

SKRIPSI

**OPTIMASI REGRESI LOGISTIK LASSO DENGAN
KRITERIA *BAYESIAN INFORMATION CRITERION* (BIC)
(STUDI KASUS : KLASIFIKASI KEMISKINAN DI
INDONESIA TAHUN 2024)**



FIRA AZZAH WIJAYAKANTHI
22106010018
STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA

2026

**OPTIMASI REGRESI LOGISTIK LASSO DENGAN
KRITERIA *BAYESIAN INFORMATION CRITERION* (BIC)
(STUDI KASUS : KLASIFIKASI KEMISKINAN DI
INDONESIA TAHUN 2024)**

Skripsi

Untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-1
Program Studi Matematika



diajukan oleh

FIRA AZZAH WIJAYAKANTHI

22106010018

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Kepada

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2026



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Fira Azzah Wijayakanthi

NIM : 22106010018

Judul Skripsi : Optimasi Regresi Logistik LASSO dengan Kriteria Bayesian Information Criterion (BIC) (STUDI KASUS : KLASIFIKASI KEMISKINAN DI INDONESIA TAHUN 2024)

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 4 Mei 2026

Pembimbing

Sri Utami Zuliana, S.Si., M.Si., Ph.D

NIP. 19741003 200003 2 002



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1166/Un.02/DST/PP.00.9/06/2026

Tugas Akhir dengan judul : OPTIMASI REGRESI LOGISTIK LASSO DENGAN KRITERIA BAYESIAN INFORMATION CRITERION (BIC)(STUDI KASUS : KLASIFIKASI KEMISKINAN DI INDONESIA TAHUN 2024)

yang dipersiapkan dan disusun oleh:

Nama : FIRA AZZAH WIJAYAKANTHI
Nomor Induk Mahasiswa : 22106010018
Telah diujikan pada : Senin, 18 Mei 2026
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.
SIGNED

Valid ID: 6a20fab3b66d0



Penguji I

Muhamad Rashif Hilmi, S.Si., M.Sc.
SIGNED

Valid ID: 6a1fe2af91c64



Penguji II

Dr. Sugiyanto, S.Si., ST., M.Si.
SIGNED

Valid ID: 6a15092299cd5



Yogyakarta, 18 Mei 2026

UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a20fe3b81830

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Fira Azzah Wijayakanthi
NIM : 22106010018
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 8 Mei 2026



Fira Azzah Wijayakanthi

HALAMAN PERSEMBAHAN

KARYA SEDERHANA INI PENULIS PERSEMBAHKAN UNTUK:
KEDUA ORANG TUA PENULIS, ADIK PENULIS DAN KELUARGA BESAR
PENULIS YANG TELAH MEMBERIKAN DO'A, KASIH
DAN CINTA NYA YANG TAK TERBATAS.
KELUARGA DEKAT, SAHABAT DAN SEMUA ORANG-ORANG YANG
PERNAH SAYA KENAL.
ALMAMATER KAMPUS TERCINTA YANG BANYAK MEMBERIKAN
WARNA DALAM HIDUP SELAMA 4 TAHUN UNIVERSITAS ISLAM NEGERI
SUNAN KALIJAGA YOGYAKARTA.

"MATEMATIKA ANGGKATAN 2022 TERCINTA"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

”Bersungguh-sungguhlah, karena sesungguhnya Allah menyukai orang yang bekerja keras.”

(HR. Thabrani)

”Orang boleh pandai setinggi langit, tapi selama ia tidak menulis, ia akan hilang di dalam masyarakat dan dari sejarah. Menulis adalah bekerja untuk keabadian.”

(Pramoedya Ananta Toer)

”Hidup tidak mudah bagi siapa pun. Tapi apa pedulinya? Kita harus memiliki kegigihan dan yang terpenting, kepercayaan pada diri sendiri.”

(Marie Curie)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan karunianya yang tak ternilai harganya berupa keimanan, kesabaran, kekuatan dan kelancaran. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Optimasi Regresi Logistik LASSO dengan Kriteria *Bayesian Information Criterion* (BIC) (Studi kasus : Klasifikasi Kemiskinan di Indonesia Tahun 2024)".

Mengenai skripsi ini, penulis ingin mengucapkan terima kasih kepada pembimbing skripsi penulis, Ibu Sri Utami Zuliana, S.Si.,M.Sc., Ph.D. Pertemuan rutin tiap dua pekan sekali memungkinkan penulis dalam mengenali progres pengerjaan skripsi ini. Bimbingan dan komentarnya mencegah penulis memilih cara yang lebih mudah. Selama menempuh perkuliahan ini penulis sangat berterima kasih kepada Tanpa mengurangi rasa hormat, penulis juga ingin menyampaikan ucapan terima kasih dan apresiasi yang sebesar-besarnya kepada:

1. Bapak Prof. Noorhaidi Hasan, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta dan selaku dosen pembimbing akademik yang telah banyak memberikan pengarahan kepada penulis dari awal perkuliahan hingga akhir.
3. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc. selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ka-

lijaga Yogyakarta.

4. Seluruh dosen program studi Matematika dan staf fakultas Sains dan Teknologi yang senantiasa memberikan ilmu dan layanan terbaik kepada penulis dari awal hingga akhir perkuliahan.
5. Kedua Orang tua penulis, Adik penulis dan keluarga besar penulis yang telah memberikan banyak nasehat, kasih sayang, doa dan dukungannya sehingga penulis sampai ke titik ini.
6. Keluarga Matematika angkatan 2022 yang telah kebersamai selama 4 tahun dibangku perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.
7. Teman-teman seperbimbingan Shun, Mutiara dan Lucky yang telah bersama-sama menemani selama bimbingan.
8. Teman-teman KKN kelompok 89 sebatang terima kasih cerita 45 hari nya selama KKN.
9. Teman dekat penulis, terima kasih sudah berada terimakasih sudah bertahan disisi saya, sudah mau menjadi teman dekat saya, sudah bersama-sama melewati masa sulit
10. Semua pihak yang penulis tidak bisa sebutkan satu-persatu, semoga dimudahkan segala urusannya.

Penulis sangat menyadari tidak ada yang sempurna di dunia ini. Oleh karena itu segala bentuk saran, kritik atas skripsi ini penulis menerimanya dengan senang hati. Penulis berharap skripsi ini dapat memberikan manfaat khususnya bagi penulis dan semua kalangan yang membutuhkan untuk dijadikan referensi dalam menulis.

Akhir kata semoga Allah SWT melimpahkan segala rahmat dan hidayah-Nya kepada kita semua, *Aamiin*.

Yogyakarta, 11 Mei 2026

Fira Azzah Wijayakanthi



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
PRAKATA	vii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMBANG	xvi
ABSTRAKxviii
ABSTRACT	xix
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
1.6. Tinjauan Pustaka	6
1.7. Sistematika Penulisan	9
II DASAR TEORI	10
2.1. Model Linear	10

2.1.1. Representasi Matriks	11
2.2. Generalized Linear Models (GLM)	12
2.3. Regresi Logistik	15
2.3.1. Model Regresi Logistik	16
2.3.2. <i>Maximum Likelihood Estimation</i> (MLE)	18
2.3.3. Iteratively Reweighted Least Squares (IRLS)	21
2.4. Multikolinearitas	25
2.5. <i>Least Absolute Shrinkage and Selection Operator</i> (LASSO)	26
2.5.1. Formulasi LASSO	26
2.5.2. Regresi Logistik LASSO	28
2.5.3. Algoritma IRLS dan <i>Coordinate Descent</i> untuk Regresi Logistik LASSO	29
2.6. <i>Kriteria Informasi Bayesian</i> (BIC)	30
2.6.1. Interpretasi BIC	31
2.6.2. BIC dalam LASSO untuk Regresi Logistik	32
2.7. Klasifikasi Kemiskinan di Indonesia	34
2.7.1. Garis Kemiskinan(GK)	34
2.7.2. Perhitungan Penduduk Miskin	34
2.7.3. Perhitungan Persentase Kemiskinan (P0)	35
III METODE PENELITIAN	38
3.1. Pendekatan Penelitian	38
3.2. Sumber Data	39
3.3. Variabel Data	39
3.3.1. Variabel Dependen	40
3.3.2. Variabel Independen	40
3.4. Metode Pengolahan Data	41

3.5. Diagram Analisis Data (<i>Flowchart</i>)	42
IV HASIL DAN PEMBAHASAN	43
4.1. Deskripsi Data Penelitian	43
4.1.1. Variabel Dependen (Persentase Penduduk Miskin (PO))	43
4.2. Statistik Deskriptif	45
4.3. Model Regresi Logistik Awal (MLE-IRLS)	47
4.3.1. Uji Signifikansi Parsial	48
4.4. Uji Multikolinearitas	53
4.4.1. Uji Multikolinearitas Model Regresi Logistik Sebelum Backward Stepwise	53
4.4.2. Uji Multikolinearitas Model Regresi Logistik Sesudah Backward Stepwise	55
4.5. Standarisasi Variabel independen	56
4.6. Hasil Regresi Logistik LASSO	57
4.6.1. Pemilihan λ berdasarkan BIC	59
4.6.2. Lintasan Koefisien Logistik LASSO	60
4.6.3. Model LASSO Terbaik	61
4.7. Perbandingan Model	64
4.7.1. Perbandingan BIC Logistik LASSO dan Logistik MLE (sebelum backward stepwise)	64
4.7.2. Perbandingan BIC Logistik LASSO dan Logistik (MLE) (backward stepwise)	65
4.7.3. Perbandingan Model Regresi Logistik MLE (backward stepwise) dan Regresi Logistik LASSO	66
4.8. Contoh Simulasi	68
4.8.1. Simulasi 2024 Provinsi Sumatera Utara	69
4.8.2. Simulasi 2025 Provinsi Sumatera Utara	71

4.8.3. Kesimpulan Simulasi	74
V PENUTUP	75
5.1. Kesimpulan	75
5.2. Saran	77
DAFTAR PUSTAKA	78
LAMPIRAN	82
A DATA PENELITIAN	82
B SOURCE CODE PROGRAM R	86
C TABEL TINJAUAN PUSTAKA	98
D CURRICULUM VITAE	100

DAFTAR TABEL

2.1	Kriteria Klasifikasi Kemiskinan	35
2.2	Jumlah dan Persentase Penduduk Miskin di Indonesia	35
2.3	Perhitungan Persentase Kemiskinan (P_0)	35
3.1	Kategorisasi Tingkat Kemiskinan	40
3.2	Daftar Variabel Independen	40
4.1	Tabel Statistik Deskriptif	45
4.2	Tabel Hasil Regresi Logistik	48
4.3	Tabel Hasil Regresi Logistik Mengeliminasi variabel TPT	49
4.4	Tabel Hasil Regresi Logistik Mengeliminasi variabel PKD	50
4.5	Tabel Hasil Regresi Logistik Mengeliminasi variabel ASL	50
4.6	Tabel Hasil Regresi Logistik Mengeliminasi variabel IPM	50
4.7	Tabel Hasil Regresi Logistik Mengeliminasi variabel AAML	51
4.8	Tabel Hasil Regresi Logistik Mengeliminasi variabel RLS	51
4.9	Hasil Estimasi Koefisien Regresi Logistik backward stepwise	52
4.10	Tabel Hasil Multikolinearitas	54
4.11	Tabel Hasil Multikolinearitas	56
4.12	Tabel Koefisien Penalti Lasso	62
4.13	Perbandingan Kriteria Model	65
4.14	Perbandingan Kriteria Model	65

DAFTAR GAMBAR

4.1	Pemilihan λ Optimasi Bersasarkan BIC	59
4.2	Lintasan Koefisien Regresi Logistik LASSO	60



DAFTAR LAMBANG

β_0	= Parameter intersep pada model regresi
β_j	= Koefisien regresi untuk variabel Independen ke- j
β	= Vektor parameter regresi berukuran $(p + 1) \times 1$
$\hat{\beta}$	= Penduga parameter regresi
Y	= Variabel dependen
Y_i	= Variabel dependen pada pengamatan ke- i
y_i	= Nilai realisasi variabel dependen pada pengamatan ke- i
X_j	= Variabel independen ke- j
x_{ij}	= Nilai variabel independen ke- j pada pengamatan ke- i
\mathbf{x}_i	= Vektor variabel independen pada pengamatan ke- i
\mathbf{X}	= Matriks variabel independen berukuran $n \times (p + 1)$
p_i	= Peluang kejadian $Y_i = 1$ pada pengamatan ke- i
\hat{p}_i	= Penduga peluang kejadian $Y_i = 1$
μ_i	= Nilai harapan dari variabel dependen Y_i
η_i	= Independen linier pada pengamatan ke- i
$g(\cdot)$	= Fungsi penghubung (<i>link function</i>)
$g^{-1}(\cdot)$	= Fungsi invers dari fungsi penghubung
ε	= Galat (<i>error</i>) pada model regresi linier

α	= tingkat signifikansi
θ	= parameter proporsi
L	= <i>likelihood</i>
e	= eksponensial
Γ	= fungsi <i>gamma</i>
$p(\theta y)$	= distribusi <i>posterior</i>
$p(\theta)$	= distribusi <i>prior</i>
$f(y \theta)$	= nilai <i>likelihood</i> dari sampel
$p(y)$	= <i>normalized constant</i> yang dapat diabaikan
y_t	= data pengamatan pada periode waktu ke- t
x_t	= vektor berdimensi $1 \times K$ dengan elemen variabel bebas
z_t	= vektor berdimensi $1 \times p$ yang mengandung input komponen
β	= koefisien regresi dari variabel bebas
θ_t	= vektor berdimensi $p \times 1$ yang berisi p persamaan <i>state</i> parameter
h	= parameter presisi yang didefinisikan sebagai $h = 1/\sigma^2$
T_t	= matriks transisi dengan dimensi $p \times p$
$z_t = z, T_t = T.v_t$	= vektor berdimensi $p \times 1$ independen

ABSTRAK

Optimasi Regresi Logistik LASSO dengan Kriteria *Bayesian Information Criterion* (BIC) (STUDI KASUS : KLASIFIKASI KEMISKINAN DI INDONESIA TAHUN 2024)

Oleh

Fira Azzah Wijayakanthi

22106010018

Regresi Logistik LASSO adalah metode regresi yang efektif untuk mengatasi multikolinearitas dan melakukan seleksi variabel secara simultan melalui penerapan penalti L1 pada fungsi log-likelihood. Metode ini digunakan untuk menganalisis klasifikasi kemiskinan, yang pembentukannya melibatkan faktor sosial, ekonomi, kesehatan, pendidikan, dan pekerjaan secara bersamaan. Penelitian ini bertujuan menerapkan Regresi Logistik LASSO dengan optimisasi Bayesian Information Criterion (BIC) untuk mengidentifikasi faktor penentu klasifikasi kemiskinan provinsi di Indonesia tahun 2024. Data yang digunakan bersumber dari Badan Pusat Statistik (BPS), mencakup 38 provinsi dengan sembilan variabel independen dan satu variabel dependen berupa klasifikasi tingkat kemiskinan berbasis rata-rata nasional. Tahapan analisis meliputi uji multikolinearitas dengan VIF, standarisasi data, pemodelan LASSO, lalu pemilihan nilai penalti optimal berdasarkan BIC. Dari hasil analisis, Regresi logistik tidak ada variabel yang signifikan dan ditemukan multikolinearitas pada data awal. Pada LASSO didapatkan penalti optimal $\lambda = 0,017373$ dengan nilai BIC minimum 41,695. Model akhir memuat tiga variabel, ada Umur Harapan Hidup (UHH), Tingkat Partisipasi Angkatan Kerja (TPAK), dan Produk Domestik Regional Bruto (PDRB). Regresi Logistik LASSO berbasis BIC terbukti mampu mengatasi multikolinearitas sekaligus menghasilkan model yang lebih ringkas dan stabil untuk mengidentifikasi faktor-faktor yang memengaruhi klasifikasi kemiskinan di Indonesia pada tahun 2024.

Kata kunci: Kemiskinan, Regresi Logistik, LASSO, BIC, Klasifikasi.

ABSTRACT

LASSO Logistic Regression Optimization Using the Bayesian Information Criterion (BIC) (CASE STUDY: POVERTY CLASSIFICATION IN INDONESIA IN 2024)

By

Fira Azzah Wijayakanthi

22106010018

LASSO Logistic Regression is an effective regression method for addressing multicollinearity and performing simultaneous variable selection through the application of an L1 penalty to the log-likelihood function. This method is used to analyze poverty classification, the formation of which involves social, economic, health, education, and employment factors simultaneously. This study aims to apply LASSO Logistic Regression with Bayesian Information Criterion (BIC) optimization to identify the determinants of provincial poverty classification in Indonesia in 2024. The data used is sourced from the Central Statistics Agency (BPS), covering 38 provinces with nine independent variables and one dependent variable in the form of poverty level classification based on the national average. The analysis stages include a multicollinearity test using VIF, data standardization, LASSO modeling, and selection of the optimal penalty value based on BIC. From the analysis results, logistic regression showed no significant variables, and multicollinearity was found in the initial data. In LASSO, the optimal penalty was $\lambda = 0,017373$ with a minimum BIC value of 41,695. The final model includes three variables, life expectancy (LE), labor force participation rate (LFPR), and regional gross domestic product (RGDP). BIC-based LASSO logistic regression has proven capable of addressing multicollinearity while producing a more concise and stable model for identifying the factors influencing poverty classification in Indonesia in 2024.

Keywords: Poverty, Logistic Regression, LASSO, BIC, Classification.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Regresi logistik adalah metode statistik yang sering digunakan dalam analisis data kategorikal ketika variabel respon bersifat biner. Metode ini menggunakan fungsi logistik dari variabel prediktor linier untuk memodelkan probabilitas suatu kategori tertentu. Karena itu, nilai prediktor berada dalam rentang probabilitas (0, 1), sehingga secara statistik lebih akurat daripada regresi linier. Prinsip dasar model ini menyatakan bahwa tidak ada *multikolinearitas* antara variabel prediktor, kemandirian antar pengamatan, serta ukuran sampel yang konsisten dan bebas bias. Fleksibilitas regresi logistik memungkinkan variabel prediktor bersifat kontinu, ordinal, atau nominal tanpa melanggar asumsi normalitas variabel respon (Mursyidah et al. (2025)). Dengan demikian, model ini menunjukkan adaptabilitas yang tinggi terhadap data sosioekonomi yang umumnya tidak mengikuti distribusi normal.

Least Absolute Shrinkage and Selection Operator (LASSO) adalah Teknik regresi seperti regresi logistik digunakan untuk mengatasi multikolinearitas, overfitting, dan ketidakstabilan dalam model regresi (Tibshirani (1996)). Metode ini menerapkan penalti L1 pada fungsi log-likelihood, sehingga memungkinkan pemilihan variabel dan identifikasi koefisien yang tidak signifikan secara bersamaan. Dalam konteks regresi logistik, LASSO menghasilkan model yang lebih akurat dengan hanya berfokus pada variabel-variabel utama, sehingga sangat cocok untuk

data berdimensi tinggi. Parameter kunci dalam LASSO adalah λ , yang menentukan kekuatan penalti; seiring meningkatnya nilainya, jumlah variabel dalam model akan berkurang. Analisis LASSO digunakan dalam penelitian ini karena kemampuannya menghasilkan solusi yang jarang (koefisien bernilai nol secara langsung), yang memungkinkan pemilihan variabel dan estimasi parameter dalam satu langkah optimisasi. Hal ini sangat relevan ketika data mengandung beberapa prediktor yang mungkin berlebihan, seperti pada data sosioekonomi multidimensi (Bühlmann & Van De Geer (2011)). Estimasi parameter LASSO dalam regresi logistik dapat dilakukan dengan menggunakan algoritma *Cyclic Coordinate Descent* (CCD) atau algoritma *Iteratively Reweighted Least Squares* (IRLS), di mana setiap koefisien dihitung secara bergantian dengan mengurangi koefisien lainnya hingga tercapai konvergensi (Kamila et al. (2021)). Parameter kunci dalam LASSO adalah λ , yang menentukan kekuatan penalti. Seiring dengan meningkatnya nilai λ , jumlah variabel dalam model akan berkurang, sehingga pemilihan nilai λ yang sesuai menjadi langkah penting dalam menentukan kualitas model akhir (Fitria & Rozci (2022)).

Bayesian Information Criterion (BIC) adalah metode seleksi model yang dikembangkan oleh Schwarz (1978) dengan menggunakan pendekatan Bayesian untuk menentukan kompleksitas model. Secara matematis, BIC menghubungkan nilai log-likelihood dengan penalti yang didasarkan pada ukuran sampel, sehingga memungkinkan pemilihan model yang lebih akurat dibandingkan dengan AIC. Dalam implementasi LASSO, BIC digunakan untuk menentukan nilai λ yang optimal dengan menyesuaikan ketepatan dan kesederhanaan model tanpa memerlukan validasi komputasi yang lebih rumit. Selain itu, BIC memiliki landasan teoritis yang kuat karena konsistensinya dalam memilih model yang tepat, terutama dalam hal ukuran sampel. Pemilihan BIC sebagai kriteria optimisasi didasarkan pada sifatnya

yang konsisten, yaitu BIC cenderung memilih model yang secara asimtotik baik ketika ukuran sampel meningkat, sedangkan AIC cenderung memilih model yang lebih kompleks. Menggabungkan BIC dan LASSO menghasilkan analisis yang lebih efektif dan komprehensif terhadap data berdimensi tinggi, seperti data sosial-ekonomi (Wang et al. (2009), Chen & Chen (2008)).

Kemiskinan merupakan fenomena yang memiliki berbagai aspek dan telah lama menjadi fokus penelitian ekonomi, analisis statistik, serta analisis kebijakan publik. Fenomena ini berkaitan erat dengan kesehatan manusia, pendidikan, sanitasi, dan akses terhadap layanan dasar. Dari perspektif statistik modern, penting untuk mengidentifikasi faktor-faktor yang menunjukkan tingkat kemiskinan relatif suatu provinsi jika dibandingkan dengan seluruh negara. Menurut data dari Badan Pusat Statistik, populasi Indonesia pada tahun 2024 adalah sekitar 25 juta orang, dengan distribusi yang tidak merata di antara provinsi-provinsi. Situasi ini menyoroti perlunya analisis yang dapat menangani data sosioekonomi yang kompleks, heterogen, dan berskala besar (Alkire & Foster (2011), Syifana et al. (2025)).

Penelitian ini menggunakan kombinasi regresi logistik, LASSO, dan BIC dengan tiga komponen utama. Pertama, regresi logistik berdasarkan variabel respons biner (di atas atau di bawah rata-rata nasional). Kedua, LASSO diperlukan karena data indikator sosioekonomi provinsi mengandung beberapa variabel prediktor yang sering kali berkorelasi (multikolinear). Akibatnya, metode konvensional tanpa regularisasi rentan menghasilkan estimasi yang tidak stabil dan sulit dipahami. Ketiga, BIC digunakan sebagai kriteria ideal untuk memilih λ karena konsisten dan tidak memerlukan proses validasi silang yang intensif, sehingga lebih efisien untuk ukuran sampel provinsi yang kecil ($n = 38$).

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana mekanisme regresi logistik LASSO dalam melakukan seleksi variabel dan estimasi model?
2. Bagaimana langkah-langkah optimisasi pada model Regresi Logistik LASSO dengan menggunakan *Bayesian Information Criterion* (BIC)?
3. Bagaimana penerapan optimisasi Regresi Logistik LASSO menggunakan *Bayesian Information Criterion* (BIC) terhadap data faktor Klasifikasi Kemiskinan di Indonesia Tahun 2024?

1.3. Batasan Masalah

Untuk menghindari perluasan pembahasan dan menjaga fokus penelitian, maka penelitian ini dibatasi pada hal-hal berikut:

1. Data yang digunakan adalah Data dan Informasi Kemiskinan Provinsi Tahun 2024 yang tersedia dalam dataset publik dari BPS. Variabel respon yang dianalisis bersifat biner, misalnya diklasifikasikan sebagai Klasifikasi miskin diatas rata-rata atau Klasifikasi miskin dibawah rata-rata.
2. Model yang digunakan terbatas pada regresi logistik LASSO. Kriteria pemilihan model terbatas pada *Bayesian Information Criterion* (BIC).
3. *Software* yang digunakan adalah R versi 4.4.5

1.4. Tujuan Penelitian

1. Menjelaskan mekanisme estimasi parameter pada regresi logistik LASSO menggunakan algoritma *Iteratively Reweighted Least Squares* (IRLS).
2. Mengetahui langkah-langkah optimisasi pada model Regresi Logistik LASSO dengan menggunakan *Bayesian Information Criterion* (BIC)
3. Menerapkan optimisasi Regresi Logistik LASSO menggunakan *Bayesian Information Criterion* (BIC) terhadap data faktor Klasifikasi Kemiskinan di Indonesia Tahun 2024

1.5. Manfaat Penelitian

Diharapkan hasil penelitian ini dapat memberikan manfaat. Secara teoritis, penelitian ini berkontribusi pada pengembangan dan penerapan metode statistik, khususnya dalam pemodelan regresi logistik menggunakan metode LASSO dan teknik seleksi model dengan menggunakan *Bayesian Information Criterion* (BIC). Penelitian ini menggambarkan bagaimana kombinasi kedua pendekatan tersebut dapat menghasilkan model klasifikasi yang lebih efisien dan lebih terkendali dalam hal kompleksitas. Secara langsung, hasilnya dapat digunakan sebagai alat untuk menemukan informasi mengenai pola-pola pembagian terstruktur terkait kemiskinan di Indonesia berdasarkan variabel prediktor yang relevan. Berdasarkan analisis multidimensi terhadap kondisi sosial-ekonomi, hal ini dapat menjadi cara untuk mendukung keputusan pengentasan kemiskinan yang lebih terfokus dan tepat sasaran. memberikan pemahaman praktis tentang bagaimana regresi logistik LASSO yang dioptimalkan dengan menggunakan BIC dapat digunakan untuk menganalisis data kategorikal kompleks dengan banyak variabel prediktor. Kombinasi ini dapat menjadi solusi yang tepat untuk mengatasi masalah multikolinearitas dan *overfitting*,

serta memudahkan interpretasi hasil. Memberikan panduan dan landasan metodologis bagi para peneliti di masa mendatang yang ingin menganalisis data kategorikal dengan menggunakan regularisasi dan model statistik yang efisien.

1.6. Tinjauan Pustaka

Pada penelitian yang dilakukan oleh (Kamila et al. (2021)) dengan judul *Metode Least Absolute Shrinkage and Selection Operator (LASSO) untuk Pendugaan Parameter Regresi Logistik Biner (Studi Kasus: Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021)* membahas penerapan metode *Least Absolute Shrinkage and Selection Operator (LASSO)* dalam estimasi parameter regresi logistik untuk menganalisis faktor-faktor yang memengaruhi kemiskinan di Indonesia pada tahun 2021. Penelitian ini menggunakan metode LASSO, yang dapat melakukan seleksi variabel dan regularisasi sekaligus, untuk mengatasi masalah *multikolinearitas* pada variabel prediktor. Artikel ini juga membahas proses pemecahan masalah standar menggunakan metode bootstrap, uji Wald untuk signifikansi parameter, dan algoritma *Cyclic Coordinate Descent (CCD)* untuk estimasi parameter. Matriks kebingungan dan Tingkat Kesalahan Tampak (APER) digunakan untuk mengevaluasi akurasi model klasifikasi akhir. Temuan penelitian menunjukkan bahwa variabel pendapatan per kapita dan Produk Domestik Regional Bruto (PDRB) merupakan faktor yang memengaruhi tingkat kemiskinan. Model LASSO dapat digunakan untuk mengukur akurasi prediksi dalam menentukan kriteria kemiskinan di berbagai wilayah di Indonesia.

Pada penelitian yang dilakukan oleh (Fitria & Rozci (2022)) dengan judul *Penerapan Metode Regresi Least Absolute Shrinkage and Selection Operator (LASSO) dan Regresi Linier untuk Memprediksi Tingkat Kemiskinan di Indonesia* yang bertujuan, membandingkan kinerja metode regresi *Least Absolute Shrinkage*

and Selection Operator (LASSO) dengan metode regresi linier dalam memprediksi tingkat risiko di Indonesia menggunakan data sekunder dari Badan Pusat Statistik (BPS) untuk periode tahun 2017–2021. Metode yang digunakan adalah analisis deskriptif kuantitatif dengan memanfaatkan teknik data mining melalui prapemrosesan, transformasi data, dan evaluasi model menggunakan indikator *Mean Squared Error* (MSE) serta koefisien determinasi (R^2). Temuan penelitian ini menunjukkan bahwa regresi linier memiliki tingkat akurasi yang lebih tinggi daripada regresi LASSO, yang ditandai dengan nilai MSE yang lebih rendah dan nilai (R^2) yang lebih tinggi. Selain itu, perlu dicatat bahwa pendidikan dan Indeks Pembangunan Manusia (IPM) merupakan variabel yang memiliki dampak terbesar terhadap tingkat kemiskinan di Indonesia. Dengan demikian, hasil penelitian ini dapat menjadi landasan bagi kebijakan pengentasan kemiskinan yang lebih efektif dan tepat sasaran.

Penelitian yang dilakukan (Syifana et al. (2025)) dengan judul *REGRESI LOGISTIK MULTINOMIAL BAYESIAN DENGAN ALGORITMA GIBBS SAMPLING UNTUK MENENTUKAN FAKTOR-FAKTOR TINGKAT KEMISKINAN DI INDONESIA* yang membahas dalam menganalisis faktor-faktor yang memengaruhi tingkat kemiskinan di Indonesia dengan menggunakan algoritma Gibbs Sampling *Markov Chain Monte Carlo* (MCMC) untuk regresi Bayesian multinomial. Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) tahun 2024 dengan variabel respon seperti pertumbuhan penduduk, pengangguran, PDRB, rasio Gini, rata-rata lama sekolah, dan pendapatan minimum per kapita provinsi. Hasil analisis menunjukkan bahwa setelah 20.000 iterasi, model mencapai keadaan yang stabil dan konvergen. Selain itu, berdasarkan interval kredibel dan uji deviance, ditentukan bahwa variabel yang secara signifikan mempengaruhi tingkat kemiskin-

an adalah Produk Domestik Regional Bruto (PDRB) pada harga konstan dan rata-rata lama sekolah. Hal ini menunjukkan bahwa faktor ekonomi dan pendidikan berperan penting dalam menentukan tingkat kemiskinan, sehingga regresi Bayesian dengan regresi logistik multinomial dapat memberikan model yang lebih akurat dalam mengidentifikasi faktor-faktor signifikan dibandingkan dengan regresi konvensional.

Penelitian yang dilakukan oleh (Mursyidah et al. (2025)) dengan judul *Implementation of The Logistic Regression Algorithm to Analyze Poverty Factors in Aceh Province* membahas tentang menganalisis faktor-faktor yang memengaruhi kondisi kemiskinan di Provinsi Aceh dengan menggunakan metode regresi logistik biner melalui analisis inferensial dan prediktif. Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) untuk periode 2019–2023 dengan variabel-variabel prediktor seperti tingkat pengangguran terbuka, PDRB, pengeluaran per kapita, jumlah penduduk, dan rata-rata lama sekolah. Hasil analisis menunjukkan bahwa, dari semua variabel yang diteliti, hanya rata-rata lama sekolah dan jumlah penduduk yang memiliki dampak signifikan terhadap kondisi kemiskinan, di mana peningkatan tingkat pendidikan yang signifikan dapat mengurangi tingkat kemiskinan hingga sekitar 75%, sedangkan jumlah penduduk memiliki dampak negatif yang relatif kecil. Selain itu, model klasifikasi yang dikembangkan memiliki kinerja yang sangat baik dengan akurasi sekitar 70% dan AUC sekitar 0,69, yang menunjukkan kemampuan moderat dalam membedakan antara kategori miskin dan non-miskin. Penelitian ini menegaskan bahwa faktor pendidikan memiliki peran paling dominan dalam menurunkan kemiskinan serta menunjukkan bahwa pendekatan regresi logistik efektif digunakan untuk analisis faktor kemiskinan berbasis data.

1.7. Sistematika Penulisan

Sistematika pada penelitian ini yaitu:

BAB 1 PENDAHULUAN

Pada bab ini memuat uraian mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan dan tinjauan pustaka.

BAB II DASAR TEORI

Pada bab ini berisi penjelasan mengenai landasan teori yang dijadikan rujukan, mencakup konsep-konsep dasar, hipotesis, serta rancangan penelitian.

BAB III METODE PENELITIAN

Pada bab ini membahas terkait metode penelitian yang akan digunakan dalam pemecahan masalah yang berisi tentang pendekatan penelitian, sumber data, metode pengolahan data dan diagram analisis data.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan dilakukan proses analisis data yang kemudian hasilnya diinterpretasikan secara terperinci.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dan saran dari penelitian yang dilakukan oleh penulis.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan terhadap data Kemiskinan Provinsi tahun 2024, dapat ditarik kesimpulan sebagai berikut:

1. Regresi Logistik LASSO bekerja dengan menambahkan penalti norma-L1 ke fungsi log-likelihood, sehingga kontribusi variabel yang kecil secara otomatis berkurang hingga nol. Metode ini memungkinkan pemilihan variabel dan estimasi koefisien dilakukan secara bersamaan dalam satu langkah. Berbeda dengan regresi logistik konvensional yang menghasilkan nilai p-value untuk setiap koefisien, signifikansi variabel dalam LASSO ditentukan oleh ada atau tidaknya koefisien yang bernilai nol pada λ optimal. Proses estimasi dilakukan dengan menggabungkan algoritma IRLS dan Coordinate Descent, yang menghasilkan sifat penalti L1 yang tidak dapat didiferensiasikan pada titik nol. Akibatnya, LASSO berhasil mengurangi jumlah variabel dalam model menjadi hanya tiga variabel dominan yaitu UHH, TPAK, dan PDRB.
2. Optimasi dilakukan dengan membuat solusi LASSO untuk berbagai nilai λ menggunakan paket glmnet di R dengan data yang telah distandarisasi (z-score). Selanjutnya, hitung BIC untuk setiap nilai λ menggunakan rumus $BIC(\lambda) = -2 \ell(\hat{\beta}_\lambda) + df_\lambda \ln(n)$, di mana df_λ adalah jumlah koefisien yang tidak nol. Terakhir, pilih nilai λ yang menghasilkan BIC terkecil. Berdasarkan proses ini, λ optimal adalah 0,017373, dan BIC minimum adalah 41,695.

BIC digunakan sebagai kriteria karena memberikan penalti yang lebih besar terhadap kompleksitas model ketika ukuran sampel ditingkatkan, yang mendorong penggunaan model yang lebih sederhana namun tetap akurat.

3. Penerapan metode ini pada data dari 38 provinsi menghasilkan beberapa temuan penting. Regresi logistik pertama menunjukkan bahwa tidak ada variabel signifikan (semua nilai $p > 0,05$) yang mengindikasikan multikolinearitas. Dengan IPM (24, 21), PKD (18, 99), dan UHH (18, 05) melampaui ambang batas kritis 10, Uji VIF mengonfirmasi hal ini. Setelah standarisasi dan penerapan LASSO dengan λ optimal, model akhir mencakup tiga variabel UHH (koefisien $-1,591$), TPAK (koefisien $1, 100$), dan PDRB (koefisien $-1,224$). Berdasarkan kriteria pemilihan model, model Regresi Logistik LASSO memiliki nilai BIC sekitar 41,695 dengan empat parameter, yang lebih baik daripada model Regresi Logistik yang diestimasi menggunakan metode Estimasi Kemungkinan Maksimum (MLE) sebelum seleksi variabel backward stepwise, yang memiliki nilai BIC sekitar 59,49542 dengan tujuh parameter. Hal ini menunjukkan bahwa LASSO dapat menghasilkan model yang lebih sederhana tanpa mengorbankan kecocokan model. Namun, setelah seleksi variabel menggunakan metode backward stepwise, model Regresi Logistik MLE menghasilkan nilai BIC yang sedikit lebih rendah, yaitu 40,509 dengan tiga parameter. Namun, model LASSO tetap memiliki keunggulan karena mampu mengidentifikasi variabel PDRB yang secara teoritis relevan namun tereliminasi melalui prosedur backward stepwise, serta menghasilkan estimasi koefisien yang lebih stabil dalam kondisi multikolinearitas. Oleh karena itu, tujuan analisis ini bukan hanya untuk menemukan model dengan nilai BIC terendah, tetapi juga untuk menemukan model yang memiliki prediksi

lebih robust dan tahan terhadap multikolinearitas. Oleh karena itu, model Regresi Logistik LASSO merupakan pilihan yang baik meskipun memiliki nilai BIC yang sedikit lebih tinggi.

5.2. Saran

Berdasarkan hasil penelitian ini, beberapa saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya dapat mengeksplorasi metode regularisasi lain seperti *Ridge Regression* atau *Elastic Net* untuk membandingkan kinerja dan efektivitasnya dalam mengatasi multikolinearitas pada data kemiskinan, sehingga dapat diperoleh model yang lebih *robust*.
2. Disarankan untuk melakukan analisis sensitivitas terhadap pemilihan nilai parameter regularisasi λ dengan menggunakan kriteria seleksi model lain seperti AIC (*Akaike Information Criterion*) atau *cross-validation* untuk memvalidasi stabilitas model yang dihasilkan.
3. Penelitian lanjutan dapat mempertimbangkan penambahan variabel independen lain yang relevan dengan kemiskinan, seperti tingkat literasi digital, akses terhadap teknologi informasi, atau indikator ketimpangan ekonomi, guna memperkaya analisis dan meningkatkan kemampuan prediksi model.
4. Mengingat data yang digunakan hanya mencakup satu tahun (2024), penelitian selanjutnya dapat menggunakan data panel atau data *time series* untuk menganalisis dinamika perubahan tingkat kemiskinan antar provinsi dari waktu ke waktu, sehingga dapat mengidentifikasi tren dan pola temporal.

DAFTAR PUSTAKA

- Agresti, A. (2013). *Categorical data analysis*. John Wiley & Sons.
- Alkire, S. & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of public economics*, 95(7-8):476–487.
- Becker, G. (1993). Human capital: A theoretical and empirical analysis, university of chicago press. *Chicago and london*.
- Bühlmann, P. & Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.
- Chen, J. & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771.
- Creswell, J. W. & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Fitria, E. R. & Rozci, F. (2022). Penerapan metode regresi least absolute shrinkage and selection operator (lasso) dan regresi linier untuk memprediksi tingkat kemiskinan di indonesia. *Jurnal Ilmiah Sosio Agribis*, 22(2):123–132.
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). glmnet: lasso and elastic-net regularized generalized linear models, 2010b. URL <http://CRAN.R-project.org/package=glmnet>. R package version, pages 1–1.
- Ghozali, I. (2018). Aplikasi analisis multivariate dengan ibm spss 25. Semarang: Badan Penerbit Universitas Diponegoro.

- Gujarati, D. N. (2012). *Basic econometrics* 4th ed.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis*.
- Hastie, T. (2009). *The elements of statistical learning: data mining, inference, and prediction*.
- Hosmer, D. & Lemeshow, S. (2000). *Applied logistic regression* 2nd edn John Wiley & sons. *Inc.: New York, NY, USA*, pages 160–164.
- K Robert, Y. (2018). *Case study research and applications design and methods*. Library of Congress Cataloging-in-Publication Data.
- Kamila, D. R., Imro'ah, N., & Sulistianingsih, E. (2021). Metode least absolute shrinkage and selection operator (lasso) untuk pendugaan parameter regresi logistik biner (studi kasus: Faktor-faktor tingkat kemiskinan di Indonesia tahun 2021). *BIMASTER: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 14(1).
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill/Irwin, New York, 5th edition, ISBN: 9780073108742.
- Mankiw, N. G. (1998). *Principles of Economics*. South-Western College Publishing, 3rd edition.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, 2 edition.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Wiley, 5 edition.

- Mursyidah, M., Dinata, R. K., & Yunizar, Z. (2025). Implementation of the logistic regression algorithm to analyze poverty factors in aceh province. *Journal of Applied Informatics and Computing*, 9(4):1451–1462.
- Nelder, J. A. & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384.
- Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, pages 758–765.
- Park, M. Y. & Hastie, T. (2007). L 1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(4):659–677.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, pages 111–163.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.
- Searle, S. R. (1997). *Linear models*. John Wiley & Sons.
- Stake, R. (1995). *The art of case study research*, sage publications, inc.
- Syifana, H., Gusriani, N., & Parmikanti, K. (2025). Regresi logistik multinomial bayesian dengan algoritma gibbs sampling untuk menentukan faktor-faktor tingkat kemiskinan di indonesia. *Jurnal Matematika Integratif*, 21(1):89–102.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.
- Todaro, M. P. & Smith, S. C. (2009). *Economic development*. Pearson education.

- Townsend, P. (1987). Deprivation. *Journal of social policy*, 16(2):125–146.
- Wang, H., Li, B., & Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(3):671–683.
- Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika*, 61(3):439–447.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.