

SKRIPSI

**OPTIMISASI ANALISIS REGRESI LASSO DENGAN
ALGORITMA LARS MENGGUNAKAN BIC
(STUDI KASUS : JUMLAH PENDUDUK MISKIN DI PROVINSI
JAWA TIMUR TAHUN 2022)**



USWATUN KHASANAH

20106010018

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

2024

**OPTIMISASI ANALISIS REGRESI LASSO DENGAN
ALGORITMA LARS MENGGUNAKAN BIC
(STUDI KASUS : JUMLAH PENDUDUK MISKIN DI PROVINSI
JAWA TIMUR TAHUN 2022)**

Skripsi

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



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

diajukan oleh

USWATUN KHASANAH

20106010018

Kepada

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

2024



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 : Uswatun Khasanah

NIM : 20106010018

Judul Skripsi : Optimisasi Analisis Regresi LASSO dengan Algoritma LARS
menggunakan BIC

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, 23 Juli 2024

Pembimbing

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

NIP. 19741003 200003 2 002

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Uswatun Khasanah
NIM : 20106010018
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, 23 Juli 2024



Uswatun Khasanah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



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-1618/Un.02/DST/PP.00.9/08/2024

Tugas Akhir dengan judul : Optimisasi Analisis Regresi LASSO dengan Algoritma LARS menggunakan BIC (Studi Kasus : Jumlah Penduduk Miskin di Provinsi Jawa Timur Tahun 20)

yang dipersiapkan dan disusun oleh:

Nama : USWATUN KHASANAH
Nomor Induk Mahasiswa : 20106010018
Telah diujikan pada : Selasa, 20 Agustus 2024
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: 66cbf8958caad



Penguji I
Deddy Rahmadi, M.Sc.
SIGNED

Valid ID: 66cc28143b190



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

Valid ID: 66cc17472bf3d



Yogyakarta, 20 Agustus 2024
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 66cc395e1176c

HALAMAN PERSEMBAHAN

KARYA SEDERHANA INI PENULIS PERSEMBAHKAN UNTUK:
KEDUA ORANG TUA PENULIS BAPAK SURIP DAN ALMH. IBU HARTINI,
KAKAK PENULIS RARAS KURNIANINGSIH, KELUARGA BESAR, SAHABAT
DAN SEMUA ORANG-ORANG YANG SELALU MEMBERIKAN
SUPPORT, DO'A, DAN SEMANGAT.
ALMAMATER KAMPUS TERCINTA YANG BANYAK MEMBERIKAN
CERITA SELAMA 4 TAHUN MENUNTUT ILMU DI UNIVERSITAS ISLAM
NEGERI SUNAN KALIJAGA YOGYAKARTA.
"MATEMATIKA ANGKATAN 2020 TERCINTA"



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

"Sesungguhnya sholatku, ibadahku, hidupku, dan matiku hanyalah untuk Allah, Tuhan seluruh alam."

"Allah tidak membebani seseorang melainkan sesuai dengan kadar kesanggupannya"

(Al-Baqarah:286)

"Janganlah engkau bersedih, sesungguhnya Allah bersama kita."

(At-Taubah:40)

"Nikmati prosesnya, jalani dan ikuti arusnya. Terkait hasil, kita serahkan pada yang Maha Kuasa."

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PRAKATA

Allhamdulillahirabbil'alamin, puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya yang tak ternilai berupa keimanan, kesabaran, kekuatan dan kelancaran. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Optimisasi Analisis Regresi LASSO dengan Algoritma LARS menggunakan BIC."

Penulis mengucapkan terima kasih kepada pembimbing skripsi penulis, Ibu Sri Utami Zuliana, S.Si.,M.Sc.,Ph.D., atas bimbingan, saran serta masukannya kepada penulis demi terselesaikannya skripsi ini. Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat berbagai kendala dan kekurangan. Namun, berkat dukungan, motivasi, serta bantuan dari berbagai pihak, *alhamdulillah* skripsi ini dapat terselesaikan. Dengan penuh rasa syukur, penulis ingin mengucapkan terima kasih dan apresiasi yang sebesar-besarnya kepada:

1. Bapak Prof. Noorhaidi, M.A., M.Phil., Ph.D., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Prof. Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Bapak Muchammad Abrori, S.Si.,M.Kom., selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
4. Ibu Aulia Khifah Futhona, M.Sc., selaku dosen pembimbing akademik yang telah banyak memberikan pengarahan kepada penulis dari awal perkuliahan

hingga akhir.

5. 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.
6. Ibu tercinta almh. Hartini, seseorang yang biasa saya sebut mamak. Kepergianmu membuatku mengerti bahwa rindu yang paling menyakitkan adalah merindukan seseorang yang telah tiada, ragamu memang tidak disini ragamu memang sudah tidak ada dan tidak bisa aku jangkau tapi namamu akan tetap jadi motivasi terkuat sampai detik ini. Ibu alhamdulillah kini penulis sudah berada di tahap ini, menyelesaikan karya tulis sederhana ini sebagai perwujudan terakhir sebelum engkau benar-benar pergi. Terimakasih sudah mengantarkan saya berada ditempat ini, walaupun pada akhirnya saya harus berjuang tertatih sendiri tanpa kau temani lagi.
7. Bapak Surip, beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan. Namun beliau mampu mendidik penulis, memotivasi, memberikan dukungan hingga penulis mampu menyelesaikan studinya sampai sarjana.
8. Saudara kandung penulis Raras Kurnianingsih, S.E., yang telah memberikan semangat, doa, dan cinta sehingga penulis sampai ke titik ini.
9. Keluarga Matematika angkatan 2020 yang telah kebersamai selama 4 tahun dibangku perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.
10. Teman-teman seperbimbingan Adelia, Anisa, Linggar, Yumna, Sari, Fatma yang telah bersama-sama menemani selama bimbingan. Akhirnya selesai juga ya.

11. Teman-teman SMA ataupun dunia maya yang telah menyemangati serta memberikan dukungan dan bantuan hingga penulis dapat menyelesaikan dengan baik.
12. Terakhir semua pihak yang penulis tidak bisa sebutkan satu-persatu, semoga dimudahkan segala urusannya.

Penulis sangat menyadari tidak ada kesempurnaan pada skripsi ini. Oleh karena itu, penulis menerima segala bentuk saran dan kritik yang membangun untuk penulis demi mengembangkan penelitian. Penulis berharap skripsi ini dapat memberikan manfaat khususnya bagi penulis dan semua kalangan yang membutuhkan untuk dijadikan referensi dalam menulis.

Yogyakarta, 23 Juli 2024

Uswatun Khasanah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PERNYATAAN KEASILAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMBANG	xv
INTISARI	xvii
ABSTRACT	.xviii
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	4
1.4. Manfaat Penelitian	5
1.5. Sistematika Penulisan	5
1.6. Tinjauan Pustaka	6
II DASAR TEORI	9
2.1. Analisis Regresi Linear	9
2.2. Asumsi Klasik	10
2.3. <i>Maximum Likelihood Estimaton</i> (MLE)	12
2.4. Regresi Terpenalti	14
2.5. <i>Least Absolute Shrinkage and Selection Operator</i> (LASSO)	16
2.6. <i>Least Angle Regression</i> (LARS)	17

2.7. <i>Bayesian Information Criterion (BIC)</i>	18
III METODE PENELITIAN	20
3.1. Metode Penelitian	20
3.2. Studi Kasus	20
3.3. Variabel Penelitian	20
3.4. Software yang Digunakan	21
3.5. Metode Pengolahan Data	21
3.6. Source code	22
3.7. Diagram Analisis Data (<i>Flowchart</i>)	22
IV HASIL DAN PEMBAHASAN	24
4.1. Statistik Deskriptif	24
4.2. Standarisasi Data	26
4.3. Analisis Regresi Linear	28
4.3.1. Uji Asumsi	28
4.3.2. Estimasi Analisis Regresi Linear	32
4.4. Uji Signifikansi	34
4.4.1. Uji F (Uji Serentak)	34
4.4.2. Uji t (Uji Parsial)	35
4.5. Analisis Regresi LASSO dengan Algoritma LARS	36
4.6. Menentukan Tahapan Masuknya Variabel Bebas ke Dalam Model	36
4.7. Menentukan Model LASSO	39
4.8. Menentukan Nilai Lambda Optimal dengan BIC	41
4.9. Pemilihan Model Terbaik Regresi LASSO	42
4.10. Evaluasi Model dengan <i>Mean Square Error</i> (MSE)	44
V PENUTUP	45
5.1. Kesimpulan	45
5.2. Saran	46
DAFTAR PUSTAKA	47
LAMPIRAN	49
A Data Penelitian	49

B Model Regresi LASSO	51
C Nilai Lambda dan Nilai BIC	53
D <i>Source Code</i> Program R	55
E TABEL TINJAUAN PUSTAKA	61
F Biodata Penulis	63



DAFTAR TABEL

3.1	Variabel Penelitian	21
4.1	Statistik Deskriptif	26
4.2	Output Uji Normalitas	28
4.3	Output Uji Autokorelasi	29
4.4	Output Uji Heteroskedastisitas	30
4.5	Nilai VIF	31
4.6	Nilai p -value Setiap Variabel	35
4.7	Tahapan-Tahapan Masuknya Variabel Bebas pada Regresi LASSO .	37
4.8	Koefisien (β) Variabel Setiap Tahapan Seleksi Variabel Bebas ke- dalam Model	39
4.9	Model Regresi LASSO	40
4.10	Model Regresi LASSO dengan BIC (λ)	43
4.11	Koefisien Regresi LASSO Berdasarkan BIC	43
4.12	Nilai MSE	44

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

3.1	Diagram Alir Penelitian	23
4.1	Boxplot Data Setiap Variabel	25
4.2	Boxplot Data Setiap Variabel	27
4.3	Ilustrasi Plot Tahapan Variabel Bebas kedalam Model	37
4.4	Plot Nilai BIC dan Lambda	42



DAFTAR LAMBANG

Y_1	= Variabel tak bebas
X_1, X_2, \dots, X_n	= variabel bebas
β_0	= konstanta
β_1, \dots, β_n	= parameter regresi
ε	= error
Y	= vektor variabel respon ukuran $n \times 1$
X	= matrix variabel prediktor ukurn $n \times (K+1)$
β	= vektor parameter regresi ukuran $(k+1) \times 1$
ε_i	= vektor residual ukuran $n \times 1$
L	= Kuadrat residual
$\varepsilon_i^T \varepsilon_i =$	perkalian matriks <i>error</i>
$\hat{\beta}$	= vektor koefisien
n	= Jumlah observasi
p	= Banyaknya variabel independen di tambah intersep
σ^2	= Varians sisaan
$Var(b_{k-1})$	= Varians b_{k-1}
r_{XX}	= Korelasi antar variabel independen
R_j^2	= koefisien determinasi antar X_j dengan variabel yang lain.
j	= banyaknya model regresi.
$f(t_i)$	= fungsi <i>spline</i> dengan pendekatan titik knot

- t = nilai parameter tuning dengan $t \leq 0$
- k = jumlah parameter yang diestimasi dalam model regresi
- θ = Parameter yang diestimasi
- $L(\theta)$ = Fungsi Likelihood



INTISARI

Optimisasi Analisis Regresi LASSO dengan Algoritma LARS Menggunakan BIC

(Studi kasus : Jumlah Penduduk Miskin di Provinsi Jawa Timur Tahun 2022)

Uswatun Khasanah

20106010018

Salah satu metode pemodelan yang umumnya digunakan adalah dengan regresi linear. Untuk membuat model linear harus memenuhi beberapa asumsi. Dalam pengujian asumsi tersebut terdapat masalah-masalah yang sering muncul, salah satunya adalah multikolinearitas, yang terjadi ketika antara variabel independen memiliki korelasi yang kuat. Salah satu cara untuk mengatasi masalah multikolinearitas yaitu dengan metode regresi *Least Absolute Shrinkage And Selection Operator* (LASSO). Metode ini dapat menyusutkan koefisien regresi pada variabel bebas yang memiliki korelasi dan error yang tinggi sampai mendekati nol atau sama dengan nol. Koefisien regresi LASSO dicari dengan menggunakan algoritma *Least Angle Regression* (LARS). Pemilihan lambda optimal pada regresi LASSO pada penelitian ini menggunakan kriteria *Bayesian Information Criterion* (BIC), model yang dipilih diidentifikasi dengan nilai minimum dari BIC. Penelitian ini akan membahas penerapan metode regresi LASSO pada data jumlah penduduk miskin di Jawa Timur tahun 2022. Dalam penyelesaiannya, regresi LASSO akan menghasilkan 79 nilai lambda, kemudian dipilih lambda yang optimal dengan nilai BIC sebesar 2715.977. Dengan dua dari lima variabel independen yang masuk ke dalam model yaitu rata-rata lama sekolah (RLS) dan tingkat pengangguran terbuka (TPT).

Kata Kunci: Multikolinearitas, LASSO, LARS, BIC.

ABSTRACT

Optimization of LASSO Regression Analysis with LARS Algorithm Using BIC

(Case study : The number of Poor People in East Java Province in 2022)

Uswatun Khasanah

20106010018

One commonly used modeling method is linear regression. To create a linear model, several assumptions must be met. In testing these assumptions there are problems that often arise, one of which is multicollinearity, which occurs when the independent variables have a strong correlation. One way to solve the multicollinearity problem is with the regression method Least Absolute Shrinkage And Selection Operator (LASSO). This method can shrink the regression coefficients on independent variables that have high correlation and error until they are close to zero or equal to zero. The LASSO regression coefficient is found using the LAST Angle Regression (LARS) algorithm. The selection of the optimal lambda in LASSO regression in this study uses the Bayesian Information Criterion (BIC) criterion, the selected model is identified by the minimum value of the BIC. This research will discuss the application of the LASSO regression method on data on the number of poor people in East Java in 2022. In its completion, LASSO regression will produce 79 lambda values, then the optimal lambda is selected with a BIC value of 2715.977. With two of the five independent variables included in the model, namely the average length of schooling (RLS) and the open unemployment rate (TPT).

Keywords: Multicollinearity, LASSO, LARS, BIC.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Analisis regresi merupakan salah satu analisis data yang digunakan untuk menentukan hubungan antara variabel dependen dan satu atau lebih variabel independen. Selain itu, variabel independen mana sajakah yang berpengaruh secara signifikan terhadap variabel dependen dapat diketahui. Sebagai tambahan, nilai variabel dependen dapat diestimasi melalui informasi yang diperoleh dari variabel-variabel independennya.

Model analisis regresi didapatkan dengan mengestimasi parameter. Koefisien regresi dapat diestimasi dengan menggunakan *Maximum Likelihood Estimation* (MLE). Perlu adanya empat asumsi yang bersifat BLUE (*Best, Linear, Unbiased, and Estimator*).

Cara untuk melakukan estimasi parameter pada MLE yaitu dengan mencari fungsi parameter yang memaksimumkan fungsi likelihood. Metode maksimum *likelihood* ini harus melakukan uji asumsi klasik untuk mendapatkan model terbaik. Terdapat empat uji asumsi diantaranya adalah uji normalitas, uji heterokedastisitas, uji autokorelasi dan uji multikolinearitas. Salah satu masalah yang sering muncul yaitu multikolinearitas. Multikolinearitas adalah masalah yang terjadi ketika antara dua atau lebih variabel independen memiliki korelasi atau hubungan kuat. Multikolinearitas dapat ditunjukkan dengan adanya korelasi yang cukup tinggi antara beberapa variabel bebas dengan melihat nilai VIF (*Variance Inflation Factor*) (Gujarati, 1978). Salah satu cara untuk mengatasi masalah multikolinearitas adalah dengan

metode regresi *Least Absolute Shrinkage And Selection Operator* (LASSO).

Menurut (Pekhimenko, 2006) regresi terpenalti yang umum digunakan terdapat dua jenis yaitu regresi L_1 (Lasso) dan regresi L_2 (Ridge). Metode regresi LASSO adalah suatu pengembangan dari regresi ridge yang diperkenalkan oleh Tibshirani pada tahun 1996. Regresi *Least Absolute Shrinkage And Selection Operator* (LASSO) merupakan metode yang dapat menyusutkan koefisien regresi pada variabel independen yang memiliki korelasi tinggi dan *error* yang besar hingga mendekati nol atau sama dengan nol (Tibshirani, 1996). Perbedaan regresi LASSO dan regresi ridge adalah LASSO dapat menyusutkan nilai koefisien regresi menjadi tepat nol, sedangkan ridge hanya dapat menyusutkan koefisien regresi sampai mendekati nol (Tibshirani, 1996). Keakuratan perhitungan metode regresi LASSO dengan menggunakan algoritma *Least Angle Regression* (LARS).

Algoritma *Least Angle Regression* (LARS) ditemukan oleh Bradley Efron, Trevor Hastie, Johnstone, dan Robert Tibshirani pada tahun 2002. Algoritma *Least Angle Regression* (LARS) merupakan algoritma yang dapat dimodifikasi menjadi algoritma komputasi pada metode regresi LASSO dan sebuah algoritma untuk menghasilkan model linear. Hasil modifikasi ini menghasilkan algoritma yang lebih efisien dalam menduga koefisien LASSO. Algoritma LARS secara bertahap mengarahkan setiap variabel prediktor menuju nilai optimalnya, sehingga mengurangi korelasi dengan kesalahan secara berkelanjutan dalam nilai absolut. Salah satu metode yang dapat digunakan untuk pemilihan model terbaik regresi LASSO adalah metode *Bayesian Information Criterion* (BIC).

Metode *Bayesian Information Criterion* (BIC) merupakan suatu metode yang digunakan untuk pemilihan model statistik. Model yang dipilih dalam pemilihan diidentifikasi dengan nilai minimum dari BIC. Nilai BIC dihitung dengan mengguna-

kan *likelihood* dari model dan jumlah parameter model. Metode BIC diperkenalkan oleh Gideon Schwarz pada tahun 1978. Metode BIC memiliki sifat konsistensi, yang berarti bahwa saat ukuran sampel mendekati tak terbatas, maka semakin besar ukuran sampel, semakin mendekati kebenaran estimasi yang diperoleh dari model tersebut, metode BIC akan cenderung memilih model yang memiliki struktur yang benar.

Beberapa data kemiskinan terdapat multikolinearitas, maka pada penelitian ini menggunakan data kemiskinan. Kemiskinan adalah ketidakmampuan seseorang dalam memenuhi kebutuhan dasar seperti makanan, pakaian, pendidikan, tempat tinggal dan kesehatan karena tingginya tingkat ketergantungan dan rendahnya jumlah orang yang bekerja. Kemiskinan menjadi masalah penting yang masih dihadapi oleh hampir di setiap negara, termasuk di Indonesia. Kemiskinan di Indonesia, khususnya di Jawa Timur menjadi masalah serius yang harus dihadapi.

Penyebab kemiskinan di suatu wilayah dapat dianalisis menggunakan pendekatan analisis spasial yang digunakan untuk mengidentifikasi pengaruh faktor geografis di suatu wilayah. Perbedaan lokasi geografis akan mempengaruhi potensi yang dimiliki atau dimanfaatkan oleh suatu wilayah. Penelitian ini akan membahas mengenai optimisasi analisis regresi LASSO dengan algoritma LARS menggunakan BIC untuk mengetahui faktor-faktor yang mempengaruhi kemiskinan di Jawa Timur tahun 2022.

1.2. Rumusan Masalah

Berdasarkan pada latar belakang yang telah diuraikan diperoleh rumusan masalah dari penelitian ini yaitu sebagai berikut:

1. Bagaimana langkah-langkah model regresi LASSO?

2. Bagaimana langkah-langkah optimisasi model regresi LASSO menggunakan BIC?
3. Bagaimana mengaplikasikan optimisasi model regresi LASSO dengan algoritma LARS menggunakan BIC pada data jumlah penduduk miskin menurut Kabupaten/Kota di Provinsi Jawa Timur pada tahun 2022?

1.3. Batasan Masalah

Dilakukan pembatasan masalah dalam penelitian ini agar fokus pada sasaran yang diharapkan, diantaranya:

1. Optimisasi model regresi LASSO yang digunakan adalah BIC.
2. Variabel yang digunakan yaitu data jumlah penduduk miskin dan faktor yang mempengaruhi kemiskinan yaitu indeks pembangunan manusia, garis kemiskinan, rata-rata lama sekolah, tingkat pengangguran terbuka, dan pendapatan per kapita menurut Kabupaten/Kota di Jawa Timur pada tahun 2022.
3. Algoritma yang digunakan yaitu algoritma LARS
4. *Software* yang digunakan adalah R versi R.4.4.1.

Tujuan Penelitian Tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengetahui langkah-langkah pemodelan regresi LASSO.
2. Untuk mengetahui langkah-langkah optimisasi model regresi LASSO menggunakan BIC.
3. Untuk mengaplikasikan optimisasi model regresi LASSO dengan algoritma LARS menggunakan BIC pada data jumlah penduduk miskin menurut Kabupaten/Kota di Provinsi Jawa Timur pada tahun 2022.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik dari segi teoritis maupun segi praktis. Secara teoritis, penelitian ini dapat menjadi tambahan acuan dalam penelitian dan menambah dinamika keilmuan dalam mengembangkan metode analisis dengan regresi *Least Absolute Shrinkage Selection and Operator* (LASSO). Secara praktik, penelitian ini akan menunjukkan penerapan analisis regresi LASSO pada data jumlah penduduk miskin dalam menangani kasus terkait kemiskinan terhadap faktor yang mempengaruhi. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya.

1.5. Sistematika Penulisan

Sistematika pada penelitian ini yaitu:

BAB 1 PENDAHULUAN

Pada bab ini menjelaskan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan, dan tinjauan pustaka.

BAB II DASAR TEORI

Pada bab ini menjelaskan landasan teori yang menjadi acuan, berisi tentang konsep, hipotesis dan rancangan penelitian.

BAB III METODE PENELITIAN

Pada bab ini menjelaskan 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 melakukan proses analisis data dan hasilnya akan diinterpretasikan secara terperinci.

BAB V PENUTUP

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

1.6. Tinjauan Pustaka

Penelitian yang dilakukan oleh (Sartika et al., 2020) menggunakan metode LASSO dengan menyusutkan koefisien regresi dari variabel bebas yang memiliki korelasi yang tinggi menjadi korelasi tepat nol atau mendekati nol. Koefisien LASSO dihitung menggunakan algoritma *Least Angle Regression* (LARS) karena lebih efisien. Penelitian tersebut menggunakan data kasus *pneumonia* pada balita di 34 puskesmas di Kota Pontianak dan Kabupaten Mempawah, terdapat dua variabel yang mengalami masalah multikolinearitas, yaitu variabel pemberian imunisasi DPT-Hb3-Hib3 (X2) dan pemberian imunisasi campak (X3) yang akan diatasi menggunakan metode LASSO. Kesimpulan yang diperoleh yaitu adanya perubahan koefisien regresi mendekati nol dan tepat nol, yaitu pada variabel X2 dari 0.285 berkurang menjadi 0.192 dan variabel X3 dari 0.024 menjadi 0. Hal ini menunjukkan masalah multikolinearitas yang terdeteksi dapat diatasi menggunakan metode LASSO.

Penelitian yang dilakukan oleh (Pardede et al., 2022) menggunakan metode regresi LASSO dengan algoritma LAR dikarenakan pada data *stunting* di Indonesia terdapat masalah multikolinearitas diantara variabel bebas. Metode LASSO digunakan untuk mengatasi masalah tersebut dengan mengecilkan koefisien regresi variabel bebas yang memiliki korelasi tinggi hingga mencapai nol. Kesimpulan yang diperoleh yaitu variabel ASI eksklusif, konsumsi protein, imunisasi DPT-HB, tinggi badan, dan diare memiliki pengaruh terhadap data *stunting* di Indonesia pada tahun 2018.

Penelitian yang dilakukan oleh (Robbani et al., 2019) bertujuan untuk meng-

identifikasi faktor-faktor yang mempengaruhi inflasi di Indonesia pada tahun 2014-2017 dengan menggunakan regresi LASSO dan algoritma LARS, yang dinilai efektif dalam menyelesaikan secara komputasi. Kesimpulan yang diperoleh yaitu variabel yang berpengaruh terhadap inflasi adalah Produk Domestik Bruto, Ekspor Bersih, Nilai Tukar Rupiah, Suku Bunga, Harga Beras, dan Upah Buruh Tani.

Penelitian yang dilakukan oleh (Mahalani & Rifai, 2022) membahas penerapan metode regresi LASSO dalam menganalisis data jumlah penduduk miskin di Indonesia pada tahun 2020. Hasil dari nilai koefisien korelasi dan nilai VIF yang diperoleh, terdapat tiga variabel yang mengalami masalah multikolinearitas.

Penelitian yang dilakukan oleh (Fitria & Rozci, 2023) bertujuan untuk membandingkan tingkat akurasi antara metode regresi LASSO dan regresi linear dalam memprediksi tingkat kemiskinan di Indonesia, serta menentukan model terbaik dari keduanya. Hasil dari penelitian ini menunjukkan bahwa, di antara kedua algoritma yang diterapkan menunjukkan tingkat akurasi lebih tinggi dalam memprediksi tingkat kemiskinan di setiap provinsi di Indonesia. Hal ini dapat dilihat dari nilai MSE yang lebih rendah dan nilai R^2 yang mendekati 1 jika dibandingkan dengan regresi LASSO. Selain itu, juga menunjukkan bahwa variabel pendidikan dan Indeks Pembangunan Manusia (IPM) memiliki pengaruh yang signifikan terhadap tingkat kemiskinan di provinsi-provinsi Indonesia.

Berdasarkan beberapa penelitian yang telah diuraikan sebelumnya, terlihat bahwa metode regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) dan algoritma *Least Angle Regression* (LARS) banyak digunakan dalam penelitian untuk mengatasi masalah multikolinearitas. Namun belum banyak penelitian sebelumnya yang menjelaskan mengenai pemilihan parameternya, salah satunya yaitu dengan BIC. Oleh karena itu, berdasarkan uraian topik dan kasus yang telah dija-

barkan, penulis tertarik untuk melakukan penelitian tentang metode regresi LASSO dengan algoritma LARS dan metode BIC dalam pemilihan parameternya, dengan judul Optimisasi Analisis Regresi LASSO dengan Algoritma LARS Menggunakan BIC.



BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian mengenai regresi LASSO dengan algoritma LARS pada data jumlah penduduk miskin di Jawa Timur tahun 2022, maka dapat diambil kesimpulan sebagai berikut:

1. Model regresi LASSO adalah metode regresi yang menggabungkan seleksi variabel dan regularisasi untuk menghasilkan model yang lebih akurat. LASSO membantu mengatasi masalah multikolinearitas, yaitu ketika dua atau lebih variabel yang saling berkorelasi satu sama lain dan LASSO dapat mengecilkan koefisien menjadi nol. Langkah-langkah membentuk model LASSO meliputi: mempersiapkan data dan normalisasi data, menentukan parameter penalti λ , lalu melakukan pelatihan model sehingga didapat model regresi LASSO.
2. Optimisasi model regresi LASSO menggunakan kriteria *Bayesian Information Criterion* (BIC) melibatkan pemilihan parameter α yang meminimalkan BIC. Langkah - langkahnya yaitu: mempersiapkan data, membagi data lalu menormalisasikan data, memilih rentang nilai lambda yang akan diuji, melakukan pelatihan model LASSO, lalu menghitung nilai BIC dan dipilih model optimal dengan nilai BIC terendah.
3. Berdasarkan analisis regresi LASSO data jumlah penduduk miskin di Jawa Timur tahun 2022 menggunakan algoritma LARS dengan kriteria BIC, diper-

oleh model terbaik sebagai berikut:

$$\hat{Y} = -0.5460815 - 0.6728233RLS + 0.1267418TPT \quad (5.1)$$

Model 5.1 menjelaskan bahwa terdapat dua dari lima variabel independen yang masuk ke dalam model terbaik yaitu *RLS* (rata-rata lama sekolah) dan *TPT* (tingkat pengangguran terbuka). Model tersebut menunjukkan bahwa semakin meningkat rata-rata lama sekolah, maka jumlah penduduk miskin akan menurun. Begitu juga dengan semakin meningkat tingkat pengangguran terbuka, maka jumlah penduduk miskin akan meningkat. Oleh karena itu, model yang menghasilkan nilai BIC optimal dapat dianggap sebagai model optimal untuk analisis data jumlah penduduk miskin di Jawa Timur tahun 2022.

5.2. Saran

Setelah penelitian dilakukan dan kesimpulan diperoleh mengenai regresi LASSO dengan algoritma LARS pada data jumlah penduduk miskin di Jawa Timur tahun 2022, penulis memberikan beberapa saran sebagai berikut:

1. Data yang digunakan pada penelitian ini yaitu data jumlah penduduk miskin di Jawa Timur tahun 2022 dengan lima variabel independen. Saran dan harapan penulis yaitu dapat menggunakan data yang lebih baru, penelitian dapat memberikan gambaran yang lebih akurat tentang faktor-faktor yang memengaruhi kemiskinan.
2. Pemilihan model terbaik pada penelitian ini menggunakan kriteria BIC sedangkan masih ada metode lain yang dapat digunakan untuk pemilihan model terbaik regresi LASSO.

DAFTAR PUSTAKA

- Badriyah, L. (2018). *Estimasi parameter model regresi poisson diperumum dengan metode maksimum likelihood*. PhD thesis, Universitas Islam Negeri Maulana Malik Ibrahim.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression.
- Fitria, E. R. & Rozci, F. (2023). 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.
- Gunes, F. (2015). Penalized regression methods for linear models in sas/stat®. In *Proceedings of the SAS Global Forum 2015 Conference*. Cary, NC: SAS Institute Inc. <http://support.sas.com/rnd/app/stat/papers/2015/PenalizedRegression-LinearModels.pdf>.
- Kurniawan, P. S. et al. (2018). Maximum likelihood estimator untuk mengestimasi model regresi isotonik dengan pendekatan polinomial bernstein pada kasus satu variabel independen.
- Mahalani, A. J. & Rifai, N. A. K. (2022). Least absolute shrinkage and selection operator (lasso) untuk mengatasi multikolinearitas pada model regresi linear berganda. In *Bandung Conference Series: Statistics*, volume 2, pages 119–125.
- Pardede, T. T., Sumargo, B., & Rahayu, W. (2022). Penerapan regresi least absolute shrinkage and selection operator (lasso) untuk mengidentifikasi variabel yang berpengaruh terhadap kejadian stunting di indonesia. *Jurnal Statistika dan Aplikasinya*, 6(1):37–48.

- Pekhimenko, G. (2006). Penalized logistic regression for classification. *Dept. Comput. Sci., Univ. Toronto, Toronto, ON M5S3L1*.
- Robbani, M., Agustiani, F., & Herrhyanto, N. (2019). Regresi least absolute shrinkage and selection operator (lasso) pada kasus inflasi di indonesia tahun 2014-2017. *Jurnal EurekaMatika*, 7(2):1–16.
- Sartika, I., Debataraja, N. N., & Imro'ah, N. (2020). Analisis regresi dengan metode least absolute shrinkage and selection operator (lasso) dalam mengatasi multikolinearitas. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 9(1).
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.
- Van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology*, 89:31–50.