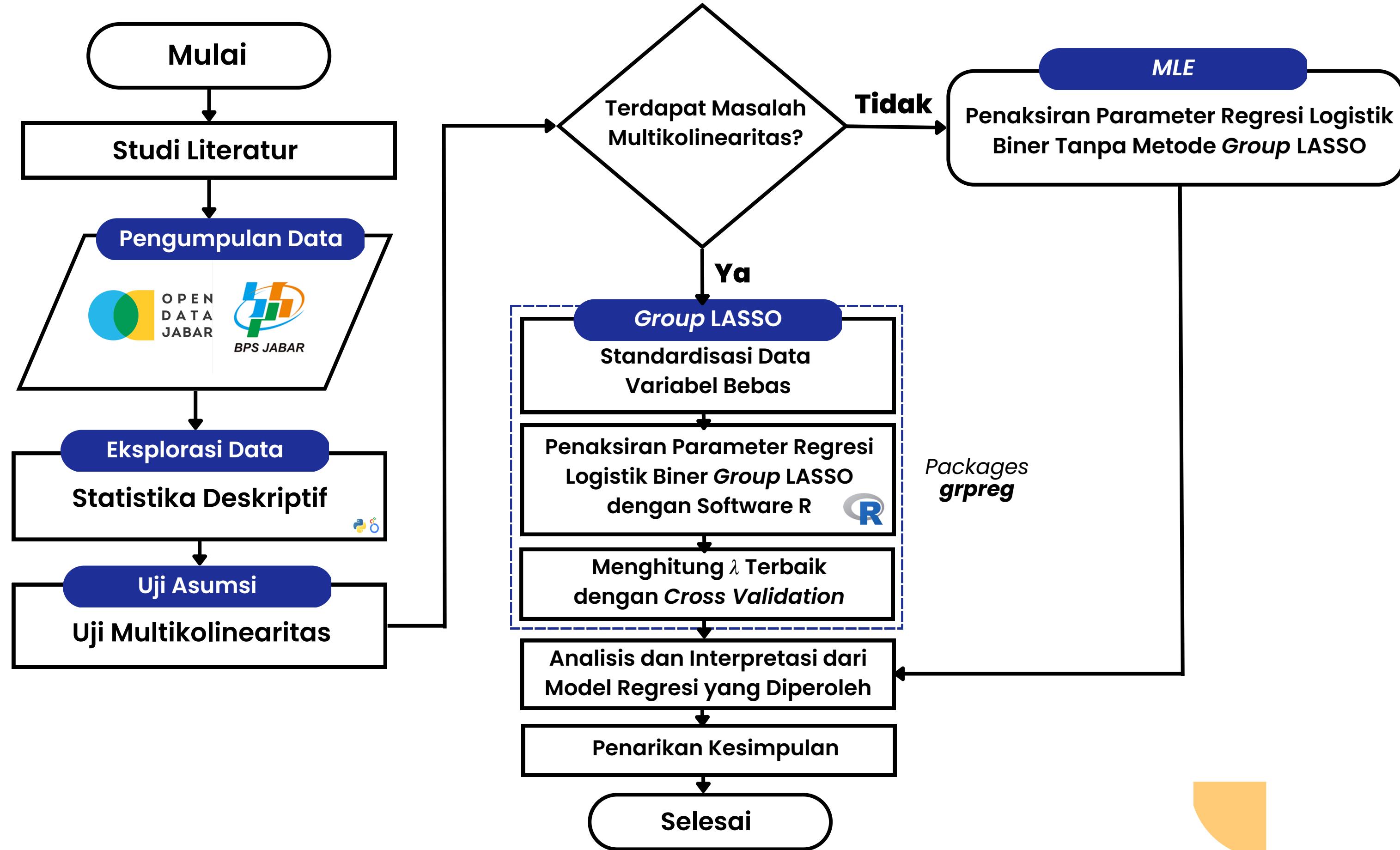
27
Amatan

40
Variabel Bebas

Variabel Penelitian



Langkah-Langkah Penelitian





Kampus
Merdeka
INDONESIA JAYA

UPI
The Education University

HASIL DAN PEMBAHASAN

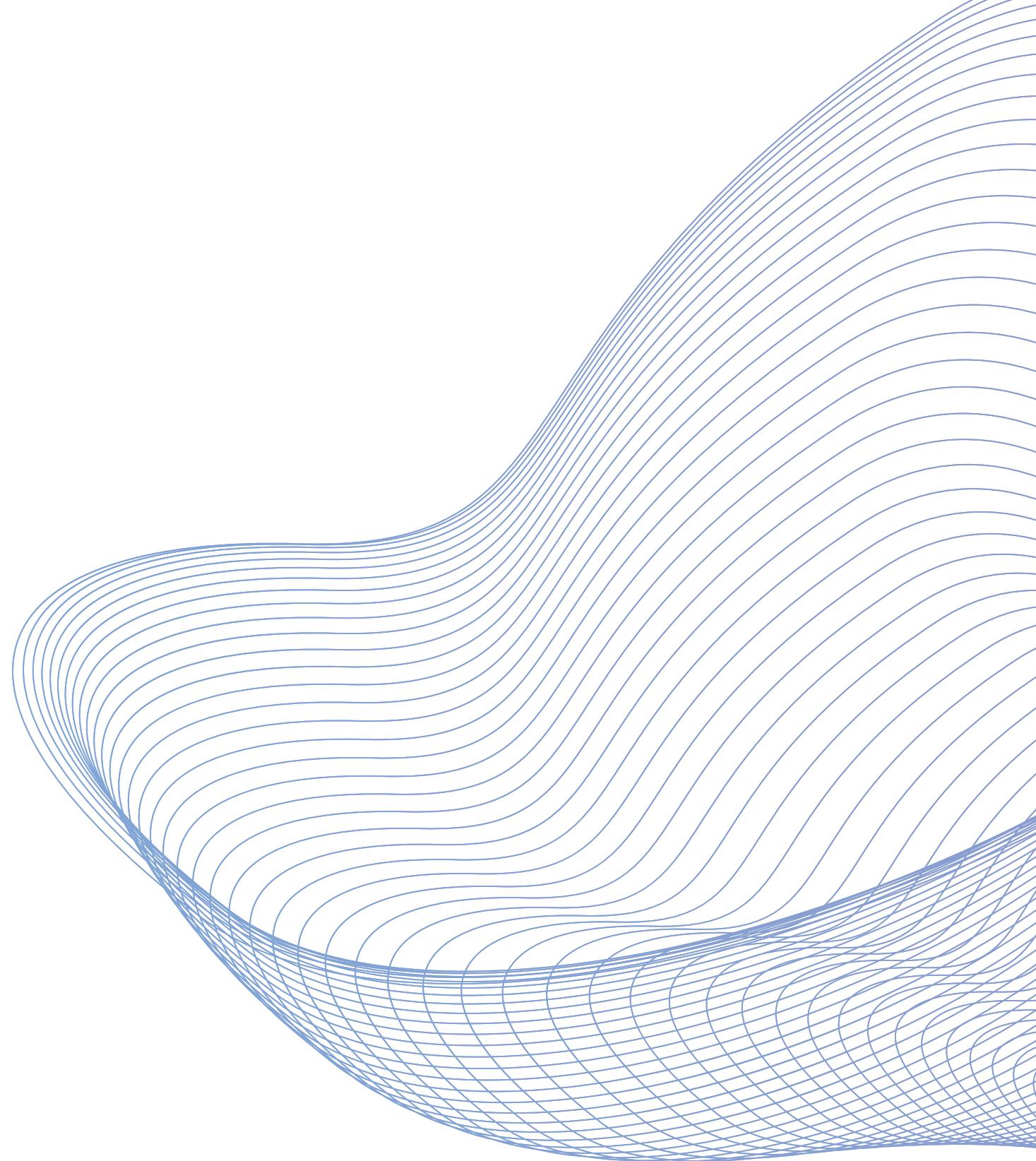
Gambaran Umum

Analisis Deskriptif

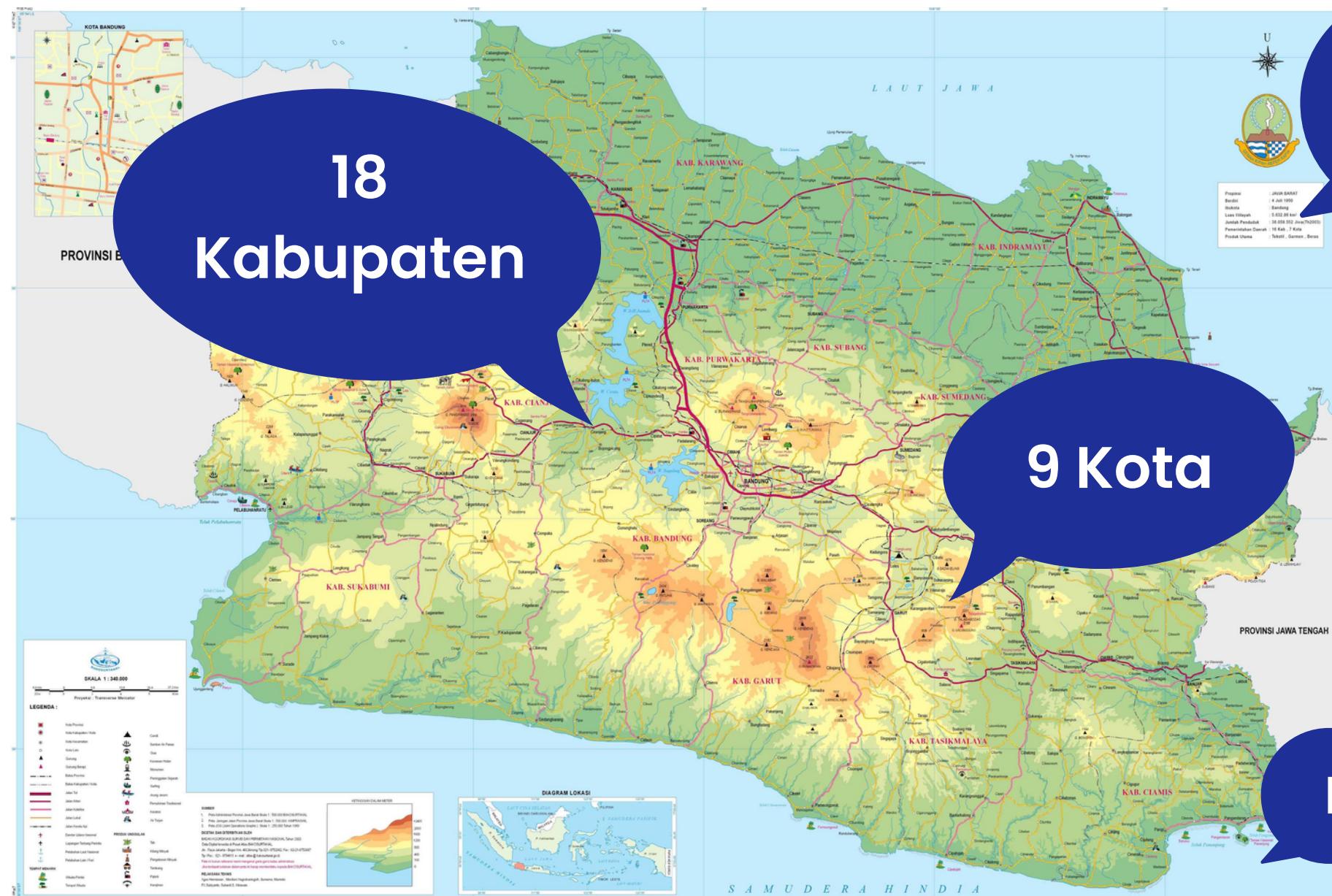
Uji Multikolinearitas

Penerapan Regresi Logistik Biner Group LASSO

Interpretasi



Gambaran Umum Provinsi Jawa Barat



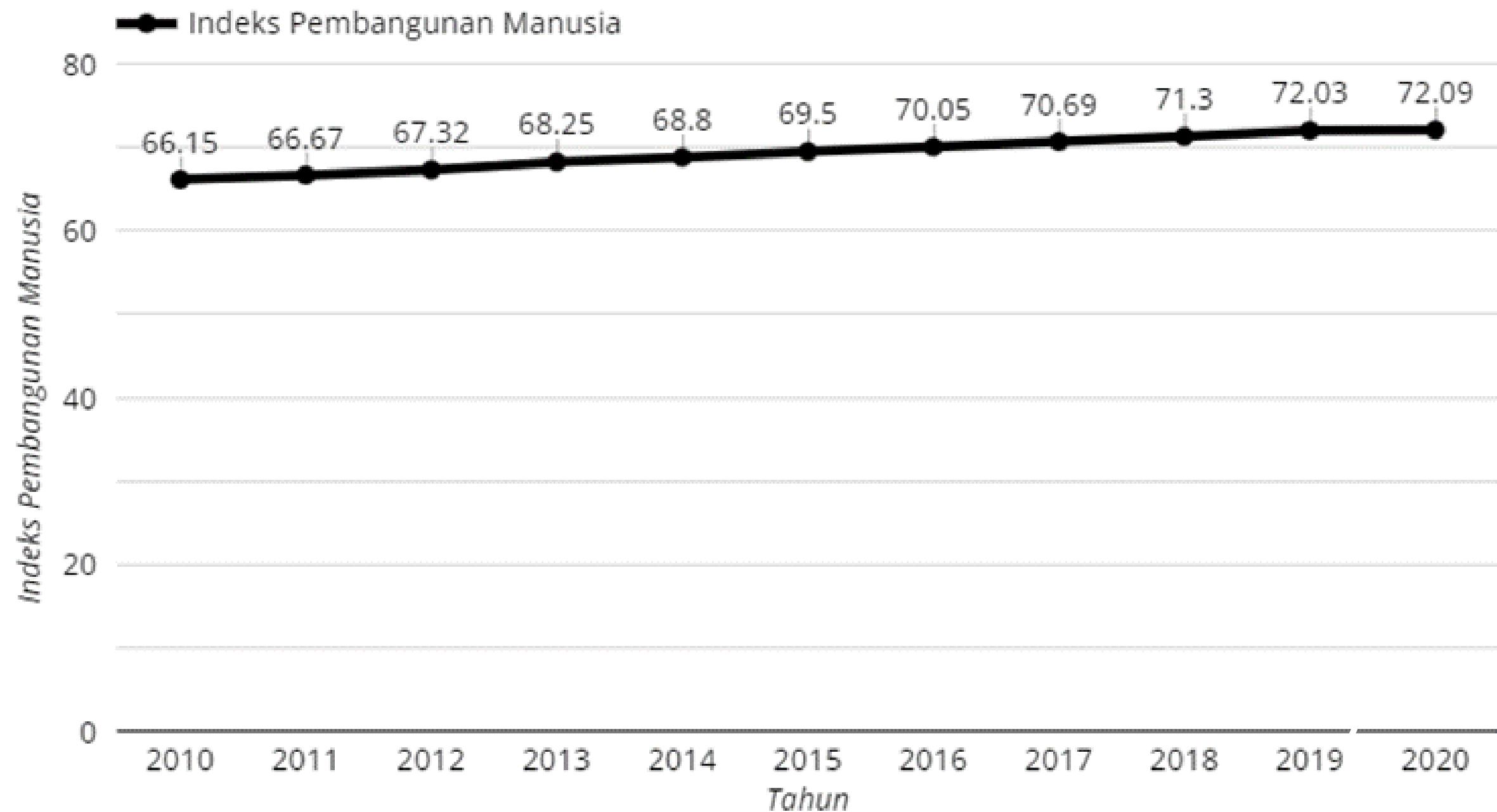
Secara astronomi, terletak pada
 $5^{\circ}50' - 7^{\circ}50'$ LS dan $104^{\circ}48' - 108^{\circ}48'$ BT

Jawa Barat merupakan salah satu provinsi yang berada di Pulau Jawa selain Provinsi DKI Jakarta, Banten, DI Yogyakarta, Jawa Tengah, dan Jawa Timur.

Luas Wilayah 35.378 km^2

Capaian Indeks Pembangunan Manusia

Provinsi Jawa Barat Tahun 2010–2020



Dari tahun ke tahun
capaian IPM Jawa Barat
selalu meningkat

Kualitas hidup manusia di
Jawa Barat masih mengalami
peningkatan dari tahun ke
tahunnya

Capaian Indeks Pembangunan Manusia Kota/Kabupaten di Jawa Barat Tahun 2020

Hijau

IPM Menengah ke Atas
Wilayah dengan capaian
IPM di atas capaian IPM
Jawa Barat

10
Wilayah

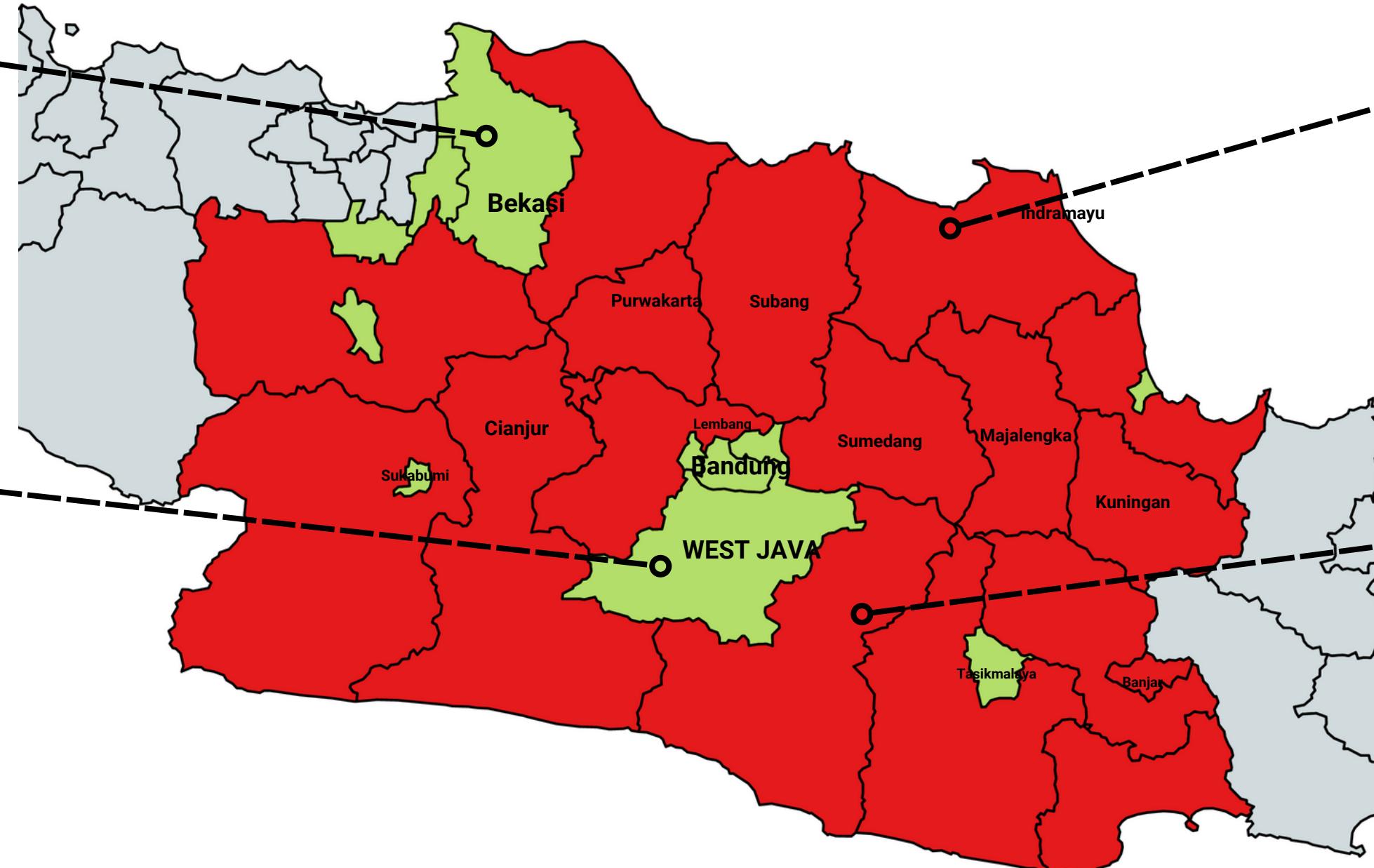
Mayoritas **wilayah Kota**
memiliki IPM Menengah
ke Atas

Merah

IPM Menengah ke Bawah
Wilayah dengan capaian
IPM di bawah capaian
IPM Jawa Barat

17
Wilayah

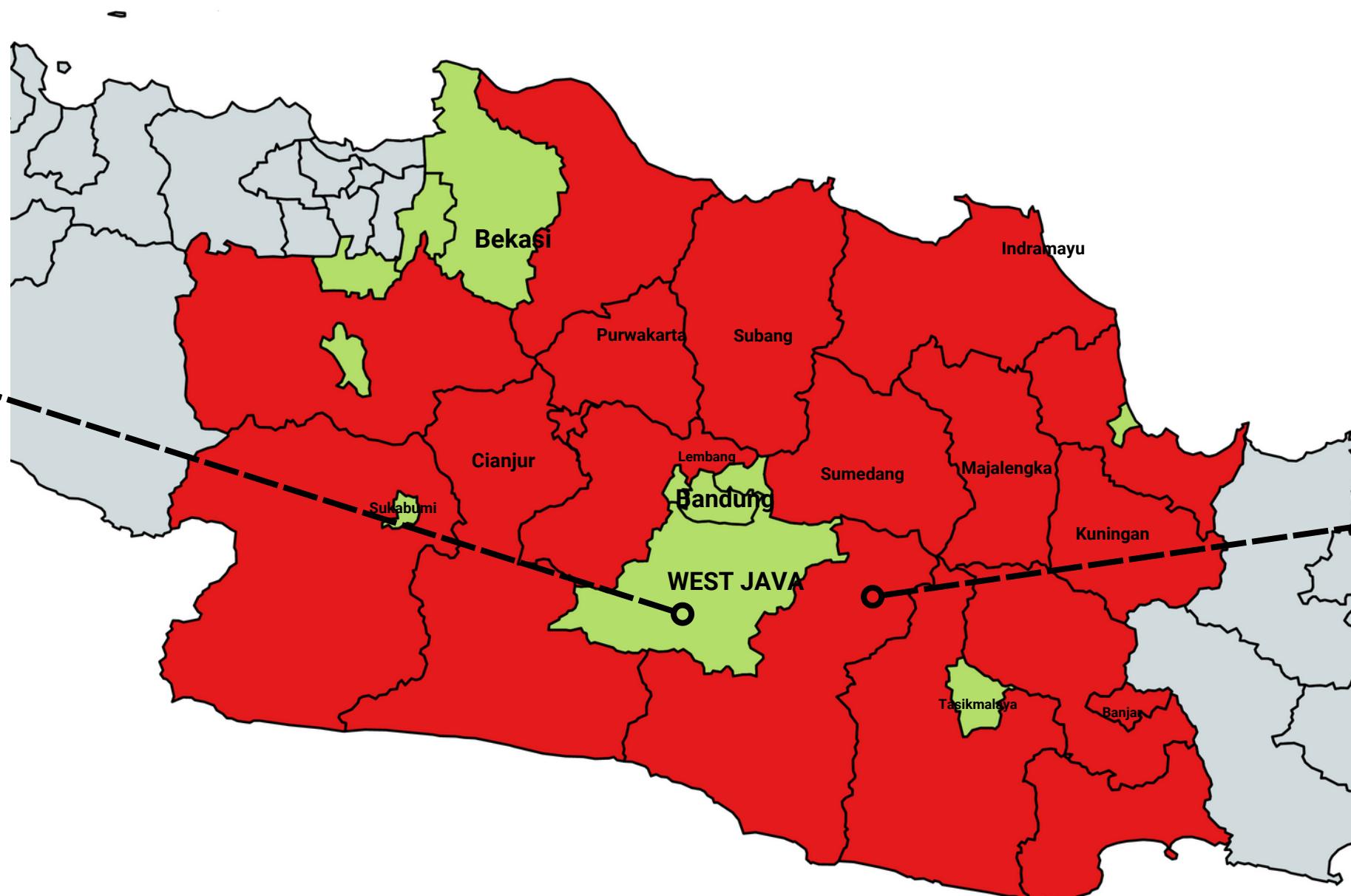
Mayoritas **wilayah**
Kabupaten memiliki IPM
Menengah ke Bawah



Deskripsi Kota/Kabupaten di Jawa Barat Berdasarkan Kelompok Pendidikan

Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Atas

Angka Partisipasi Sekolah	76,56
Angka Partisipasi Kasar	67,11
Angka Partisipasi Murni	90,24
Harapan Lama Sekolah	13,52
Jumlah Guru	14719
Jumlah Murid	294828
Jumlah Sekolah	825
Rata-Rata Lama Sekolah	10,14



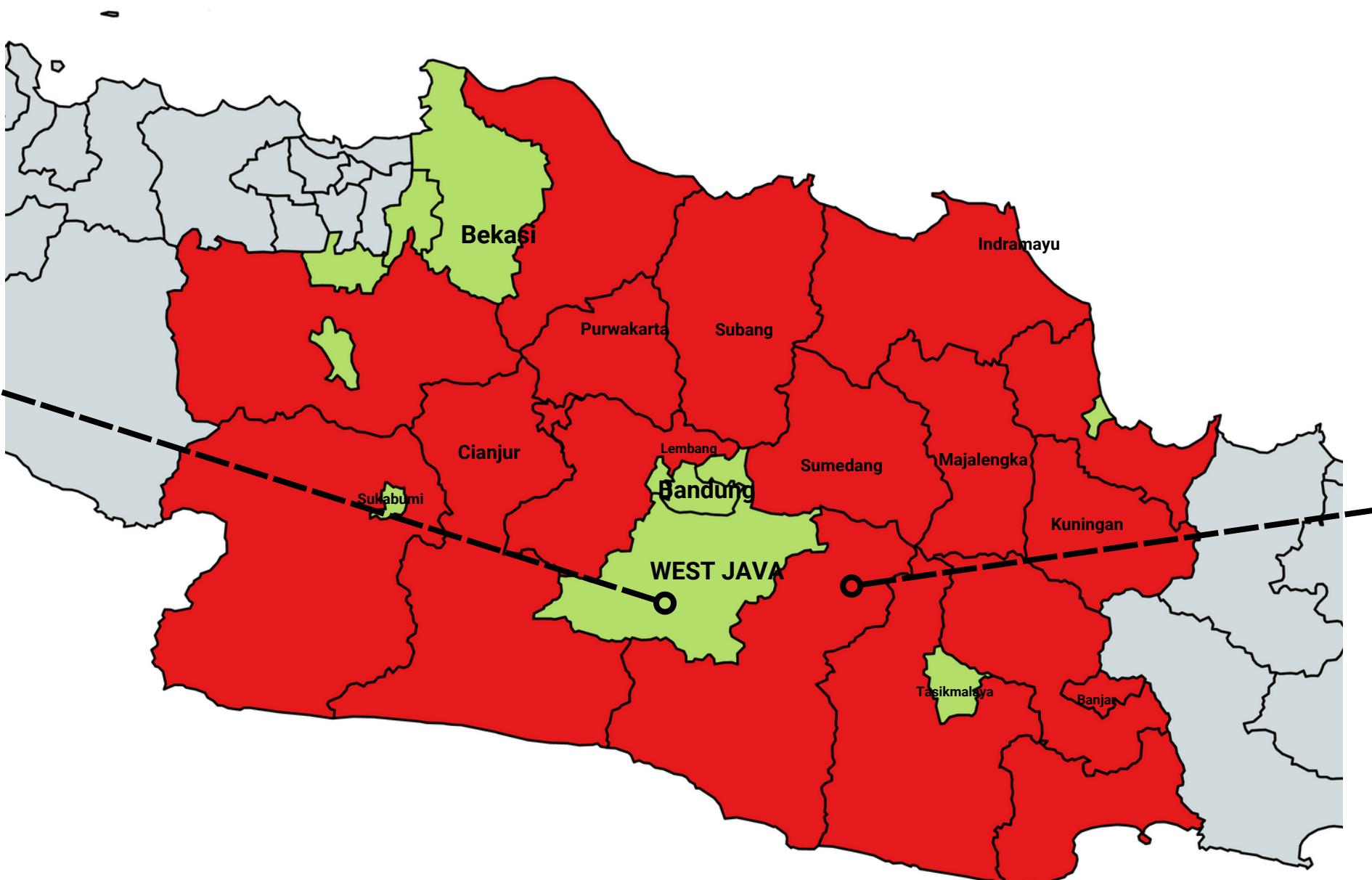
Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Bawah

Angka Partisipasi Sekolah	66,48
Angka Partisipasi Kasar	56,22
Angka Partisipasi Murni	75,81
Harapan Lama Sekolah	12,36
Jumlah Guru	15824
Jumlah Murid	306516
Jumlah Sekolah	1254
Rata-Rata Lama Sekolah	7,60

Deskripsi Kota/Kabupaten di Jawa Barat Berdasarkan Kelompok Kesehatan

Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Atas

Jumlah Rumah Sakit	22
Jumlah Dokter	603
Jumlah Tenaga Kefarmasian	372
Jumlah Bidan	868
Jumlah Perawat	2802
Jumlah Kematian Bayi	61
Persentase Penduduk yang Penerima Bantuan Jaminan Kesehatan	30,5



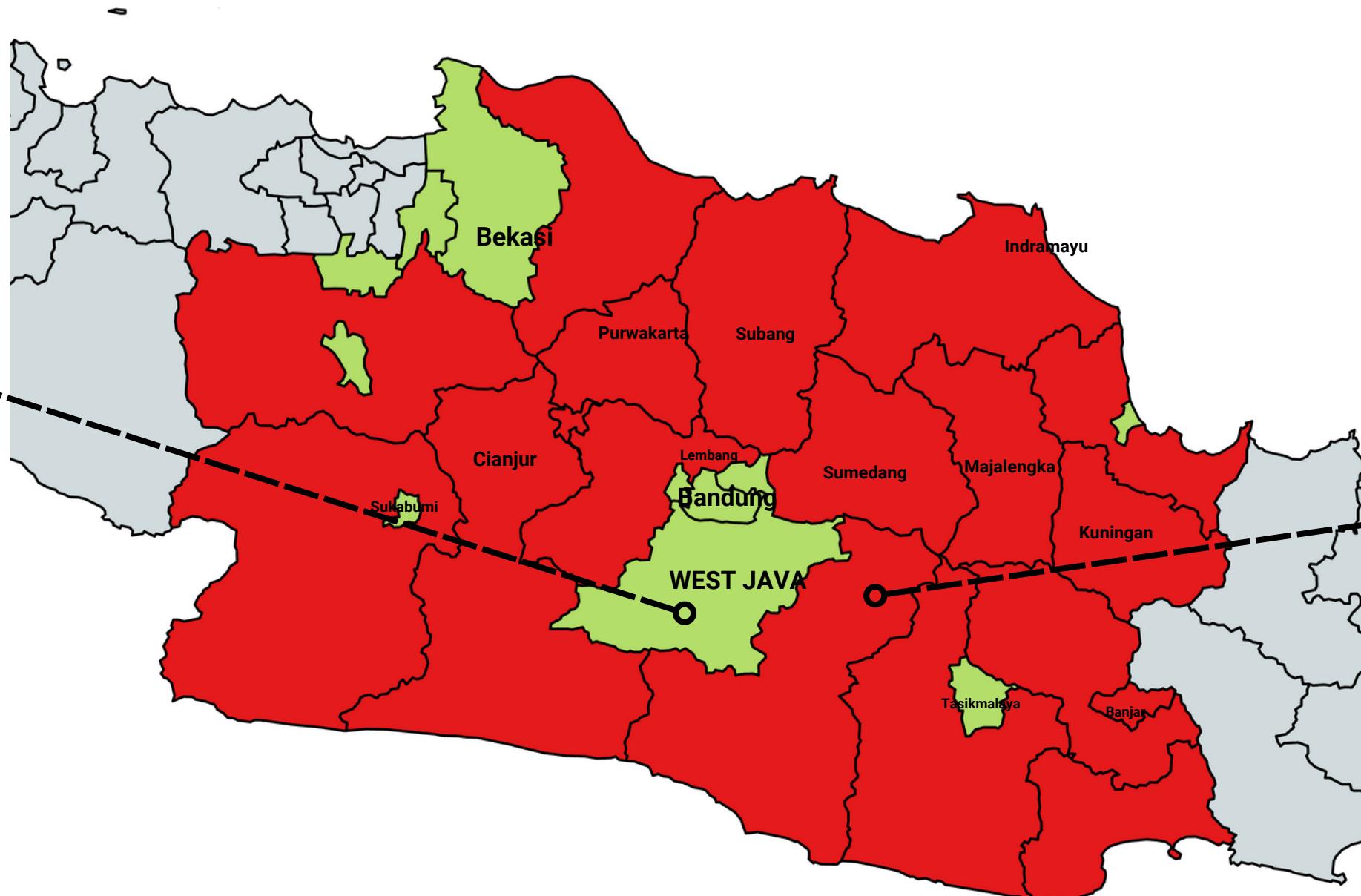
Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Bawah

Jumlah Rumah Sakit	9
Jumlah Dokter	232
Jumlah Tenaga Kefarmasian	143
Jumlah Bidan	975
Jumlah Perawat	1625
Jumlah Kematian Bayi	127
Persentase Penduduk yang Penerima Bantuan Jaminan Kesehatan	34,0

Deskripsi Kota/Kabupaten di Jawa Barat Berdasarkan Kelompok Ekonomi

Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Atas

PDRB	56252,1
Pengeluaran Per Kapita	12603,7
Banyak Usaha Mikro dan Kecil	16014
Jumlah Kunjungan Wisatawan	588932



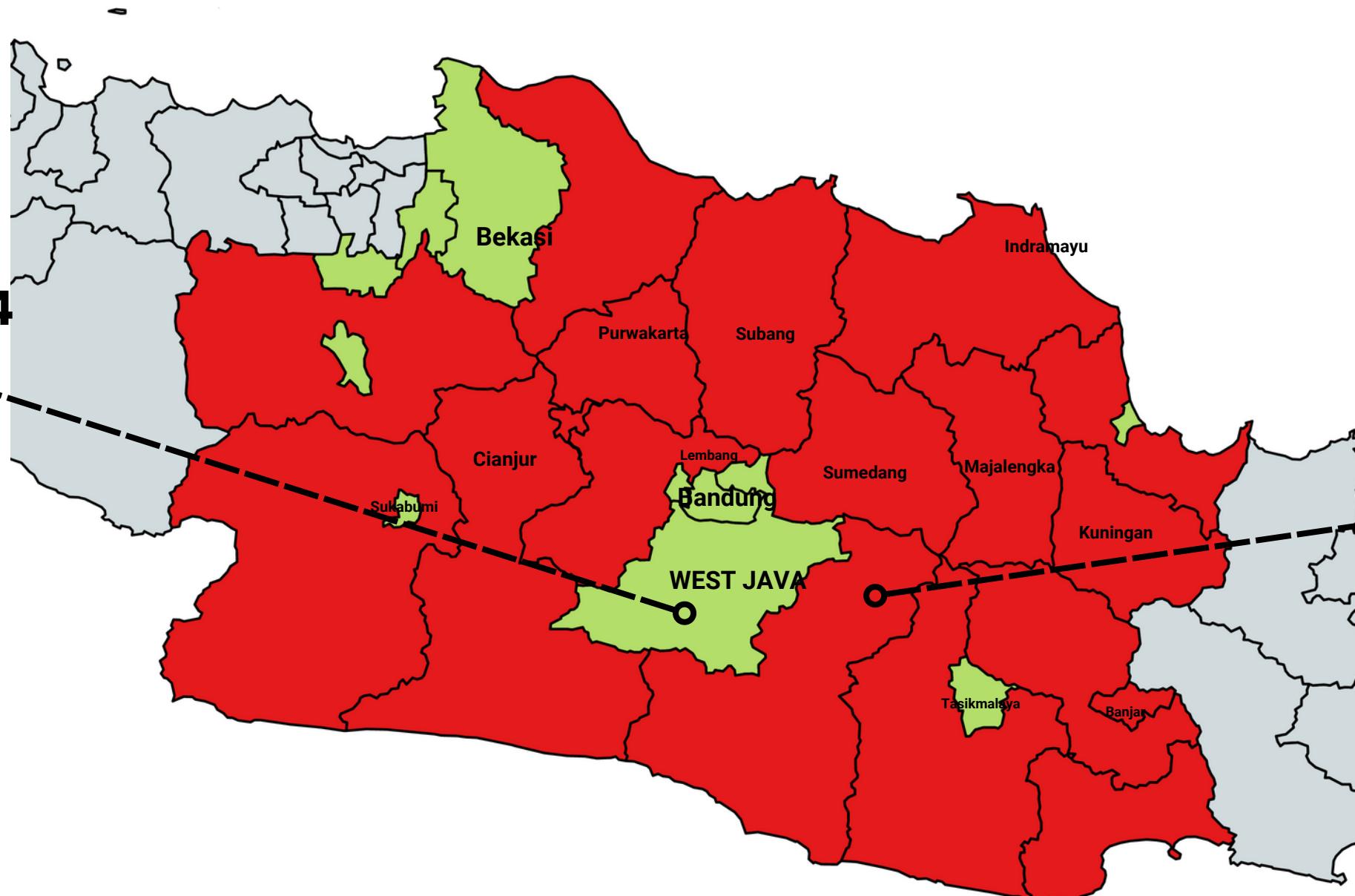
Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Bawah

PDRB	32899
Pengeluaran Per Kapita	9607,471
Banyak Usaha Mikro dan Kecil	27400,24
Jumlah Kunjungan Wisatawan	968610

Deskripsi Kota/Kabupaten di Jawa Barat Berdasarkan Kelompok Lingkungan

Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Atas

Luas Area	398,24
Luas Hutan	3461,54
Jumlah Kecamatan	14
Jumlah Kelurahan	49
Sampah Terangkut	712,317
Timbunan Sampah	1027,6
Akses Sanitasi Layak	77,58



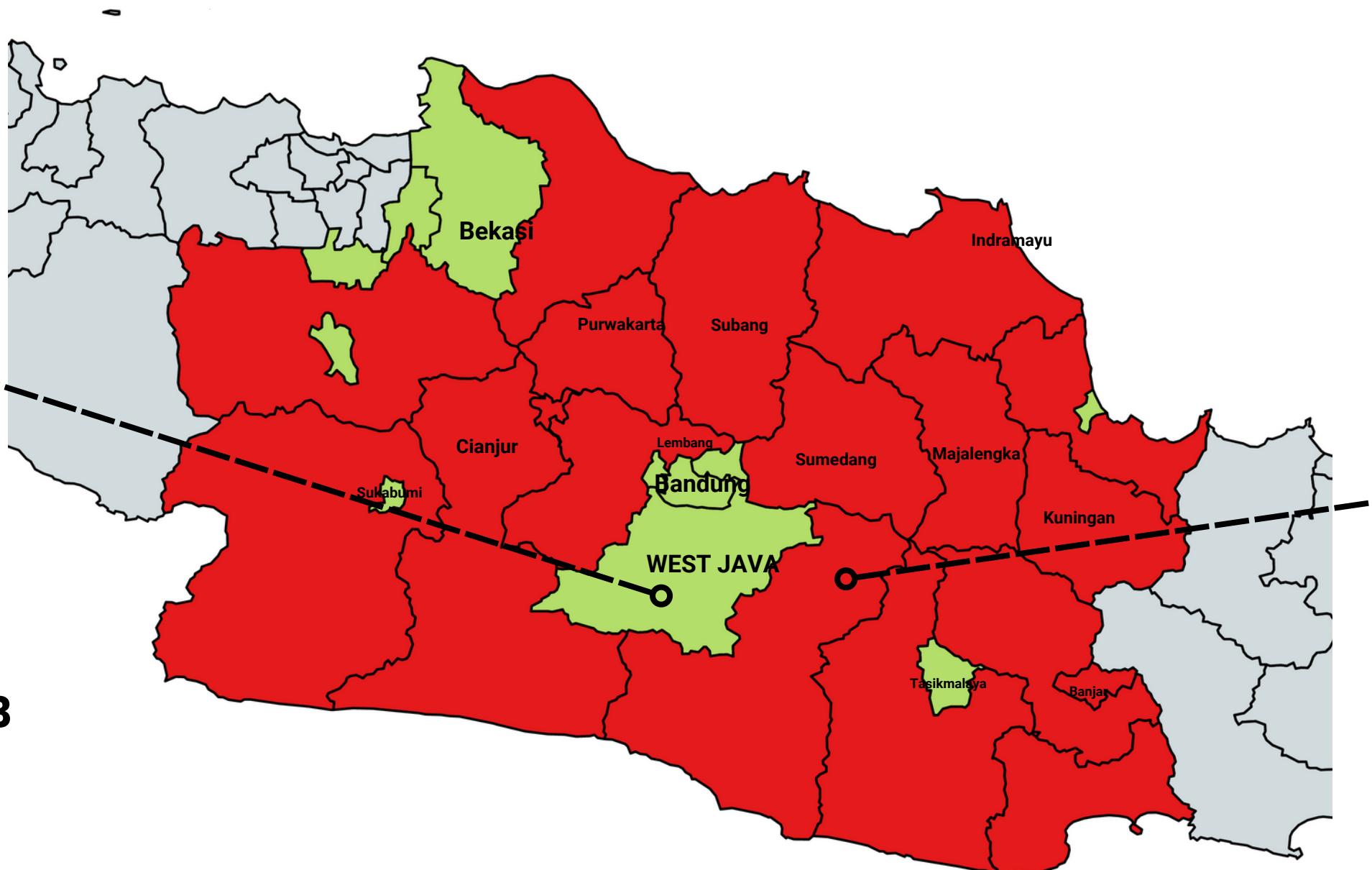
Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Bawah

Luas Area	1846,79
Luas Hutan	34012,5
Jumlah Kecamatan	29
Jumlah Kelurahan	9
Sampah Terangkut	619,469
Timbunan Sampah	885,706
Akses Sanitasi Layak	79,5294

Deskripsi Kota/Kabupaten di Jawa Barat Berdasarkan Kelompok Kependudukan

Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Atas

Jumlah Penduduk	1678823
Kepadatan Penduduk	9226,7
Laju Pertumbuhan Penduduk	1,5
Angka Ketergantungan	42,5
Jumlah Penduduk Miskin	99,3
Jumlah Keluarga Penerima Bantuan	58140,0
Jumlah Penduduk Bekerja	807384,3
Tingkat Pengangguran	10,9



Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Bawah

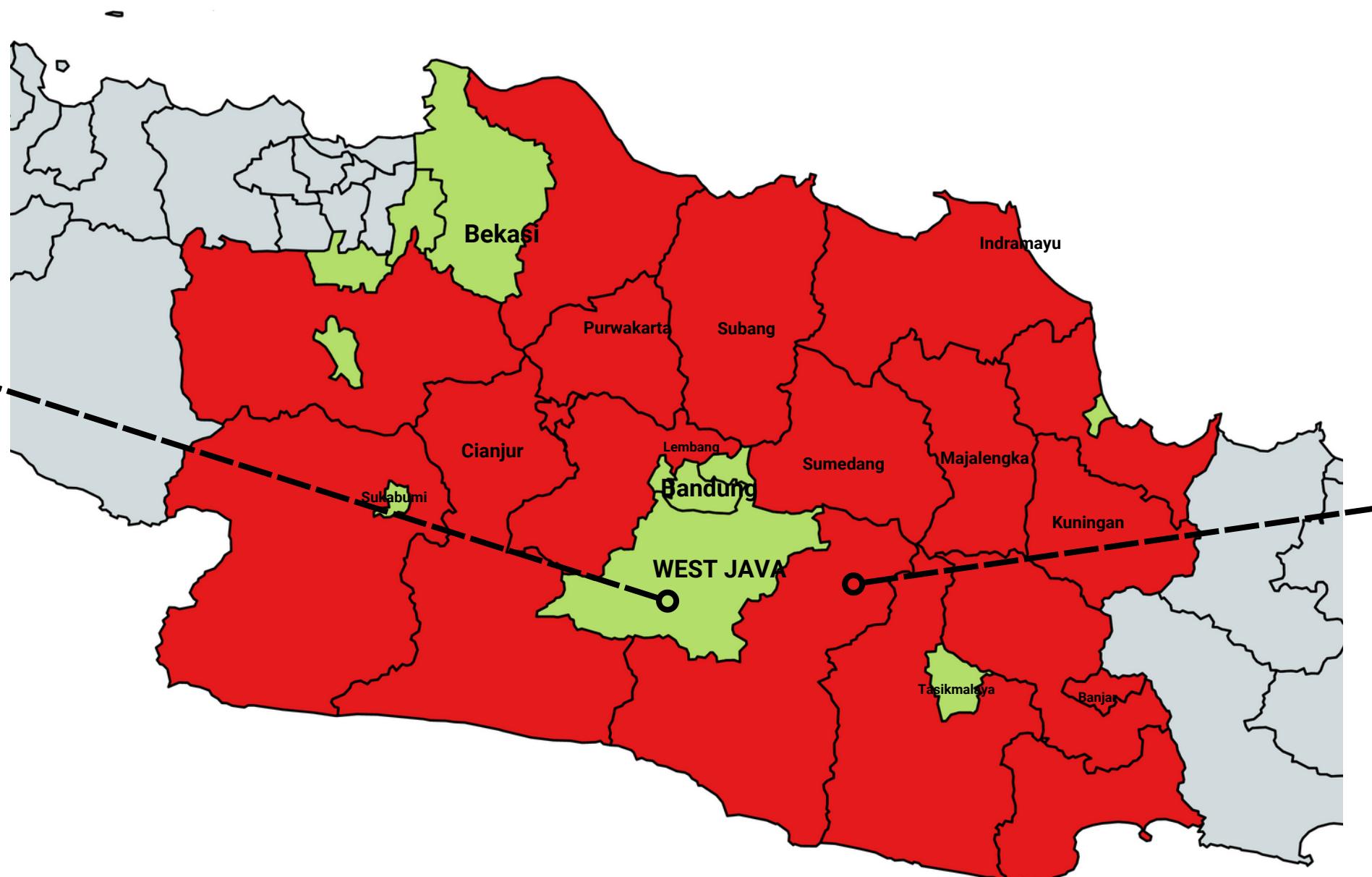
Jumlah Penduduk	1852114
Kepadatan Penduduk	1085,6
Laju Pertumbuhan Penduduk	0,6
Angka Ketergantungan	49,2
Jumlah Penduduk Miskin	172,2
Jumlah Keluarga Penerima Bantuan	120917,4
Jumlah Penduduk Bekerja	800059,5
Tingkat Pengangguran	9,4

Deskripsi Kota/Kabupaten di Jawa Barat

Berdasarkan Kelompok Penyandang Masalah Kesejahteraan Sosial

Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Atas

Jumlah Anak Terlantar	1848
Jumlah Anak Jalanan	790
Jumlah Gelandangan	309
Jumlah Pengemis	180
Jumlah Korban NAPZA	102
Jumlah Lansia Terlantar	5042



Rata-Rata Variabel pada
Wilayah dengan Kategori
IPM Menegah Ke Bawah

Jumlah Anak Terlantar	4868
Jumlah Anak Jalanan	218
Jumlah Gelandangan	230
Jumlah Pengemis	228
Jumlah Korban NAPZA	457
Jumlah Lansia Terlantar	30521



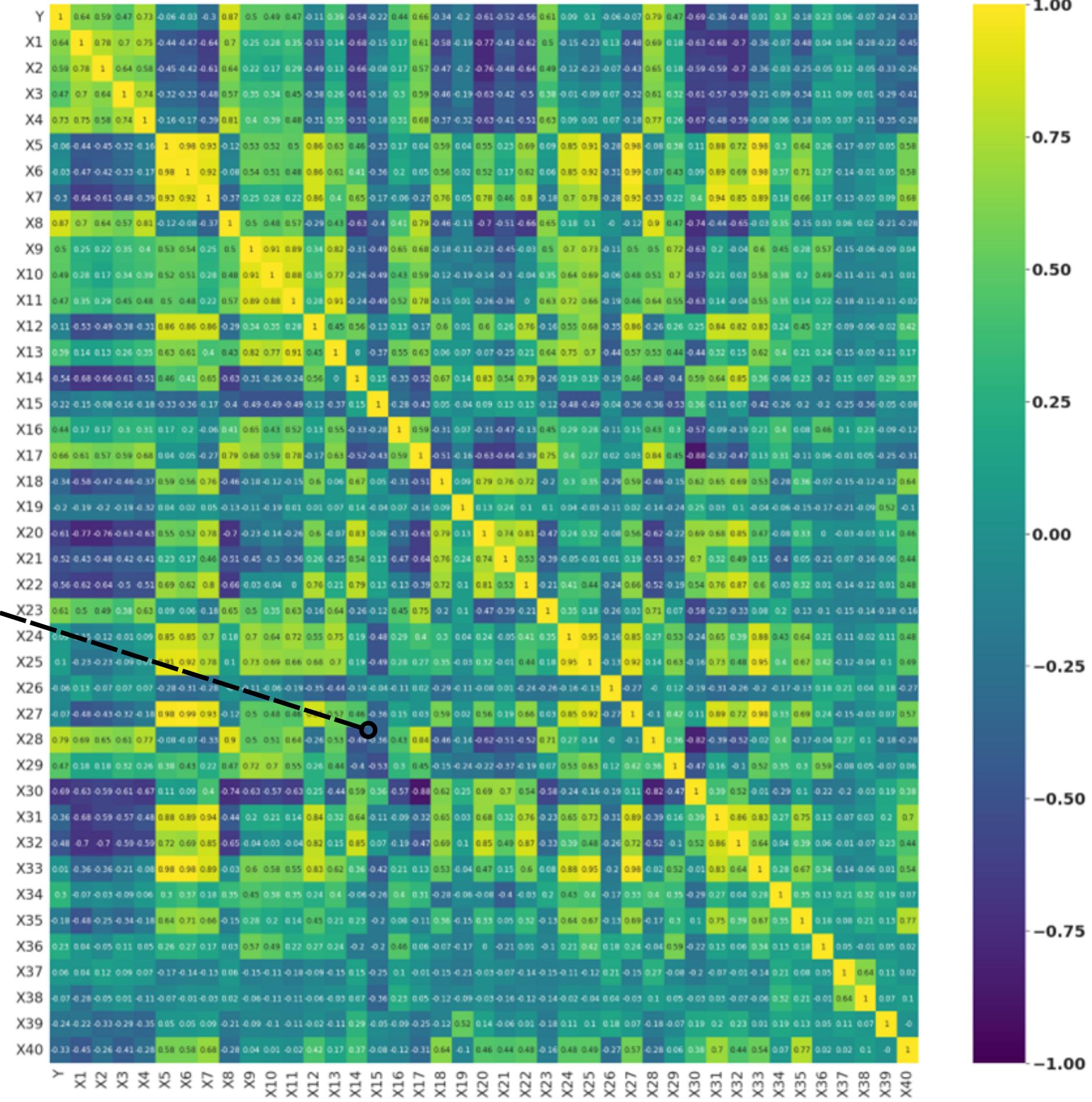
Kampus
Merdeka
INDONESIA JAYA

UPI
The Education University

Hubungan Antarvariabel

Terdapat beberapa variabel bebas yang memiliki **nilai korelasi lebih dari 0,8** atau **korelasi yang sangat kuat**.

Hal ini mengindikasikan adanya permasalahan multikolinearitas variabel bebas pada data penelitian (Senaviratna & Cooray, 2019).



Grafik Heatmap Matriks Korelasi Antarvariabel



Varians Inflation Factor

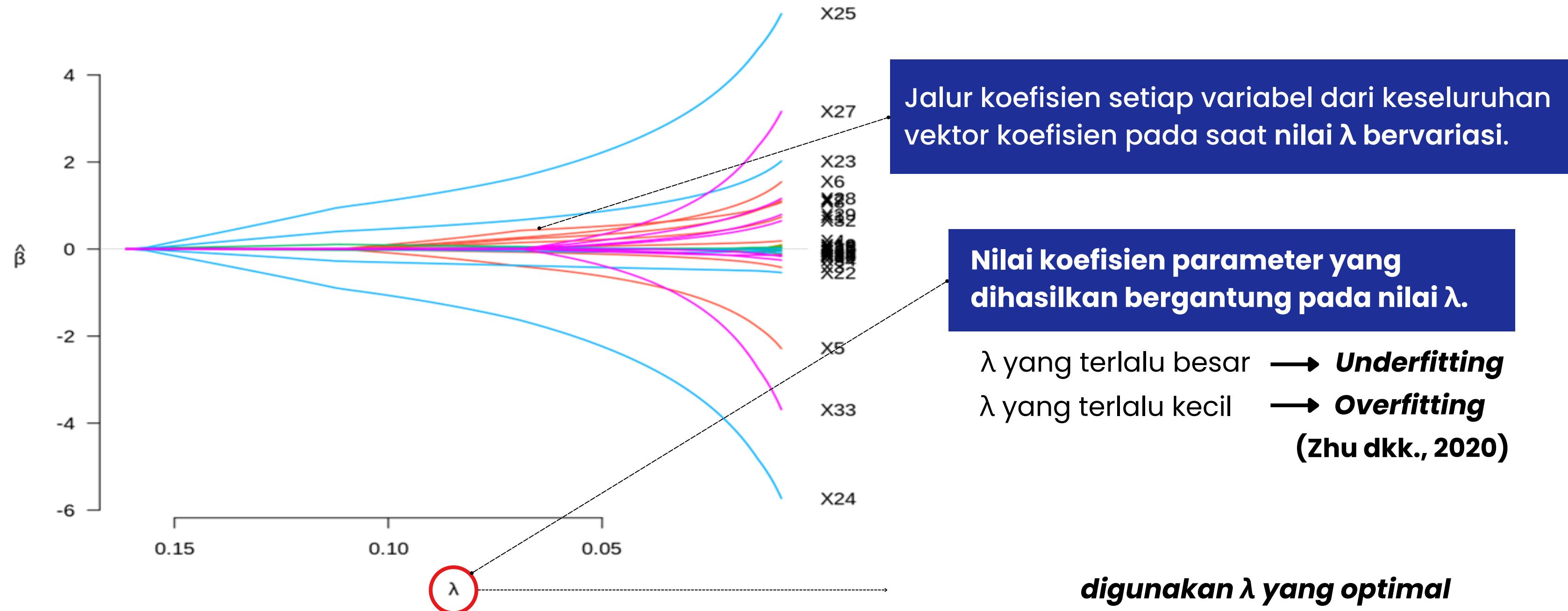
Variabel	VIF	Variabel	VIF	Variabel	VIF	Variabel	VIF
X_1	<i>inf</i>	X_{11}	<i>inf</i>	X_{21}	<i>inf</i>	X_{31}	<i>inf</i>
X_2	<i>inf</i>	X_{12}	<i>inf</i>	X_{22}	<i>inf</i>	X_{32}	<i>inf</i>
X_3	<i>inf</i>	X_{13}	<i>inf</i>	X_{23}	<i>inf</i>	X_{33}	<i>inf</i>
X_4	<i>inf</i>	X_{14}	<i>inf</i>	X_{24}	<i>inf</i>	X_{34}	<i>inf</i>
X_5	<i>inf</i>	X_{15}	<i>inf</i>	X_{25}	<i>inf</i>	X_{35}	<i>inf</i>
X_6	<i>inf</i>	X_{16}	<i>inf</i>	X_{26}	<i>inf</i>	X_{36}	<i>inf</i>
X_7	<i>inf</i>	X_{17}	<i>inf</i>	X_{27}	<i>inf</i>	X_{37}	<i>inf</i>
X_8	<i>inf</i>	X_{18}	<i>inf</i>	X_{28}	<i>inf</i>	X_{38}	<i>inf</i>
X_9	<i>inf</i>	X_{19}	<i>inf</i>	X_{29}	<i>inf</i>	X_{39}	<i>inf</i>
X_{10}	<i>inf</i>	X_{20}	<i>inf</i>	X_{30}	<i>inf</i>	X_{40}	<i>inf</i>

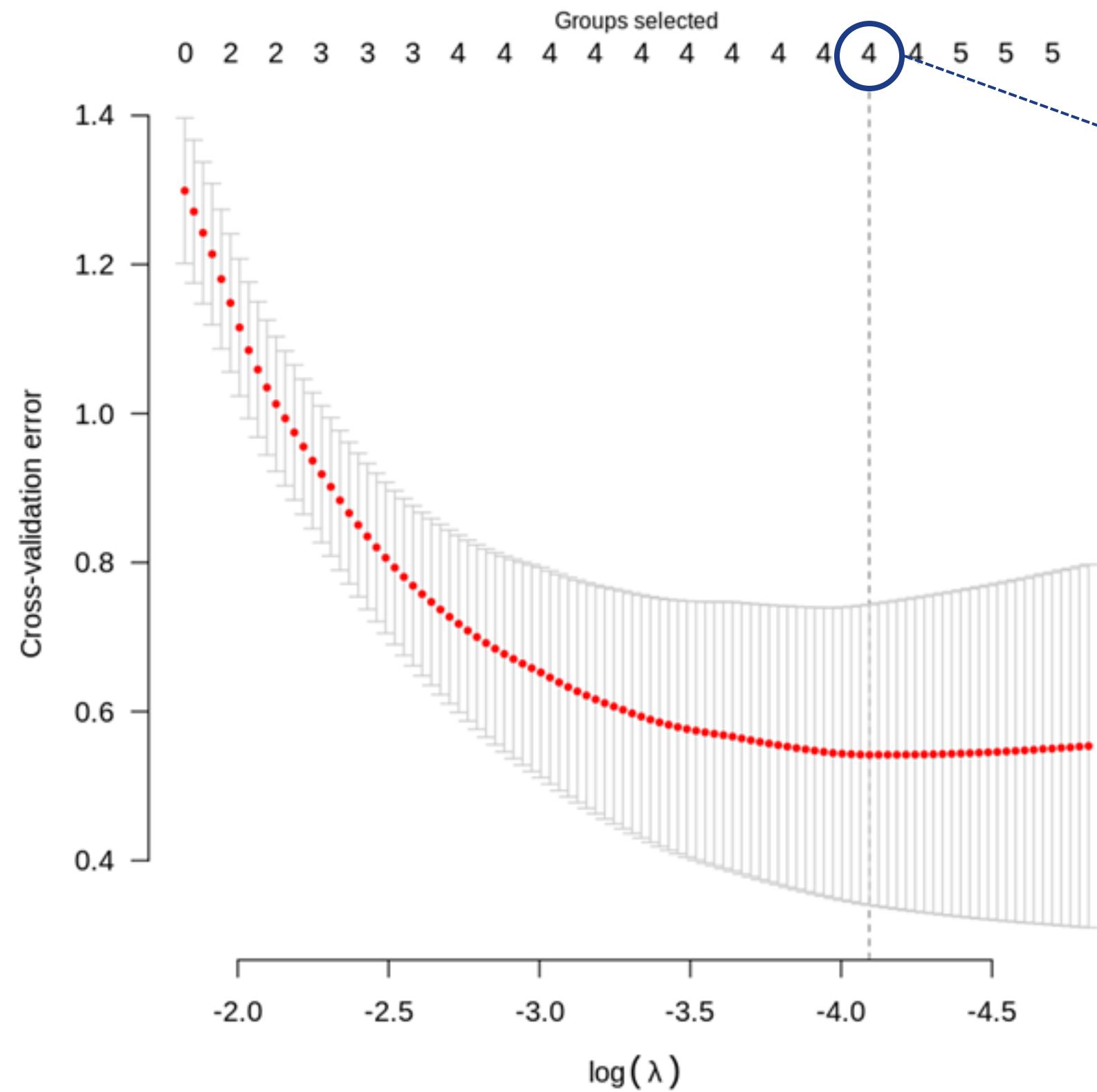
Adanya permasalahan multikolinearitas

Terdapat variabel yang bersesuaian dengan tepat oleh kombinasi linear dari variabel lain disebut juga dengan **korelasi sempurna (Kim, 2019)**.

Hal ini sejalan dengan teori mengenai data berdimensi tinggi yang cenderung mengalami masalah multikolinearitas (**Bielza dkk., 2011**)

Penerapan Regresi Logistik Biner Group LASSO





Validasi Silang

Variabel Ketika λ optimal

Keterangan	Nilai
Nonzero coefficients	27
Nonzero groups	4
kelompok pendidikan	
kelompok ekonomi	
kelompok lingkungan	
kelompok kependudukan	

Kelompok
Variabel
Terpilih

λ yang optimal saat meminimumkan
cross-validation error sebesar **0,0167**

Koefisien Parameter Saat Nilai λ Optimal

Variabel	Koefisien	Variabel	Koefisien	Variabel	Koefisien
(Intercept)	-0,998	X_{14}	0	X_{28}	0,8146
X_1	0,5089	X_{15}	0	X_{29}	0,5302
X_2	-0,05269	X_{16}	0,0211	X_{30}	-0,08243
X_3	-0,2923	X_{17}	0,02575	X_{31}	-0,14
X_4	0,1417	X_{18}	0,01115	X_{32}	0,4327
X_5	-1,59	X_{19}	-0,01526	X_{33}	-2,303
X_6	1,075	X_{20}	-0,07123	X_{34}	-0,1747
X_7	0,794	X_{21}	-0,06117	X_{35}	0
X_8	0,8525	X_{22}	-0,4945	X_{36}	0
X_9	0	X_{23}	1,551	X_{37}	0
X_{10}	0	X_{24}	-4,411	X_{38}	0
X_{11}	0	X_{25}	4,217	X_{39}	0
X_{12}	0	X_{26}	-0,003971	X_{40}	0
X_{13}	0	X_{27}	1,987		

Bentuk persamaan Regresi Logistik Biner Group LASSO dalam bentuk logit sebagai berikut.

$$\begin{aligned}
 g(x) = & -0,998 + 0,5089X_1 - 0,05269X_2 - 0,2923X_3 \\
 & + 0,1417X_4 - 1,59X_5 + 1,075X_6 + 0,794X_7 \\
 & + 0,8525X_8 + 0,0211X_{16} + 0,02575X_{17} \\
 & + 0,01115X_{18} - 0,01526X_{19} - 0,071X_{20} \\
 & - 0,061X_{21} - 0,4945X_{22} + 1,551X_{23} \\
 & - 4,411X_{24} + 4,217X_{25} - 0,0039X_{26} \\
 & + 1,987X_{27} + 0,814X_{26} + 0,53X_{29} \\
 & - 0,08243X_{30} - 0,14X_{31} + 0,432X_{32} \\
 & - 2,303X_{33} - 0,1747X_{34}
 \end{aligned}$$



Interpretasi Koefisien Model Regresi Logistik Biner **Group LASSO**

Setiap kenaikan **Variabel Bebas** sebesar 1 **satuan**, maka IPM Kota/Kabupaten di Jawa Barat berpeluang untuk tergolong IPM menengah ke atas **Nilai Odds Ratio** kali lipat dibandingkan dengan IPM menengah ke bawah.

Interpretasi Variabel Kelompok Pendidikan

Variabel	Penaksir Parameter	Odds Ratio
Angka Partisipasi Sekolah	0,5089	1,66
Angka Partisipasi Murni	-0,05269	0,95
Angka Partisipasi Kasar	-0,2923	0,75
Harapan Lama Sekolah	0,1417	1,15
Jumlah Guru	-1,59	0,20
Jumlah Murid	1,075	2,93
Jumlah Sekolah	0,794	2,21
Rata-Rata Lama Sekolah	0,8525	2,35

■ Temuan yang serupa

- **Taufik (2022)**
Akses terhadap pelayanan pendidikan menjadi salah satu upaya yang dapat meningkatkan capaian IPM
- **Fitriyah dkk. (2021)**
Rata-Rata Lama Sekolah dan Harapan Lama Sekolah memiliki pengaruh yang positif pada capaian IPM
- **Susiani dan Abadiah (2021)**
Kualitas guru menjadi penentu dalam peningkatan mutu pendidikan



Interpretasi Variabel Kelompok Lingkungan

Variabel	Penaksir Parameter	Odds Ratio
Luas Area	-0,07123	0,93
Luas Hutan	-0,06117	0,94
Jumlah Kecamatan	-0,4945	0,61
Jumlah Kelurahan	1,551	4,71
Sampah Terangkut ke TPA	-4,411	0,01
Timbunan Sampah	4,217	67,83
Persentase Rumah Tangga yang Memiliki Sanitasi Layak	-0,003971	0,996

■ Temuan yang serupa

Kualitas lingkungan yang tertata dengan baik tentunya akan berdampak baik pada setiap sektor pembangunan. **(Li & Xu, 2021)**

- **Andriansyah dkk. (2022)**
Peningkatan kondisi lingkungan tempat tinggal akan meningkatkan indeks pembangunan manusia.



Interpretasi Variabel Kelompok Ekonomi

Variabel	Penaksir Parameter	Odds Ratio
PDRB	0,0211	1,02
Pengeluaran per Kapita	0,02575	1,03
Banyaknya Usaha Mikro dan Kecil	0,01115	1,01
Banyaknya Kunjungan Wisatawan	-0,01526	0,98

■ Temuan yang serupa

- **Imron (2018)**

Pertumbuhan ekonomi meningkatkan pendapatan masyarakat, produksi umum barang dan jasa, serta Produk Domestik Bruto suatu wilayah, yang secara langsung dapat meningkatkan capaian IPM.

- **Mirza (2011)**

Tingkat pengeluaran pemerintah menggambarkan besarnya peran pemerintah dalam pembangunan manusia.



Interpretasi Variabel Kelompok Kependudukan

■ Temuan yang serupa

- **Antara dan Suryana (2020)**

Jumlah dan persebaran penduduk serta kepadatan penduduk mempengaruhi persebaran dan kualitas sarana atau prasarana yang berkaitan dengan pembangunan sumber daya manusia.

- **Zuhairoh (2018)**

Angka ketergantungan yang tinggi menyebabkan banyak tanggungan yang besar sehingga menurunkan indeks pembangunan manusia

- **Trisno dkk. (2022)**

Tingginya jumlah penduduk miskin akan menurunkan tingkat pembangunan manusia

- **Zakaria (2018)**

Pengangguran mengurangi pendapatan masyarakat sehingga berdampak pada penurunan IPM suatu wilayah

Variabel	Penaksir Parameter	Odds Ratio
Jumlah Penduduk	1,987	7,294
Kepadatan Penduduk	0,8146	2,258
Laju Pertumbuhan Penduduk	0,5302	1,699
Angka Ketergantungan	-0,08243	0,921
Jumlah Penduduk Miskin	-0,14	0,869
Jumlah Keluarga Penerima Bantuan	0,4327	1,541
Jumlah Penduduk Bekerja	-2,303	0,100
Tingkat Pengangguran	-0,1747	0,840



Ketepatan Klasifikasi Model

Amatan	Prediksi		Total	APER
	IPM Menengah ke Bawah (0)	IPM Menengah ke Atas (1)		
IPM Menengah ke Bawah (0)	17	0	17	
IPM Menengah ke Atas (1)	0	10	10	0.00 %
Total	17	10	27	

Model regresi logistik biner Group LASSO pada penelitian ini mampu memprediksi dengan tepat pengamatan sebesar 100%.

Akan tetapi,

- Group LASSO hanya melakukan pemilihan variabel pada tingkat kelompok variabel.
- Sulitnya memberikan standard error yang akurat karena limit distribusi penaksir LASSO yang sangat rumit (Chatterjee & Lahiri, 2011).



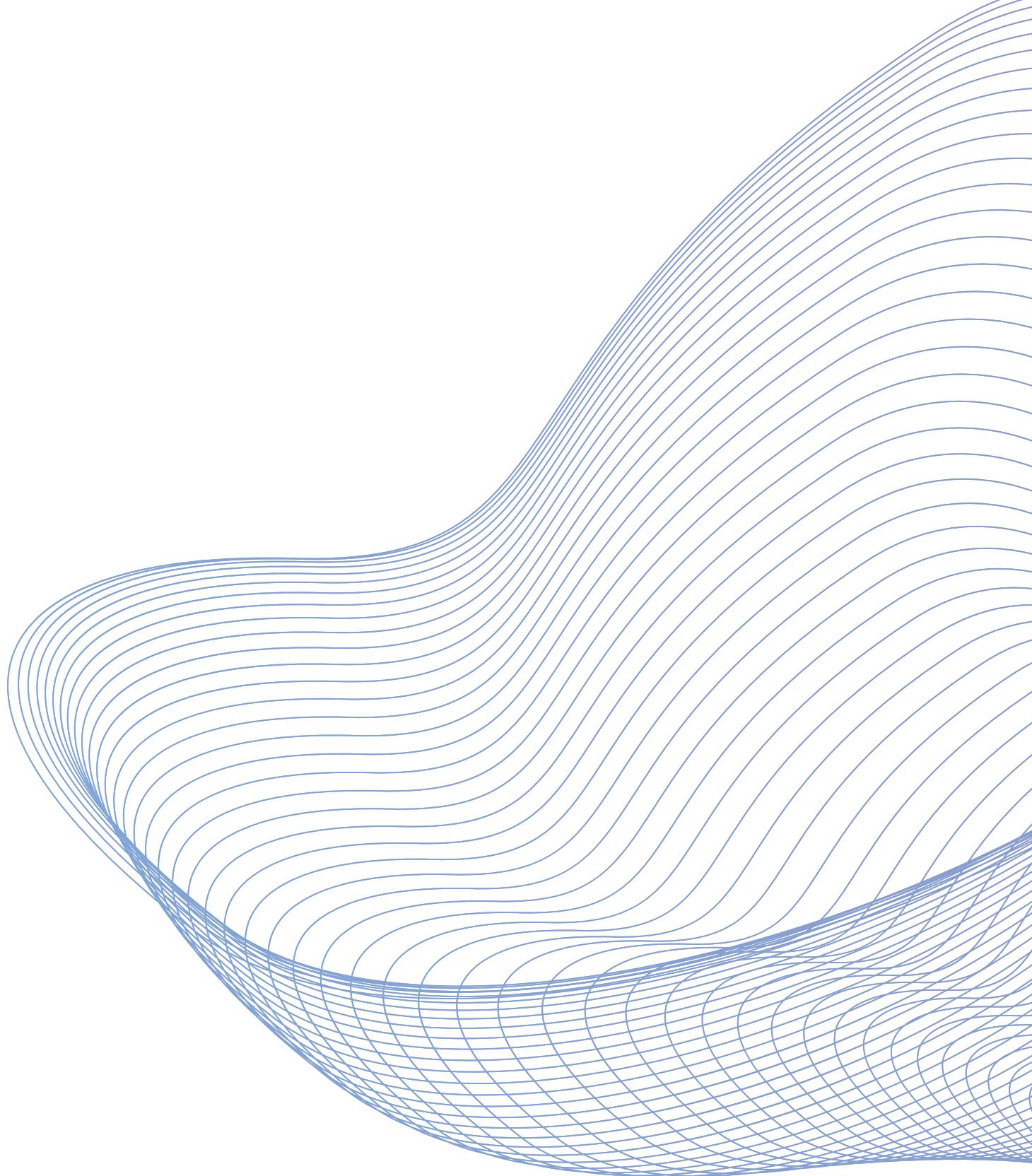
Kampus
Merdeka
INDONESIA JAYA

UPI
The Education University

KESIMPULAN DAN SARAN

Kesimpulan

Saran

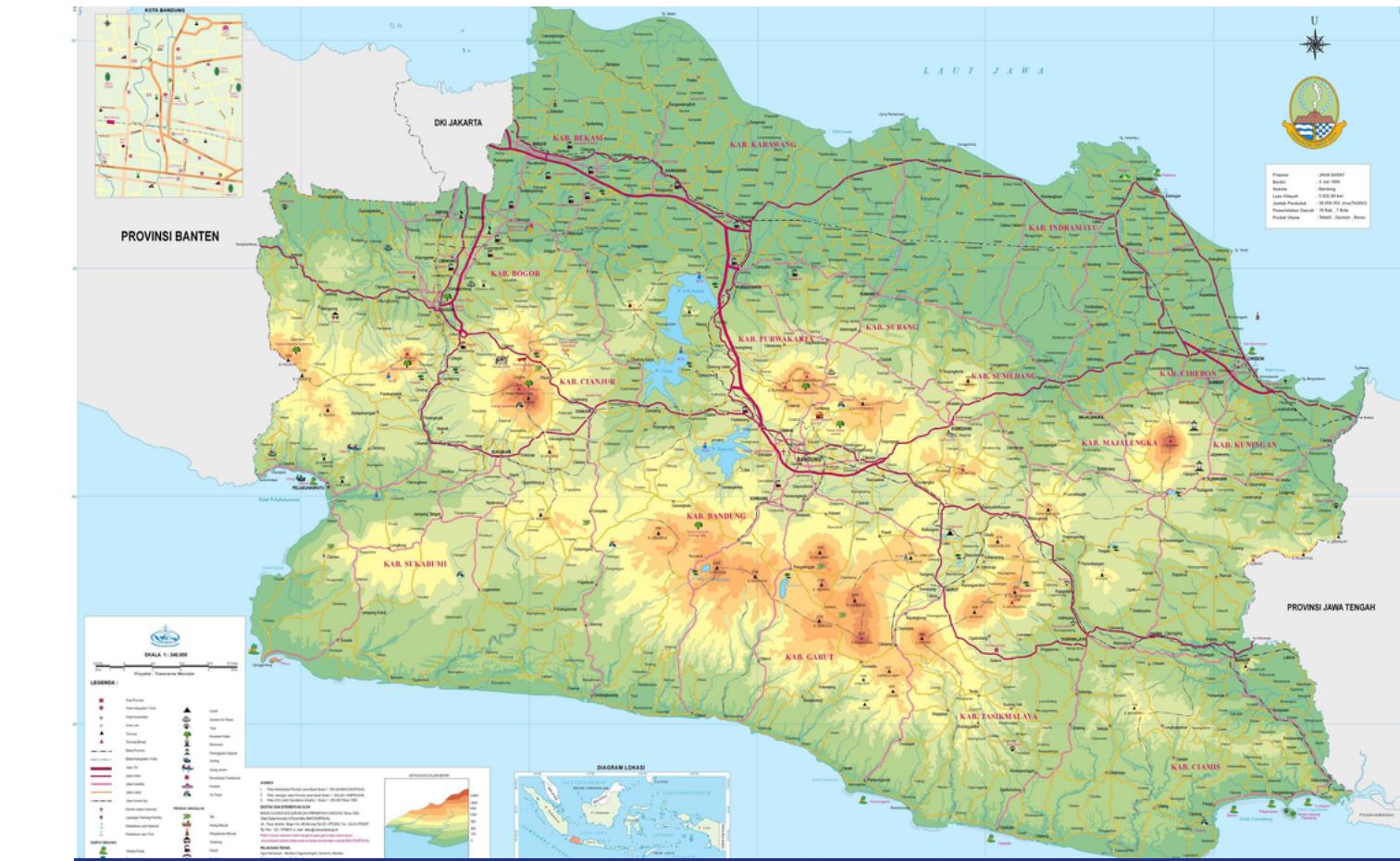


Kesimpulan

#1 —

Berdasarkan hasil dari validasi silang, **model regresi logistik biner Group LASSO** terbaik untuk capaian IPM Kota/Kabupaten di Jawa Barat dengan 6 kelompok variabel diperoleh dengan nilai λ optimal sebesar 0,0167. Adapun model regresi logistik biner Group LASSO dalam bentuk logit sebagai berikut.

$$\begin{aligned}g(x) = & -0,998 + 0,5089X_1 - 0,05269X_2 - 0,2923X_3 \\& + 0,1417X_4 - 1,59X_5 + 1,075X_6 + 0,794X_7 \\& + 0,8525X_8 + 0,0211X_{16} + 0,02575X_{17} \\& + 0,01115X_{18} - 0,01526X_{19} - 0,071X_{20} \\& - 0,061X_{21} - 0,4945X_{22} + 1,551X_{23} \\& - 4,411X_{24} + 4,217X_{25} - 0,0039X_{26} \\& + 1,987X_{27} + 0,814X_{28} + 0,53X_{29} \\& - 0,08243X_{30} - 0,14X_{31} + 0,432X_{32} \\& - 2,303X_{33} - 0,1747X_{34}\end{aligned}$$



#2 —

Faktor-faktor yang berpengaruh terhadap capaian IPM Kota/Kabupaten di Jawa Barat tahun 2020 yaitu kelompok pendidikan, kelompok ekonomi, kelompok lingkungan, dan kelompok kependudukan.

Rekomendasi

01

Mengkaji dan menggabungkan **Group LASSO** dengan metode lain yang dapat menaksir **standard error** seperti **Metode Bayesian** dan sebagainya.

02

Menggunakan metode yang lebih kompleks lagi yang mampu menyeleksi baik antarvariabel maupun antarkelompok variabel secara bersamaan seperti metode **Sparse Group LASSO**.

03

Menambah variabel-variabel lain yang diduga berpengaruh terhadap capaian IPM dan memperluas amatan wilayahnya.

04

Bagi pemerintah diharapkan lebih memperhatikan pada aspek-aspek yang mempengaruhi capaian IPM Kota/Kabupaten di Jawa Barat.



Kampus
Merdeka
INDONESIA JAYA


The Education University

TERIMA KASIH
Atas Perhatiannya!

- Abdurrahman & Rachman, F. F. (2021). Ketimpangan Indeks Pembangunan Manusia dan Komponennya antar Kabupaten/Kota di Provinsi Kalimantan Selatan 2010-2020. *BESTARI BPS Kalimantan Timur*, 1(01).
- Andriansyah, F., Wulandari, C., Hayatri, T. N., & Zahra, A. H. I. (2022). Strategy of Improvement of the Human Development Indeks in Indonesia Through Sustainable Settlement Management. *Dinamika Ekonomi*, 13(2), 259-270.
- Angelini, C. (2018). Regression analysis. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-809633-8.20360-9>
- Antara, I. G. M. Y., & Suryana, I. G. P. E. (2020). Pengaruh Tingkat Kepadatan Penduduk Terhadap Indeks Pembangunan Manusia di Provinsi Bali. *Media Komunikasi Geografi*, 21(1), 63-73.
- Apriati, Y., & Arofah, L. (2016). Profil Pengemis Di Kawasan Pemakaman Syekh Muhammad Arsyad Albanjari Di Desa Kalampaian Tengah Kecamatan Astambul Kabupaten Banjar Kalimantan Selatan.
- Arifin, M. Y. (2015). Pengaruh pengeluaran pemerintah sektor kesehatan, pengeluaran pemerintah sektor pendidikan dan pertumbuhan ekonomi terhadap indeks pembangunan manusia Jawa Timur 2006-2013.
- Bakin, S. (1999). Adaptive regression and model selection in data mining problems.
- Basri, S. M. (2017). Peranan Pasar Tradisional Dalam Meningkatkan Kesejahteraan Masyarakat (Studi Kasus Pada Pasar Pa'baeng-Baeng).
- Bazdaric, K., Sverko, D., Salaric, I., Martinovic, A., & Lucijanic, M. (2021). The ABC of linear regression analysis: What every author and editor should know. *European Science Editing*, 47, e63780.
- Bielza, C., Robles, V., & Larrañaga, P. (2011). Regularized logistic regression without a penalty term: An application to cancer classification with microarray data. *Expert Systems with Applications*, 38(5), 5110-5118.
- Breheny, P., & Huang, J. (2013). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and computing*, 25, 173-187.
- Brink-Jensen, K., & Ekstrøm, C. T. (2014). Inference for feature selection using the Lasso with high-dimensional data. *arXiv preprint arXiv:1403.4296*.
- Casella, G., Ghosh, M., Gill, J., & Kyung, M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* 5 (2) 369 – 411, June 2010. <https://doi.org/10.1214/10-BA607>
- Chatterjee, A., & Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494), 608-625.
- Chen, H., & Xiang, Y. (2017). The Study of Credit Scoring Model Based on Group Lasso. *Procedia Computer Science*, 122, 677-684. <https://doi.org/10.1016/j.procs.2017.11.423>
- Chen, S., Notodiputro, K. A., & Rahardiantoro, S. (2020). Penerapan analisis LASSO dan Group LASSO dalam mengidentifikasi faktor-faktor yang berhubungan dengan tuberkulosis di Jawa Barat. *Indonesian Journal of Statistics and Its Applications*, 4(1), 39-54.
- Daoud, J. I. (2018). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 949(1). <https://doi.org/10.1088/1742-6596/949/1/012009>
- Dewi, P. S. (2021). Dampak Pemekaran Kelurahan Dalam Meningkatkan Pelayanan Administrasi Pada Kantor Kelurahan Rapak Dalam Kota Samarinda. *PREDIKSI: Jurnal Administrasi dan Kebijakan*, 20(1), 67-70.
- Digdowiseiso, K., & Damayanti, N. (2021). Faktor-faktor Penentu Indeks Pembangunan Manusia di Provinsi Jawa Barat Tahun 2016-2020. *AKURASI: Jurnal Riset Akuntansi Dan Keuangan*, 3(3), 183-196. <https://doi.org/10.36407/akurasi.v3i3.428>
- Edgar, T. W., & Manz, D. O. (2017). Exploratory Study. *Research Methods for Cyber Security*, 95-130. <https://doi.org/10.1016/B978-0-12-805349-2.00004-2>
- El Sheikh, A. A., Barakat, S. L., & Mohamed, S. M. (2021). New aspects on the modified group lasso using the least angle regression and shrinkage algorithm. *Information Sciences Letters*, 10(3), 527-536. <https://doi.org/10.18576/isl/100317>

- Ferré, J. (2009). 3.02 – Regression Diagnostics. *Comprehensive Chemometrics*, Elsevier, 33–89
- Fitriyah, Z., Irsalina, S., & Widodo, E. (2021). ANALISIS FAKTOR YANG BERPENGARUH TERHADAP IPM MENGGUNAKAN REGRESI LINEAR BERGANDA. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 2(3), 282–291.
- Garson, G. D. (2014). *Logistic regression: Binary and multinomial*. Asheboro, NC.
- Hastie T, Thibshirani R, Friedman J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Ed ke-2. New York (US): Springer
- Hayati, N. (2019). Faktor-Faktor Mempengaruhi Ipm Di Kota Banjarmasin. *JIEP: Jurnal Ilmu Ekonomi Dan Pembangunan*, 1(1), 187–193. <https://doi.org/10.20527/jiep.v1i1.1127>
- Herrhyanto, N. (2003). *Statistika Matematis Lanjutan*. Bandung: Pustaka Setia.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Huang, J., & Zhang, T. (2010). The benefit of group sparsity. *The Annals of Statistics*, 38(4), 1978–2004.
- Iftikhar, S., & Aziz, B. (2017). Analysis Of Factors Affecting Solid Waste Management: A Study Of Cities Municipalities. *Gomal Journal of Social Sciences & Humanities*, 1(1), 48–60.
- Imanulloh, D. W., & Purwanti, L. (2017). Pengaruh belanja modal, Pendapatan Asli Daerah, dan pendapatan per kapita terhadap Indeks Pembangunan Manusia (studi pada kabupaten dan kota di Pulau Jawa tahun 2013–2015). *Jurnal Ilmiah Mahasiswa FEB*, 5(2), 1–15.
- Imron, C. (2018). Analisis Pengaruh Pertumbuhan Ekonomi, Pengeluaran Pemerintah Sektor Pendidikan, Kesehatan, Infrastruktur Terhadap Indeks Pembangunan Manusia Di Provinsi Jawa Timur 2011–2016 (Doctoral dissertation, Universitas Brawijaya).
- Iqbal, M. A. (2020). Application of regression techniques with their advantages and disadvantages. *Elektron Magazine*, 4(1), 11–17.
- Izzah, C. I., & Hendarti, I. M. (2021). Analisis Pengaruh Tenaga Kerja, Tingkat Upah, dan PDRB terhadap Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah. *Jurnal Kajian Pendidikan Ekonomi Dan Ilmu Ekonomi*, 2, 99–106.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
- Kariuki, C. W. (2015). Factors influencing solid waste management in urban centres: A case of Thika sub county, Kiambu county, Kenya (Doctoral dissertation, University of Nairobi).
- Khairunnisa, S. F., Suharni, S., & Nohe, D. A. (2022). ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI INDEKS PEMBANGUNAN MANUSIA DI JAWA BARAT MENGGUNAKAN REGRESI LOGISTIK BINER. In *Prosiding Seminar Nasional Matematika dan Statistika* (Vol. 2).
- Khoiriah, A. A., Bakri, S., & Santoso, T. (2017). Pengaruh Perubahan Lahan, Tingkat Kemiskinan Dan Pendapatan Beberapa Sektor Perekonomian Terhadap Indeks Pembangunan Manusia: Studi Di Provinsi Lampung. *Jurnal Sylva Lestari*, 5(1), 117–127.
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean journal of anesthesiology*, 72(6), 558–569.
- Koh, K., Kim, S. J., & Boyd, S. (2007). An interior-point method for large-scale ℓ_1 -regularized logistic regression. *Journal of Machine learning research*, 8(Jul), 1519–1555.
- Kulla, T., Rumapea, P., & Tampongango, D. (2018). Kualitas Sumber daya manusia dalam meningkatkan pembangunan desa tinggilbet distrik beoga kabupaten Puncak provinsi Papua. *Jurnal Administrasi Publik*, 4(58).
- Latuconsina, Z. M. Y. (2017). Analisis faktor-faktor yang mempengaruhi indeks pembangunan manusia Kabupaten Malang berbasis pendekatan perwilayah dan regresi panel. *Journal of Regional and Rural Development Planning (Jurnal Perencanaan Pembangunan Wilayah Dan Perdesaan)*, 1(2), 202–216.

- Li, X., & Xu, L. (2021). Human development associated with environmental quality in China. *Plos one*, 16(2), e0246677.
- Liu, H., & Zhang, J. (2009). Estimation consistency of the group lasso and its applications. *Journal of Machine Learning Research*, 5, 376–383.
- Marseno, B., & Mulyani, E. (2020). Pengaruh Pertumbuhan Ekonomi, Pendapatan Asli Daerah (PAD), Jumlah Penduduk Dan Luas Wilayah Terhadap Belanja Modal Pemerintah Daerah. *Jurnal Eksplorasi Akuntansi*, 2(4), 3452–3467.
- Mirza, D. S. (2011). Pengaruh kemiskinan, pertumbuhan ekonomi, dan belanja modal terhadap ipm jawa tengah. *JEJAK: Jurnal Ekonomi Dan Kebijakan*, 4(2).
- Narisetty, N. N. (2020). Bayesian model selection for high-dimensional data. In *Handbook of Statistics* (Vol. 43, pp. 207–248). Elsevier.
- Rahmi, D. M. (2019). Investasi Modal Manusia Dalam Pembangunan Ekonomi Perspektif Islam. *Jurnal Manajemen Dan Bisnis Indonesia*, 5(1), 55–64.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 5, 532–538.
- Sapaat, T. M., Lapian, A. L. C. P., & Tumangkeng, S. Y. (2020). Analisis faktor-faktor yang mempengaruhi indeks pembangunan manusia di Provinsi Sulawesi Utara tahun (2005-2019). *Jurnal Berkala Ilmiah Efisiensi*, 20(03).
- Sari, M., & Purhadi, P. (2021). PEMODELAN INDEKS PEMBANGUNAN MANUSIA PROVINSI JAWA BARAT, JAWA TIMUR DAN JAWA TENGAH TAHUN 2019 DENGAN MENGGUNAKAN METODE REGRESI LOGISTIK ORDINAL. *Jurnal Gaussian*, 10(1), 149–158.
- Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, 107(44), 776.
- Senaviratna, N. A. M. R., & Cooray, T. M. J. A. (2019). Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics*, 5(2), 1–9.
- Setiawan, S., & Sutikno, S. (2010). Bayes Wavelet Regression Approach to Solve Problems in Multivariable Calibration Modeling. *IPTEK The Journal for Technology and Science*, 21(2).
- Simanjuntak, F. N. (2017). Pendidikan untuk pembangunan berkelanjutan. *Jurnal Dinamika Pendidikan*, 10(2), 169–195.
- Sudjana. (2005). *Metoda Statistika Edisi 7*. Bandung: Tarsito
- Sunandi, E., Notodopuro, K. A., & Sartono, B. (2021). A study on group lasso for grouped variable selection in regression model. *IOP Conference Series: Materials Science and Engineering*, 1115(1), 012089. <https://doi.org/10.1088/1757-899x/1115/1/012089>
- Suryani, A., & Sartika, E. (2021). Determinasi Faktor-Faktor Indeks Pembangunan Manusia di Jawa Barat Menggunakan Regresi Data Panel. *ETHOS: Jurnal Penelitian dan Pengabdian kepada Masyarakat*, 9(1), 151–160.
- Susiani, I. R., & Abadiah, N. D. (2021). Kualitas Guru dalam Meningkatkan Mutu Pendidikan di Indonesia. *MODELING: Jurnal Program Studi PGMI*, 8(2), 292–298.
- Taufik (2022) Peningkatan Indeks Pembangunan Manusia Pada Aspek Pendidikan di Kabupaten Takalar
- Tewu, M. E. (2015). Peranan sumber daya manusia dalam meningkatkan aktivitas kelompok tani di Desa Tember. *Acta Diurna Komunikasi*, 4(3).
- Tibshirani, R. J., & Tibshirani, R. (2009). A bias correction for the minimum error rate in cross-validation.
- Trisno, T. U., & Oktarina, Y. (2022). Pengaruh kemiskinan terhadap indeks pembangunan manusia (ipm) di provinsi sumatera selatan tahun 2016–2020. *Fair Value: Jurnal Ilmiah Akuntansi dan Keuangan*, 4(8), 3560–3566.
- Tristanto, A., & Diartho, H. C. (2018). Strategi Sektor Kesehatan Dalam Meningkatkan Indeks Pembangunan Manusia Di Kabupaten Situbondo. *Wiga: Jurnal Penelitian Ilmu Ekonomi*, 8(2), 86–93.
- Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)*, 6(2).
- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.
- Yunus, M., Saefudin, A., & Soleh, A. M. (2017). Characteristics of group LASSO in handling high correlated data. *Applied Mathematical Sciences*, 11(20), 953–961. <https://doi.org/10.12988/ams.2017.7276>
- Zakaria, R. (2018). Pengaruh Tingkat Jumlah Penduduk, Pengangguran, Kemiskinan, Pertumbuhan Ekonomi, dan Belanja Modal Terhadap Indeks Pembangunan Manusia di Provinsi Jawa Tengah Tahun 2010–2016.
- Zhao, N., Xu, Q., Tang, M. L., Jiang, B., Chen, Z., & Wang, H. (2020). High-dimensional variable screening under multicollinearity. *Stat*, 9(1), e272.
- Zhu, L., Li, L., & Wang, M. H. (2020). Comparison of Regression Models on House Value Prediction.
- Zuhairoh, Z. A., & Melaniani, S. (2018). Pengaruh Angka Kematian Bayi, Angka Partisipasi Murni, Rasio Ketergantungan terhadap Indeks Pembangunan Manusia Provinsi Jawa Timur. *Jurnal Biometrika dan Kependudukan*, 7(1), 87–95.