METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO) UNTUK PENDUGAAN PARAMETER REGRESI LOGISTIK BINER
(Studi Kasus: Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021)

DIVA RAHMA KAMILA
NIM H1091211022

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO) UNTUK PENDUGAAN PARAMETER REGRESI LOGISTIK BINER
(Studi Kasus: Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021)

DIVA RAHMA KAMILA
NIM H1091211022

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk Pendugaan Parameter Regresi Logistik Biner (Studi Kasus: Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021)
Nama Mahasiswa	: Diva Rahma Kamila
NIM	: H1091211022
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 6 Maret 2025
SK Pembimbing	: No. 2601/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	: No. 568/UN22.8/TD.06/2025/Tanggal 24 Februari 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Nurfitri Imro'ah, M.Si.	Dr. Evy Sulistianingsih, M.Sc.
NIP 198907182019032021	NIP 198502172008122006

Dosen Penguji

Ketua Penguji	Anggota Penguji
Shantika Martha, M.Si.	Neva Satyahadewi, M.Sc.,CRA.,CRP.,CRMP.
NIP 198403082008122003	NIP 198212042005012001

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Nurfitri Imro'ah, M.Si.	Dr. Evy Sulistianingsih, M.Sc.
NIP 198907182019032021	NIP 198502172008122006

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama	: Diva Rahma Kamila
NIM	: H1091211022
Program Studi/ Jurusan	: Statistika/Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 6 Maret 2025



Diva Rahma Kamila
H1091211022

Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk Pendugaan Parameter Regresi Logistik Biner (Studi Kasus: Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021)

Abstrak

Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan pendekatan regresi yang efektif dalam mengatasi multikolinieritas dengan meminimalkan jumlah kuadrat galat dan menerapkan batasan penalti. Batasan ini memungkinkan LASSO mengecilkan beberapa koefisien hingga nol, sekaligus melakukan seleksi variabel. Salah satu isu strategis yang dapat dianalisis dengan metode LASSO adalah kemiskinan. Kemiskinan berkaitan erat dengan kesejahteraan masyarakat dan menjadi tantangan utama di negara berkembang seperti Indonesia, yang dipengaruhi oleh berbagai faktor. Penelitian ini menerapkan metode LASSO untuk mengatasi multikolinieritas dalam regresi logistik biner terkait faktor-faktor yang mempengaruhi persentase tingkat kemiskinan di Indonesia pada 2021. Tahapan penelitian mencakup pengumpulan data, penerapan metode, seleksi variabel yang signifikan, dan analisis hasil. Hasil menunjukkan bahwa metode LASSO efektif dalam menangani multikolinearitas melalui seleksi variabel, dengan variabel yang masuk dalam model adalah jumlah penduduk miskin, umur harapan hidup, pengeluaran per kapita, dan PDRB serta pengeluaran per kapita dan PDRB sebagai faktor signifikan. Nilai kesalahan klasifikasi model atau *Apparent Error Rate* (APER) yang diperoleh adalah sebesar 20,5882% dan ketepatan klasifikasi sebesar 79,4118%. Artinya, secara keseluruhan terdapat sebanyak 27 dari 34 provinsi dapat diprediksi secara tepat dengan model yang diperoleh. Sehingga dapat disimpulkan bahwa pengklasifikasian persentase tingkat kemiskinan berdasarkan yang berada di atas rata-rata dan di bawah rata-rata persentase tingkat kemiskinan dalam model regresi logistik biner LASSO memiliki kriteria cukup baik.

Kata kunci: Kemiskinan, Regresi Logistik Biner, LASSO, Multikolinearitas.

Least Absolute Shrinkage and Selection Operator (LASSO) Method for Binary Logistic Regression Parameter Estimation
(Case Study: Factors Poverty Levels in Indonesia in 2021)

Abstract

The Least Absolute Shrinkage and Selection Operator (LASSO) method is an effective regression approach in overcoming multicollinearity by minimizing the number of squares of errors and applying penalty constraints. This constraint allows LASSO to shrink some coefficients to zero, while performing variable selection. One of the strategic issues that can be analyzed with the LASSO method is poverty. Poverty is closely related to people's welfare and is a major challenge in developing countries such as Indonesia, which is influenced by various factors. This study applies the LASSO method to overcome multicollinearity in binary logistic regression related to factors that affect the percentage of poverty rate in Indonesia in 2021. The research stages include data collection, application of methods, selection of significant variables, and result analysis. The results show that the LASSO method is effective in dealing with multicollinearity through variable selection, with the variables included in the model being the number of poor population, life expectancy, per capita expenditure, and GDP as well as per capita expenditure and GDP as significant factors. The model classification error value or Apparent Error Rate (APER) obtained was 20.5882% and the classification accuracy was 79.4118%. This means that overall there are 27 out of 34 provinces that can be predicted accurately with the model obtained. Therefore, it can be concluded that the classification of the poverty rate percentage based on those above the average and below the average poverty rate percentage in the LASSO binary logistics regression model has quite good criteria.

Keywords: *Poverty, Binary Logistic Regression, LASSO, Multicollinearity*.

PRAKATA

Puji dan syukur atas kehadirat Allah SWT atas rahmat, hidayah, dan karunia-Nya serta shalawat serta salam selalu dijunjungkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk Pendugaan Parameter Regresi Logistik Biner (Studi Kasus: Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021)"**. Penulisan skripsi ini bertujuan untuk memenuhi syarat dalam memperoleh gelar Sarjana Statistika pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura. Dalam penyusunan skripsi ini, tentu tidak lepas dari motivasi, arahan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua penulis, Bapak Haryono, S.T. dan Ibu Henny Wardhany, S.Pd., serta saudari penulis, Devi Ardhia Pramesti, yang selama ini telah mendoakan, menyemangati, dan memberikan dukungan moril serta material kepada penulis dalam menyelesaikan skripsi.
2. Ibu Nurfitri Imro'ah, M.Si., selaku Dosen Pembimbing Akademik yang sedari awal selalu memberikan bimbingan, arahan, dan masukan selama masa perkuliahan dan selaku Dosen Pembimbing Pertama yang telah membimbing, memberi arahan, saran dan masukan yang membangun, memberi semangat dan memotivasi selama penulisan skripsi ini.
3. Ibu Dr. Evy Sulistianingsih, M.Sc., selaku Dosen Pembimbing Kedua yang telah membimbing, memberi arahan dan masukan, memberi semangat dan memotivasi dalam menyelesaikan penulisan skripsi ini.
4. Ibu Shantika Martha, M.Si., selaku Dosen Penguji Pertama yang telah memberikan masukan dan saran untuk perbaikan skripsi sehingga menjadi lebih baik.

5. Ibu Neva Satyahadewi, M.Sc., CRA., CRP., CRMP., selaku Dosen Penguji Kedua yang telah memberikan masukan dan saran untuk perbaikan skripsi sehingga menjadi lebih baik.
6. Teman-teman SMP penulis, Alayli, Desta, Ara, Aulia, dan Claudia, yang sama-sama berjuang mendapatkan gelar di jalannya masing-masing dan saling memberi *support* dalam menyelesaikan skripsi.
7. Teman-teman SMA penulis, Hana, Aulia, Nadya, Hafiza, dan Nindy, yang selalu berjuang dalam meraih gelar di tempat masing-masing dan saling memberi *support* dalam menyelesaikan skripsi.
8. Teman-teman penulis selama perkuliahan, Wina, Fadhela, Hana, dan Mira, yang selalu memberi motivasi, dukungan, dan saling berbagi ilmu dalam penyelesaian skripsi ini.
9. Teman-teman Expos Statistika Universitas Tanjungpura Angkatan 2021 yang telah membersamai selama perkuliahan dan mendukung satu sama lain dalam penyelesaian skripsi.
10. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu dalam penyusunan skripsi ini.

Penulis berharap dapat memperoleh kritik dan saran yang bersifat membangun dari semua pihak untuk kesempurnaan skripsi ini. Semoga skripsi ini dapat memberikan manfaat bagi siapa saja yang membacanya serta memberikan wawasan pemikiran yang lebih luas.

Pontianak, 6 Maret 2025

Diva Rahma Kamila

DAFTAR ISI

LEMBAR PENGESAHAN ..... **ii**
PERNYATAAN INTEGRITAS AKADEMIK ..... **iii**
Abstrak ..... **iv**
Abstract ..... **v**
PRAKATA ..... **vi**
DAFTAR ISI ..... **viii**
DAFTAR TABEL ..... **x**
DAFTAR LAMPIRAN ..... **xi**
DAFTAR SIMBOL ..... **xii**
BAB I PENDAHULUAN **1**
1.1. Latar Belakang 1
1.2. Rumusan Masalah 3
1.3. Tujuan Penelitian 3
1.4. Batasan Masalah 3
1.5. Tinjauan Pustaka 4
1.6. Metodologi Penelitian 5
BAB II LANDASAN TEORI **9**
2.1. Regresi Logistik Biner 9
2.2. Uji Multikolinearitas 11
2.3. Kemiskinan 12
BAB III REGRESI LOGISTIK BINER DENGAN METODE LASSO **14**
3.1. *Least Absolute Shrinkage and Selection Operator* (LASSO) 14
3.2. Algoritma *Cyclic Coordinate Descent* (CCD) 16
3.3. Estimasi *Standard Error* dengan Metode *Bootstrap* 17
3.4. Uji Signifikansi Parsial Parameter Regresi Logistik 18
3.5. Ketepatan Klasifikasi Regresi Logistik 19
3.6. Interpretasi Model Regresi Logistik Biner 21

BAB IV HASIL DAN PEMBAHASAN 24
4.1. Data Penelitian 24
4.2. Statistik Deskriptif 25
4.3. Pendeteksian Multikolinearitas 26
4.4. Pendugaan Parameter Regresi Logistik Biner Menggunakan Metode LASSO 27
4.5. Estimasi *Standard Error* Penduga Parameter dengan Metode *Bootstra*p 29
4.6. Uji Signifikansi Parameter 30
4.7. Ketepatan Klasifikasi 31
4.8. Interpretasi Koefisien Model 32
BAB V PENUTUP 34
5.1. Kesimpulan 34
5.2. Saran 34
DAFTAR PUSTAKA 36
LAMPIRAN 39

DAFTAR TABEL

Tabel 3.1 Pembagian Kategori Hasil Prediksi 20
Tabel 3.2 Matriks Konfusi untuk Dua Kelas 20
Tabel 3.3 Klasifikasi Model Berdasarkan Nilai Akurasi 21
Tabel 4.1 Variabel Independen Penelitian 24
Tabel 4.2 Pengelompokan Persentase Tingkat Kemiskinan 25
Tabel 4.3 Statistik Deskriptif 25
Tabel 4.4 Nilai VIF Setiap Variabel Independen 26
Tabel 4.5 Penduga Parameter dengan LASSO 27
Tabel 4.6 *Standard Error* dengan Metode *Bootstrap* 29
Tabel 4.7 Hasil Uji Signifikansi Parameter 30
Tabel 4.8 Ketepatan Klasifikasi 31
Tabel 4.9 *Odds Ratio* Setiap Variabel Independen Model LASSO 32

DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian 39
Lampiran 2 Perbandingan Hasil Klasifikasi *Y* 41

DAFTAR SIMBOL

$\pi(x_i)$: Peluang kejadian sukses dengan nilai probabilitas $0 \leq \pi(x_i) \leq 1$

x_{ji} : Nilai variabel independen ke- j pengamatan ke- i, dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$

y_i : Nilai variabel dependen pengamatan ke- i

β_0 : Koefisien konstanta (intersep)

β_j : Koefisien parameter regresi logistik variabel independen ke- j

n : Banyaknya pengamatan

p : Banyaknya variabel independen

Δ_j : Hampiran penduga β_j

Δ_{vj} : Nilai perubahan parameter β_j

$\hat{\theta}(.)$: Penduga parameter *Bootstrap*

$\hat{\theta}(x^{*b})$: Penduga parameter *Bootstrap* saat terjadi pengulangan sebanyak $b = 1, 2, \dots, B$

B : Banyaknya ulangan *Bootstrap*

$\widehat{\beta_j}$: Penduga β_j

$SE(\widehat{\beta_j})$: *Standard error* penduga β_j

$S_j(\beta)$: Turunan pertama fungsi *log-likelihood*

$Q(\beta_j, \Delta_j)$: Turunan kedua fungsi *log-likelihood*

BAB I
PENDAHULUAN

1.1. Latar Belakang

Analisis regresi merupakan analisis yang bertujuan untuk mengetahui hubungan antara satu atau lebih variabel independen dengan satu variabel dependen (Munawaroh et al., 2015). Analisis regresi dibagi menjadi dua jenis, yaitu analisis regresi sederhana dan analisis regresi berganda. Analisis regresi sederhana bertujuan untuk memprediksi nilai suatu variabel dependen berdasarkan nilai variabel independennya. Banyaknya variabel dependen dan variabel independen untuk analisis regresi sederhana adalah satu dan tidak bisa lebih dari itu (Sartika et al., 2020). Sedangkan, analisis regresi berganda bertujuan untuk memprediksi bagaimana dua atau lebih variabel independen berhubungan dengan satu variabel dependen.

Salah satu cara untuk menganalisis variabel dependen yang bersifat kategorik adalah melalui model regresi logistik. Regresi logistik terdiri atas model regresi logistik biner, ordinal, dan multinomial. Hubungan antara variabel dependen yang berupa data dikotomi atau biner dengan variabel independen yang berupa data interval dan atau kategorik dapat dijelaskan melalui regresi logistik biner (Pratama, 2019). Variabel dikotomi atau biner hanya memiliki dua kategori, yaitu kategori untuk kejadian sukses dan kategori untuk kejadian gagal.

Pemodelan regresi logistik harus memenuhi asumsi klasik untuk mendapatkan model regresi yang baik, yaitu asumsi non multikolinearitas. Multikolinearitas adalah kejadian di mana terdapat keterkaitan linier antara berbagai variabel independen, baik itu berupa variabel kontinu atau variabel kategorik (Yaldi et al., 2022). Jika terdapat multikolinearitas dalam analisis regresi, hal ini menjadi masalah serius karena dapat menjadi masalah dalam model yang terbentuk. Variansi bias pendugaan yang besar diakibatkan karena adanya multikolinearitas (Jamco et al., 2023). Parameter yang diestimasi cenderung memiliki *standard error* yang tinggi, yang mana dapat

mengurangi akurasi dari estimasi parameter. Kemungkinan banyaknya parameter yang tidak signifikan dalam pengujian parsial terhadap variabel dependen disebabkan oleh tingginya *standar error* tersebut.

Permasalahan multikolinearitas dapat diatasi dengan pendekatan agar menghasilkan model terbaik. Salah satu pendekatan yang dapat digunakan adalah metode *Least Absolute Shrinkage and Selection Operator* (LASSO). Metode LASSO merupakan metode regresi yang mengestimasi parameter dengan cara meminimalkan jumlah kuadrat galat dengan memperhatikan suatu batasan (Mahalani & Rifai, 2022). Dengan adanya batasan tersebut, LASSO dapat mengurangi sejumlah koefisien hingga mencapai nol dan secara bersamaan melakukan seleksi variabel (Prabowo et al., 2015). Penerapan metode LASSO memerlukan pemahaman masalah yang menjadi fokus dalam penelitian, yaitu faktor apa saja yang berpengaruh terhadap masalah penelitian. Salah satu isu strategis yang dapat dianalisis menggunakan metode ini adalah kemiskinan.

Kemiskinan memiliki hubungan erat dengan masalah kesejahteraan masyarakat yang merupakan salah satu masalah utama yang sering dihadapi negara berkembang seperti Indonesia. Pada September 2021, Badan Pusat Statistik (BPS) melaporkan bahwa jumlah penduduk miskin di Indonesia mencapai 26,50 juta orang, yang setara dengan 9,71% dari total populasi Indonesia. Di Indonesia sendiri terdapat banyak faktor yang dapat mempengaruhi tingkat kemiskinan, seperti jumlah penduduk miskin, sumber daya manusia, pendidikan, angka harapan hidup, pengeluaran per kapita, serta PDRB dari tiap provinsi di Indonesia. Penelitian ini akan mengkaji penggunaan metode LASSO dalam model regresi logistik biner pada persentase tingkat kemiskinan setiap provinsi di Indonesia pada tahun 2021. Berdasarkan analisis penggunaan metode LASSO yang dilakukan diharapkan dapat mengetahui faktor apa saja yang berkontribusi secara signifikan terhadap tingkat kemiskinan yang bermanfaat dalam menentukan kebijakan sosial yang tepat bagi pemerintah.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan maka rumusan masalah yang diangkat dalam penelitian ini adalah:

1. Bagaimana metode *Least Absolute Shrinkage and Selection Operator* (LASSO) diterapkan untuk menduga parameter pada model regresi logistik biner, terutama pada data yang mengalami multikolinearitas?
2. Bagaimana analisis faktor-faktor yang berpengaruh secara signifikan terhadap tingkat kemiskinan di Indonesia?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan maka tujuan yang ingin dicapai melalui penelitian ini adalah:

1. Menerapkan metode *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk menduga parameter pada model regresi logistik biner, khususnya pada data yang mengalami multikolinearitas.
2. Menganalisis faktor-faktor yang berpengaruh secara signifikan terhadap tingkat kemiskinan provinsi di Indonesia.

1.4. Batasan Masalah

Batasan masalah dalam penelitian ini, yaitu:

1. Penelitian ini memfokuskan pada pembahasan model regresi logistik biner yang mengandung multikolinearitas pada data.
2. Data yang digunakan adalah data persentase tingkat kemiskinan provinsi di Indonesia Tahun 2021, sebelum adanya pemekaran provinsi di Indonesia, yang bersumber dari Badan Pusat Statistik (BPS).