


SKRIPSI

**OPTIMISASI ANALISIS REGRESI LASSO DENGAN
ALGORITMA LARS MENGGUNAKAN AKAIKE
INFORMATION CRITERION (AIC)**

**(STUDI KASUS : INDEKS PEMBANGUNAN MANUSIA DI
PROVINSI JAWA TENGAH TAHUN 2023)**



ZAINAB YUMNA SHALIHAH
20106010030

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

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

2025

**OPTIMISASI ANALISIS REGRESI LASSO DENGAN
ALGORITMA LARS MENGGUNAKAN AKAIKE
INFORMATION CRITERION (AIC)
(STUDI KASUS : INDEKS PEMBANGUNAN MANUSIA DI
PROVINSI JAWA TENGAH TAHUN 2023)**

Skripsi

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



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA
diajukan oleh
ZAINAB YUMNA SHALIHAH
20106010030

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

2025



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp : -

Kepada:

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga

di Yogyakarta

Assalaamu'alaikum wr. wb.

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

Nama : ZAINAB YUMNA SHALIHAH

NIM : 20106010030

Judul Skripsi : OPTIMISASI ANALISIS REGRESI LASSO DENGAN ALGORITMA LARS MENGGUNAKAN *AKAIKE INFORMATION CRITERION* (AIC).
(Studi Kasus: Indeks Pembangunan Manusia Provinsi Jawa Tengah Tahun 2023)

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, 19 Desember 2025

Pembimbing

Sri Utami Zuliana, S.Si., M.Sc., 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-232/Un.02/DST/PP.00.9/01/2026

Tugas Akhir dengan judul : Optimisasi Analisis Regresi Lasso dengan Algoritma LARS menggunakan Akaike Information Criterion (AIC) (Studi Kasus: Indeks Pembangunan Manusia di Provinsi Jawa Tengah tahun 2023)

yang dipersiapkan dan disusun oleh:

Nama : ZAINAB YUMNA SHALIAH
Nomor Induk Mahasiswa : 20106010030
Telah diujikan pada : Kamis, 08 Januari 2026
Nilai ujian Tugas Akhir : A/B

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: 697b2c3157c59



Penguji I
Arya Fendha Ibnu Shina, M.Si.
SIGNED

Valid ID: 697b1ae359e26



Penguji II
Arif Munandar, M.Sc.
SIGNED

Valid ID: 697b158600f5



Yogyakarta, 08 Januari 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 697c2172d8312

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Zainab Yumna Shalihah
NIM : 20106010030
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, 22 Desember 2025



Zainab Yumna Shalihah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

KARYA SEDERHANA INI PENULIS PERSEMBAHKAN UNTUK:

KEDUA ORANG TUA PENULIS BAPAK NGADINA DAN IBU IKA WULANDARI,

ADIK-ADIK PENULIS ASMA' FAUZIYYAH DAN AHMAD MUJTABA,

KELUARGA BESAR PENULIS YANG TELAH MEMBERIKAN DO'A, KASIH

DAN CINTA YANG TAK TERBATAS.

DIRI SENDIRI YANG SUDAH SELALU MENGUSAHAKAN APAPUN.

SAHABAT TERCINTA DAN SEMUA ORANG YANG PERNAH PENULIS KENAL.

ALMAMATER KAMPUS TERCINTA DAN SEMUA YANG ADA DI DALAMNYA

YANG BANYAK MEMBERIKAN WARNA DALAM HIDUP SELAMA MASA

PERKULIAHAN

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA YOGYAKARTA.

"MATEMATIKA ANGGARAN 2020 TERCINTA"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

”Apa yang kamu lakukan saat ini adalah apa yang akan kamu dapatkan dimasa yang akan datang.”

”Tidak ada yang kita taklukan, kecuali diri sendiri. Tidak ada yang kita kalahkan, kecuali ego sendiri.”

Fiersa Besari



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PRAKATA

Puji syukur kehadirat Allah SWT atas segala limpahan rahmat, nikmat, dan karunia-Nya yang tak ternilai harganya berupa keimanan, kesabaran, kekuatan dan kelancaran. Tak lupa shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul "Optimisasi Analisis Regresi LASSO dengan Algoritma LARS Menggunakan *Akaike Information Criterion* (AIC): Studi Kasus Indeks Pembangunan Manusia Provinsi Jawa Tengah Tahun 2023".

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga. Penelitian ini bertujuan untuk mengkaji penerapan metode Regresi LASSO secara optimal melalui algoritma *Least Angle Criterion* (LARS), serta mengevaluasi kriteria pemilihan model menggunakan *Akaike Information Criterion* (AIC), dengan studi kasus pada data Indeks Pembangunan Manusia di Provinsi Jawa Tengah.

Mengenai skripsi ini, penulis menyadari bahwa keberhasilan dan kelancaran penelitian ini tidak lepas dari dukungan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Noorhaidi, 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.

3. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
4. Bapak Arif Munandar, M.Sc., selaku dosen pembimbing akademik yang telah memberikan pengarahan kepada penulis dari awal perkuliahan hingga akhir.
5. Ibu Sri Utami Zuliana, S.Si., M.Sc., Ph.D., selaku pembimbing skripsi penulis atas semua bimbingan dan arahan yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini.
6. 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.
7. Orang tua tercinta, Bapak Ngadina dan Ibu Ika Wulandari, yang menjadi alasan penulis untuk terus berjuang. Terima kasih atas doa yang tidak pernah terputus, cinta yang tidak pernah berkurang, dukungan yang tidak pernah tergantikan di setiap langkah, dan pengorbanan yang tidak akan pernah bisa penulis balas. Segala pencapaian ini juga milik kalian.
8. Kedua adik tercinta Asma' dan Ahmad, dan keluarga besar penulis yang selalu memberikan kasih sayang, nasehat dan juga doa yang tidak pernah terputus sehingga penulis bisa berada dititik ini.
9. Seseorang yang tidak kalah penting, sudah menjadi pendengar yang baik dan selalu memberikan *support* dan juga motivasi untuk segera menyelesaikan skripsi ini. Mari mengusahakan apa yang seharusnya kita usahakan selanjutnya.

10. Sahabat dekat penulis selama dibangku perkuliahan Riri, Alfiah, dan Reni terima kasih sudah menjadi teman berbagi keluh kesah, tawa, dan segala cerita, tanpa kalian bangku perkuliahan tidak akan seberwarna ini. Tetap jaga silaturahmi apapun keadaannya ya!
11. Keluarga Matematika angkatan 2020 yang telah kebersamai selama dibangku perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.
12. Teman-teman seperbimbingan Linggar, Sari, Adel, Uswa, Anisa, Fatma yang telah berbagi ilmu.
13. Teman-teman KKN 111 Desa Surengede yang telah memberikan pengalaman hidup nyata di luar kampus, terima kasih untuk kerja sama, tawa, dan pembelajaran yang begitu berharga.
14. Rekan dan Rekanita Pengurus PC IPNU-IPPNU Kabupaten Klaten, yang menjadi rumah organisasi tempat penulis bertumbuh, belajar kepemimpinan, tanggungjawab dan berjuang bersama. Khususnya, Mas Hanif yang sudah menjadi partner ketua selama satu periode.
15. Zainab Yumna Shalihah, yang sudah berusaha kuat meski sering lelah, tetap berdiri ketika ingin menyerah, dan terus mencoba meski banyak hal terasa berat. Terima kasih sudah sampai dititik ini dan selalu mengusahakan apapun karena tidak ada yang bisa diandalkan selain diri sendiri. *Finally you did it!*.
16. Terakhir untuk semua pihak yang telah berkontribusi dalam perjalanan penulis khususnya selama dibangku perkuliahan yang tidak bisa disebutkan satu-persatu, semoga selalu dimudahkan segala urusannya.

Penulis sangat menyadari bahwa tidak ada yang sempurna di dunia ini. Oleh karena itu, segala bentuk saran, kritik yang membangun 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*.

Wallahul muwaffiq ilaa aqwamith thariq

Yogyakarta, 12 Desember 2025

Zainab Yumna Shalihah



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMBANG	xvi
INTISARI	xvii
ABSTRACTxviii
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah	5
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	6
1.6. Sistematika Penulisan	6
1.7. Tinjauan Pustaka	7
II DASAR TEORI	11
2.1. Analisis Regresi	11
2.1.1. Analisis Regresi Linier Sederhana	12
2.1.2. Analisis Regresi Linier Berganda	12
2.2. Uji Hipotesis dalam Regresi Linier	13

2.2.1. Uji F (Uji Serentak)	14
2.2.2. Uji t (Uji Parsial)	14
2.3. Uji Asumsi Klasik Regresi Linier	14
2.3.1. Uji Normalitas	15
2.3.2. Uji Autokorelasi	15
2.3.3. Uji Heteroskedastisitas	16
2.3.4. Uji Multikolinearitas	16
2.4. <i>Maximum Likelihood Estimation</i> (MLE)	17
2.5. Multikolinearitas	17
2.6. Standarisasi Data	18
2.7. Regresi Terpenalti	19
2.8. <i>Least Absolute Shrinkage and Selection Operator</i> (LASSO)	19
2.9. Algoritma <i>Least Angle Regression</i> (LARS)	20
2.10. <i>Akaike Information Criterion</i>	23
III METODE PENELITIAN	25
3.1. Metode Penelitian	25
3.2. Jenis dan Sumber Literatur	25
3.3. Studi Kasus	26
3.4. Software yang Digunakan	27
3.5. Langkah-langkah Analisis	27
3.6. Diagram Analisis Data (<i>Flowchart</i>)	28
IV HASIL DAN PEMBAHASAN	29
4.1. Statistik Deskriptif	29
4.2. Standarisasi Data	32
4.3. Analisis Regresi Linier	33
4.4. Uji Hipotesis	35

4.4.1. Uji F (Uji Serentak)	35
4.4.2. Uji t (Uji Parsial)	36
4.5. Uji Asumsi Klasik	37
4.5.1. Uji Normalitas	38
4.5.2. Uji Autokorelasi	38
4.5.3. Uji Heteroskedastisitas	40
4.5.4. Uji Multikolinearitas	40
4.6. <i>Least Absolute Shrinkage and Selection Operator Operator (LASSO)</i> dengan Algoritma LARS	42
4.7. Menentukan Model LASSO	44
4.8. Menentukan Nilai λ Optimal dengan AIC	46
4.9. Pemilihan Model Terbaik Regresi LASSO	48
V PENUTUP	51
5.1. Kesimpulan	51
5.2. Saran	52
DAFTAR PUSTAKA	54
LAMPIRAN	57
A Data Penelitian	57
B Source Code Program R	61
C Tabel Tinjauan Pustaka	67

DAFTAR TABEL

1.1	Tabel Tinjauan Pustaka	10
3.1	Variabel Data Penelitian	26
4.1	Statistik Deskriptif	30
4.2	Nilai <i>p-value</i> Setiap Variabel	36
4.3	Uji Normalitas	38
4.4	Output Uji Autokorelasi	39
4.5	Output Uji Heteroskedastisitas	40
4.6	Nilai VIF	41
4.7	Tahapan Masuknya Variabel Bebas pada Regresi LASSO	43
4.8	Nilai Lambda	45
4.9	Nilai Lambda dan Nilai AIC	47
4.10	Model Regresi LASSO dengan AIC (λ)	48
4.11	Koefisien Regresi LASSO Berdasarkan AIC	49

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

3.1	Flowchart Penelitian	28
4.1	Boxplot Data	32
4.2	Boxplot Standarisasi Data	33
4.3	Plot Autokorelasi	39
4.4	Ilustrasi Plot Tahapan Variabel Bebas kedalam Model	42
4.5	Plot Nilai AIC terhadap Lambda	48



DAFTAR LAMBANG

Y	= Variabel terikat
X_n	= Variabel bebas
β_0	= Konstanta
β_1, \dots, β_n	= Koefisien regresi
ϵ	= <i>Error</i>
n	= Banyaknya observasi
R_j^2	= koefisien determinasi
$VIF(j)$	= Faktor perubahan variansi dalam variabel bebas ke-j
L_1	= Regularisasi LASSO
L_2	= Regularisasi ridge
λ	= Parameter penyusutan
$\hat{\beta}^{lasso}$	= koefisien regresi LASSO
p	= Banyaknya variabel bebas dalam model
x_{ij}	= Variabel bebas ke-j pada pengamatan ke-i
t	= Nilai parameter tuning dengan $t \geq 0$
N	= Banyaknya pengamatan
L	= Fungsi likelihood
k	= Jumlah parameter bebas yang diestimasi

INTISARI

Optimisasi Analisis Regresi LASSO dengan Algoritma LARS Menggunakan

Akaike Information Criterion (AIC)

(Studi kasus : Indeks Pembangunan Manusia di Provinsi Jawa Tengah Tahun 2023)

Zainab Yumna Shalihah

20106010030

Multikolinearitas merupakan permasalahan umum dalam regresi linier berganda akibat adanya korelasi tinggi antar variabel independen, yang dapat menyebabkan ketidakstabilan estimasi parameter, meningkatnya variansi koefisien, serta menurunnya signifikansi statistik model. Penelitian ini bertujuan mengatasi permasalahan multikolinearitas melalui penerapan metode *Least Absolute Shrinkage and Selection Operator (LASSO)* dengan algoritma *Least Angle Regression (LARS)* serta pemilihan parameter penalti optimal menggunakan *Akaike Information Criterion (AIC)*. Proses analisis diawali dengan standarisasi data dan identifikasi multikolinearitas menggunakan *Variance Inflation Factor (VIF)*, kemudian dilanjutkan dengan estimasi regresi LASSO berbasis LARS. Nilai parameter penalti (λ) ditentukan berdasarkan nilai AIC minimum untuk memperoleh keseimbangan optimal antara ketepatan model dan kompleksitas parameter. Metode ini diaplikasikan pada data Indeks Pembangunan Manusia Provinsi Jawa Tengah tahun 2023 menurut kabupaten/kota. Hasil penelitian menunjukkan bahwa regresi LASSO mampu mereduksi dampak multikolinearitas melalui seleksi variabel yang efektif, sehingga menghasilkan model yang lebih sederhana, stabil, dan mudah diinterpretasikan dibandingkan regresi linier konvensional.

Kata kunci: multikolinearitas, regresi LASSO, algoritma LARS, Akaike Information Criterion, Indeks Pembangunan Manusia.

ABSTRACT

Optimization of LASSO Regression Analysis with LARS Algorithm Using

Akaike Information Criterion (AIC)

(Case study : Human Development Index of Central Java Province 2023)

Zainab Yumna Shalihah

20106010030

Multicollinearity is a common problem in multiple linear regression due to the high correlation between independent variables, which can cause instability in parameter estimates, increase coefficient variance, and decrease the statistical significance of the model. This study aims to address the problem of multicollinearity by applying the Least Absolute Shrinkage and Selection Operator (LASSO) method with the Least Angle Regression (LARS) algorithm and selecting the optimal penalty parameter using the Akaike Information Criterion (AIC). The analysis process begins with data standardization and multicollinearity identification using the Variance Inflation Factor (VIF), followed by LARS-based LASSO regression estimation. The penalty parameter (λ) value is determined based on the minimum AIC value to achieve an optimal balance between model accuracy and parameter complexity. This method is applied to the 2023 Human Development Index data of Central Java Province by district/city. The results of the study show that LASSO regression is able to reduce the impact of multicollinearity through effective variable selection, resulting in a simpler, more stable, and easier to interpret model than conventional linear regression.

Keywords: multicollinearity, LASSO regression, LARS algorithm, Akaike Information Criterion, Human Development Index.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk mengetahui dan memodelkan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Tujuannya adalah untuk mengetahui pengaruh dari variabel bebas terhadap variabel terikat, serta melakukan prediksi terhadap variabel dependen berdasarkan nilai-nilai variabel independen tersebut. Dalam praktiknya, regresi linear berganda digunakan untuk model yang melibatkan lebih dari satu prediktor. Namun, ketika jumlah variabel meningkat atau terjadi korelasi tinggi antar prediktor (multikolinearitas), keakuratan dan stabilitas model regresi konvensional menjadi menurun.

Salah satu pendekatan yang sering digunakan untuk mengestimasi parameter dalam model regresi adalah metode *Maximum Likelihood Estimation* (MLE). Metode ini bekerja dengan memaksimalkan fungsi likelihood untuk mendapatkan nilai parameter yang paling mungkin menghasilkan data yang diamati. Estimasi MLE memiliki sifat konsisten dan efisien jika asumsi-asumsi klasik regresi terpenuhi, seperti normalitas residual, tidak terdapat autokorelasi, homoskedastisitas, dan tidak ada multikolinearitas. Namun, ketika multikolinearitas terjadi, MLE menghasilkan estimasi parameter yang tidak stabil dan berdampak pada kualitas model regresi. Cara mengatasi permasalahan multikolinearitas ini dengan menggunakan metode regresi terpenalti. (Purba, 2021).

Regresi terpenalti merupakan salah satu pendekatan dalam analisis regresi yang menerapkan penalti terhadap besarnya koefisien regresi pada setiap variabel dalam model. Pemberian penalti ini bertujuan untuk mengurangi risiko terjadinya *overfitting*, yaitu kondisi ketika model menjadi terlalu kompleks sehingga kurang mampu melakukan generalisasi. Melalui mekanisme penalti, kontribusi variabel dengan pengaruh relatif kecil akan ditekan hingga mendekati nol, bahkan dapat dieliminasi dari model. (Pekhimenko, 2006) Beberapa metode regresi terpenalti yang umum digunakan antara lain regresi Ridge, regresi *Least Absolute Shrinkage and Selection Operator* (LASSO), dan Elastic Net.

Regresi LASSO bekerja dengan menambahkan penalti berupa nilai absolut koefisien regresi ke dalam fungsi objektif. Karakteristik utama dari LASSO adalah kemampuannya untuk menyusutkan beberapa koefisien regresi hingga bernilai nol. Dengan demikian, LASSO tidak hanya berfungsi sebagai metode regularisasi untuk mengatasi multikolinearitas, tetapi juga sebagai metode seleksi variabel yang efektif. Keunggulan ini menjadikan LASSO sangat relevan untuk analisis data dengan jumlah prediktor yang relatif banyak dan saling berkorelasi, karena mampu menghasilkan model yang lebih sederhana, stabil, dan mudah diinterpretasikan. (Tibshirani, 1996).

Dalam penelitian ini, regresi LASSO diestimasi secara efisien dengan bantuan algoritma *Least Angle Regression* (LARS). Algoritma LARS dikembangkan oleh Efron et al. (2004) sebagai metode efisien dalam mengestimasi koefisien regresi dengan pendekatan seleksi variabel bertahap. Algoritma ini bekerja dengan menambahkan variabel independen satu per satu berdasarkan korelasi tertinggi terhadap residual sehingga korelasi terhadap galat akan menurun. Keunggulan utama algoritma LARS adalah kemampuannya menghasilkan jalur solusi LASSO hanya

dalam satu proses komputasi. Efisiensi LARS menjadikannya sangat sesuai untuk regresi dengan dimensi data tinggi dan kompleksitas tinggi.

Salah satu aspek penting dalam regresi LASSO adalah penentuan nilai parameter penalti (λ), karena nilai λ mengontrol tingkat penyusutan koefisien regresi pada variabel independen sekaligus kompleksitas model. Semakin besar nilai λ , semakin kuat penalti yang diberikan, sehingga lebih banyak koefisien yang disusutkan menuju nol. Sebaliknya, nilai λ yang terlalu kecil dapat menghasilkan model yang kompleks dan berpotensi mengalami *overfitting*. Oleh karena itu, pemilihan nilai λ yang optimal menjadi tahap krusial dalam pemodelan regresi LASSO.

Terdapat beberapa metode yang umum digunakan untuk menentukan nilai parameter penalti, antara lain *Cross-Validation (CV)*, *Generalized Cross Validation (GCV)*, *Bayesian Information Criterion (BIC)*, dan *Akaike Information Criterion (AIC)*. Dalam penelitian ini dipilih *Akaike Information Criterion (AIC)* sebagai dasar pemilihan nilai parameter penalti (λ) optimal. Tujuan utama AIC adalah meminimalkan kesalahan prediksi model dengan mempertimbangkan kualitas model dan kompleksitasnya. AIC mengevaluasi model berdasarkan keseimbangan antara kecocokan model dan kompleksitas model, di mana model dengan nilai AIC paling kecil dianggap sebagai model terbaik. AIC menjadi sangat penting dalam pemodelan LASSO karena membantu menentukan nilai λ yang memberikan prediksi terbaik tanpa *overfitting*. (Fathurahman, 2010)

Sebagai bentuk penerapan metode, penelitian ini menggunakan Indeks Pembangunan Manusia (IPM) sebagai studi kasus. IPM merupakan indikator penting untuk mengukur kualitas pembangunan manusia dari tiga dimensi utama: umur panjang dan hidup sehat, pengetahuan, dan standar hidup layak. Keterkaitan yang kuat antarindikator tersebut menyebabkan potensi multikolinearitas dalam pemodelan.

delan regresi, sehingga menjadikan IPM sebagai studi kasus yang relevan untuk menguji efektifitas regresi LASSO.

Sejalan dengan karakteristik tersebut, data Indeks Pembangunan Manusia (IPM) memenuhi kriteria yang sesuai untuk penerapan regresi LASSO karena terdiri atas beberapa variabel prediktor yang berskala kontinu dan saling berkorelasi. Kondisi ini berpotensi menimbulkan multikolinearitas apabila dianalisis menggunakan regresi linier konvensional, sehingga diperlukan metode yang mampu melakukan penyusutan koefisien sekaligus seleksi variabel. Oleh karena itu, IPM menjadi studi kasus yang relevan untuk menguji kemampuan regresi LASSO dalam menghasilkan model yang lebih sederhana, stabil, dan mudah diinterpretasikan melalui proses penyusutan koefisien dan seleksi variabel secara simultan.

Pemilihan IPM Provinsi Jawa Tengah tahun 2023 didasarkan pada adanya variasi capaian nilai IPM antar kabupaten/kota menunjukkan variasi yang cukup besar. Variasi ini mencerminkan adanya perbedaan dalam akses terhadap layanan dasar dan kesejahteraan sosial masyarakat. Dalam penelitian ini, IPM digunakan sebagai studi kasus untuk menerapkan metode analisis regresi LASSO menggunakan algoritma LARS dan pemilihan model berdasarkan *Akaike Information Criterion* (AIC). Fokus utama dari penelitian ini adalah pada metode analisis dan pendekatan statistik yang digunakan, sedangkan IPM hanya digunakan sebagai data penerapan.

Dengan menerapkan kombinasi metode LASSO, algoritma LARS, dan kriteria AIC, diharapkan dapat diperoleh model regresi yang optimal, efisien, dan akurat dalam kondisi multikolinearitas. Penelitian ini tidak hanya berkontribusi dalam pengembangan metode statistik, tetapi juga membuka ruang aplikasi yang luas dalam bidang sosial ekonomi dan kebijakan publik.

1.2. Rumusan Masalah

Berdasarkan pemaparan yang telah disampaikan pada bagian latar belakang, maka permasalahan yang menjadi fokus dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana langkah-langkah analisis regresi LASSO dengan algoritma LARS?
2. Bagaimana penentuan parameter penalti optimal pada regresi LASSO menggunakan kriteria *Akaike Information Criterion* (AIC)?
3. Bagaimana penerapan analisis regresi LASSO dengan algoritma LARS menggunakan *Akaike Information Criterion* (AIC) pada data indeks pembangunan manusia di Provinsi Jawa Tengah menurut kabupaten/kota tahun 2023?

1.3. Batasan Masalah

Agar penelitian ini tetap terarah dan sesuai dengan tujuan yang telah ditetapkan, maka ruang lingkup pembahasan perlu dibatasi melalui sejumlah batasan masalah sebagai berikut:

1. Penelitian ini menggunakan data Indeks Pembangunan Manusia Provinsi Jawa Tengah tahun 2023.
2. Metodologi yang digunakan yaitu regresi LASSO dengan algoritma LARS.
3. Pemilihan model terbaik menggunakan *Akaike Information Criterion* (AIC).
4. *Software* yang digunakan adalah R versi R.4.4.2.

1.4. Tujuan Penelitian

1. Mengetahui langkah-langkah analisis regresi LASSO dengan algoritma LARS.

2. Menentukan parameter penalti optimal pada regresi LASSO berdasarkan kriteria AIC.
3. Mengaplikasikan analisis regresi LASSO dengan algoritma LARS menggunakan AIC pada data indeks pembangunan manusia Provinsi Jawa Tengah menurut kabupaten/kota tahun 2023.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis. Secara teoritis, penelitian ini dapat memperkaya kajian statistik dengan menggabungkan regresi LASSO, algoritma LARS dan kriteria AIC untuk membentuk model regresi yang efisien. Secara praktis, metode ini membantu mengatasi masalah multikolinearitas dan seleksi variabel, khususnya dalam menganalisis Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah tahun 2023. Hasilnya diharapkan dapat menjadi acuan bagi pengambilan kebijakan serta referensi akademis bagi penelitian selanjutnya.

1.6. Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini memuat uraian mengenai latar belakang penelitian, perumusan masalah, pembatasan masalah, tujuan dan manfaat penelitian, kajian pustaka yang relevan, serta sistematika penulisan laporan penelitian.

BAB II DASAR TEORI

Bab ini memuat uraian mengenai teori yang relevan dengan penelitian seperti regresi LASSO, algoritma LARS, *Akaike Information Criterion* (AIC), serta studi terdahulu yang mendukung pembahasan dalam penelitian ini.

BAB III METODE PENELITIAN

Bab ini menguraikan metode penelitian yang digunakan untuk menyelesaikan permasalahan, meliputi pendekatan penelitian, sumber data, teknik pengolahan data, serta diagram alur analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan tahapan analisis data yang dilakukan, disertai dengan pemaparan dan interpretasi hasil secara rinci dan sistematis..

BAB V PENUTUP

Bab ini memuat simpulan yang diperoleh dari hasil penelitian serta saran yang diajukan penulis sebagai bahan pertimbangan untuk penelitian selanjutnya.

1.7. Tinjauan Pustaka

Dalam penelitian ini, fokus utama adalah pada optimisasi analisis regresi LASSO dengan algoritma LARS menggunakan *Akaike Information criterion (AIC)* dalam konteks Indeks Pembangunan Manusia (IPM). Sejumlah penelitian terdahulu telah dilakukan dengan pendekatan yang serupa dan dapat memberikan gambaran tentang perkembangan metode ini serta aplikasinya dalam konteks yang relevan.

Salah satu studi yang relevan adalah penelitian oleh (Fitria & Fatchur, 2022) yang berjudul Penerapan Metode Regresi *Least Absolute Shrinkage Selection Operator (LASSO)* dan Regresi Linier untuk Memprediksi Tingkat Kemiskinan di Indonesia. Penelitian ini berisi terkait membandingkan regresi LASSO dengan regresi linier dalam memprediksi tingkat kemiskinan di Indonesia. Hasil penelitian menunjukkan bahwa regresi linier memberikan akurasi lebih tinggi dibandingkan LASSO, ditunjukkan oleh nilai *Mean Squared Error (MSE)* yang lebih kecil serta nilai koefisien determinasi (R^2) yang lebih tinggi. Namun demikian, regresi LASSO tetap efektif dalam mengidentifikasi variabel signifikan, di mana IPM dan pendidikan

menjadi faktor dominan. Penelitian ini relevan karena menggunakan regresi LASSO dan melibatkan IPM, meskipun fokus utamanya adalah pada kemiskinan, bukan IPM secara langsung.

Selanjutnya, penelitian oleh (Mahalani & Rifai, 2022) dengan judul *Least Absolute Shrinkage Selection Operator (LASSO) untuk Mengatasi Multikolinearitas pada Model Regresi Linier Berganda*. Penelitian ini membahas penerapan regresi LASSO untuk mengatasi multikolinearitas dalam model regresi linier berganda. Hasilnya menunjukkan bahwa LASSO mampu menyederhanakan model dengan menyeleksi variabel yang paling berpengaruh dan mengeliminasi variabel dengan korelasi tinggi. Hal ini menunjukkan potensi LASSO dalam meningkatkan efisiensi model prediktif. Penelitian ini selaras dengan tujuan studi ini, terutama dalam konteks penanganan multikolinearitas, meskipun belum mengintegrasikan AIC atau algoritma LARS.

Penelitian serupa juga dilakukan oleh (Sartika et al., 2020) dengan judul *Analisis Regresi dengan Metode Least Absolute Shrinkage Selection Operator (LASSO) dalam Mengatasi Multikolinearitas*. Pada penelitian ini menekankan bahwa regresi LASSO efektif dalam menyusutkan koefisien variabel yang berkorelasi tinggi menjadi nol. Dengan demikian, LASSO dinilai mampu mengatasi multikolinearitas yang sering menjadi kendala dalam regresi linier. Meski tidak menggunakan AIC dalam pemilihan model, penelitian ini memberikan dasar kuat mengenai keunggulan LASSO dalam konteks penyederhanaan model yang kompleks.

Selanjutnya, penelitian yang dilakukan oleh (Robbani et al., 2019) dengan judul *Regresi Least Absolute Shrinkage Selection Operator (LASSO) pada Kasus Inflasi di Indonesia Tahun 2014-2017*. Pada penelitian ini menerapkan regresi LASSO untuk menganalisis faktor-faktor yang memengaruhi inflasi di Indonesia. Mere-

ka menunjukkan bahwa LASSO tidak hanya dapat menyederhanakan model dengan mengurangi jumlah variabel, tetapi juga dapat menangani multikolinearitas dengan baik. Penelitian ini memiliki kesamaan dalam penggunaan LASSO untuk pemodelan regresi dan pemilihan variabel, walaupun berbeda dari segi topik dan tidak mengintegrasikan AIC ataupun LARS.

Kemudian, penelitian yang dilakukan oleh (Pardede et al., 2022) dengan judul Penerapan Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk Mengidentifikasi Variabel yang Mempengaruhi Kejadian *Stunting* di Indonesia bertujuan menganalisis data *stunting* dan menarik kesimpulan mengenai hubungan antar variabel. Penelitian ini menyoroti asumsi penting dalam analisis regresi linear, seperti distribusi normal data dan ketiadaan korelasi dengan kesalahan. Salah satu permasalahan utama yang dihadapi adalah adanya multikolinearitas di antara variabel independen. Untuk mengatasi hal tersebut, digunakan metode LASSO dengan algoritma LAR. Hasil analisis menunjukkan bahwa variabel ASI eksklusif, asupan protein, imunisasi DPT-HB, tinggi badan, serta kejadian diare memiliki pengaruh yang signifikan terhadap kejadian *stunting* di Indonesia pada tahun 2018.

Dari beberapa studi tersebut, dapat disimpulkan bahwa regresi LASSO telah banyak digunakan dalam berbagai studi sosial ekonomi, terutama dalam menangani multikolinearitas dan seleksi variabel. Namun, belum banyak penelitian yang secara eksplisit menggabungkan regresi LASSO dengan algoritma LARS dan kriteria AIC dalam konteks prediksi atau analisis IPM. Oleh karena itu, penelitian ini akan memberikan kontribusi dengan mengintegrasikan ketiga pendekatan tersebut dalam satu kerangka analisis yang lebih optimal dan komprehensif.

Tabel 1.1 Tabel Tinjauan Pustaka

NO	JUDUL PENELITIAN	PENELITI	PERBEDAAN
1	Penerapan Metode Regresi Least Absolute Shrinkage and Selection Operator (LASSO) dan Regresi Linier untuk Memprediksi Tingkat Kemiskinan di Indonesia (Fitria & Fatchur, 2022)	Eka Rossalina Fitria, Fatchur Rozci	Perbandingan antara tingkat akurasi metode LASSO dengan regresi linear dalam memprediksi tingkat kemiskinan dan memilih model terbaik dari kedua metode tersebut, menggunakan data kemiskinan untuk data penelitian dan metode penentuan koefisien regresi LASSO hanya menggunakan algoritma LARS.
2	Least Absolute Shrinkage and Selection Operator (LASSO) untuk Mengatasi Multikolinearitas pada Model Regresi Linier Berganda. (Mahalani & Rifai, 2022).	Annisa Juwita Mahalani, Nur Azizah Komara Rifai	Data penelitian menggunakan data penduduk miskin di Indonesia pada tahun 2020. Dan metode penentuan koefisien regresi LASSO hanya menggunakan algoritma LARS.
3	Analisis Regresi dengan Metode Least Absolute Shrinkage Operation (LASSO) dalam Mengatasi Multikolinearitas (Sartika et al., 2020).	Imi Sartika, Naomi Nessyana Debataraja, Nurfitri Imro'ah	Pada penelitian ini data yang digunakan adalah data kasus <i>pneumonia</i> pada balita di Kota Pontianak dan Kabupaten Mempawah. Penentuan koefisien hanya menggunakan algoritma LARS.
4	Regresi Least Absolute Shrinkage and Selection Operator (LASSO) Pada Kasus Inflasi di Indonesia Tahun 2014-2017 (Robbani et al., 2019).	Muhammad Robbani, Fitriani Agustiani, Nar Herrhyanto	Data penelitian menggunakan kasus inflasi di Indonesia tahun 2014-2017 dan metode penentuan koefisien model hanya menggunakan algoritma LARS.
5	Penerapan Regresi Least Absolute shrinkage and Selection Operator (LASSO) untuk Mengidentifikasi Variabel yang Berpengaruh terhadap kejadian <i>Stunting</i> di Indonesia (Pardede et al., 2022).	Tesa Trilonika Pardede, Bagus Sumargo, Widyanti Rahayu	Data penelitian menggunakan data <i>stunting</i> di Indonesia dan analisis yang digunakan yaitu metode regresi LASSO dengan algoritma LAR.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil analisis serta pembahasan yang telah dilakukan, dapat disimpulkan bahwa regresi LASSO dengan algoritma *Least Angle Regression* (LARS) menggunakan *Akaike Information Criterion* (AIC) mampu mengoptimalkan proses analisis regresi pada data Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah tahun 2023. Langkah-langkah analisis dilakukan secara sistematis, dimulai dari standarisasi data, identifikasi masalah multikolinearitas pada regresi linier berganda, penerapan regresi LASSO dengan algoritma LARS, serta pemilihan nilai parameter penalti λ optimal berdasarkan nilai AIC minimum. Penerapan metode ini terbukti efektif dalam mengatasi masalah multikolinearitas melalui mekanisme penalti L_1 yang mampu menyusutkan koefisien regresi dan melakukan seleksi variabel secara otomatis, sehingga menghasilkan model yang lebih sederhana dan stabil.

Penentuan parameter penalti optimal (λ) pada regresi LASSO dilakukan menggunakan kriteria *Akaike Information Criterion* (AIC) dengan mengevaluasi nilai AIC pada setiap kandidat λ yang dihasilkan. Nilai λ optimal diperoleh pada kondisi ketika AIC mencapai nilai minimum yaitu sebesar 0,01214 dengan nilai λ sebesar 0,01446, karena pada titik tersebut tercapai keseimbangan terbaik antara ketepatan model dan kompleksitas parameter. Model regresi LASSO dengan λ optimal menghasilkan struktur model yang lebih sederhana, stabil, dan mampu mengurangi dampak multikolinearitas melalui mekanisme penyusutan dan seleksi

variabel secara simultan.

Berdasarkan analisis regresi LASSO data indeks pembangunan manusia Provinsi Jawa Tengah tahun 2023 menggunakan algoritma LARS dengan kriteria AIC, mendapatkan model terbaik sebagai berikut:

$$\hat{Y} = 3,1112 \times 10^{-16} + 2,0024 \times 10^{-01} X_1 + 1,7862 \times 10^{-1} X_2 + 3,8607 \times 10^{-1} X_3 + 3,2125 \times 10^{-1} X_4 + 2,2914 \times 10^{-2} X_6 + 6,7526 \times 10^{-03} X_9 \quad (5.1)$$

Berdasarkan Persamaan 5.1 menjelaskan bahwasanya terdapat enam dari sepuluh variabel independen yang masuk ke dalam model terbaik yaitu umur harapan hidup saat lahir (X_1), harapan lama sekolah (X_2), rata-rata lama sekolah (X_3), pengeluaran per kapita (X_4), tingkat pengangguran terbuka (X_6), dan jumlah penduduk (X_9). Model terbaik ditentukan berdasarkan kombinasi variabel yang menghasilkan nilai lambda dan AIC yang paling optimal. Oleh karena itu, model dengan nilai λ dan AIC optimal dapat dinyatakan sebagai model yang sesuai untuk analisis data indeks pembangunan manusia Provinsi Jawa Tengah tahun 2023.

5.2. Saran

Berdasarkan hasil penelitian yang telah diperoleh, penelitian selanjutnya disarankan untuk menggunakan data lintas waktu, baik *time series* maupun data panel, agar dinamika Indeks Pembangunan Manusia (IPM) dapat dianalisis secara lebih mendalam serta memungkinkan evaluasi kestabilan model regresi dalam jangka panjang. Selain itu, metode seleksi model dapat dikembangkan dengan membandingkan *Akaike Information Criterion* (AIC) dengan kriteria lain, seperti *Bayesian Information Criterion* (BIC), *Cross Validation*, atau *Generalized Cross Validation* (GCV), sehingga diperoleh gambaran yang lebih komprehensif mengenai kinerja berbagai pendekatan dalam menentukan parameter penalti optimal. Pengembang-

an penelitian juga dapat dilakukan dengan membandingkan regresi LASSO dengan metode regularisasi lain, seperti Ridge Regression dan Elastic Net, untuk menilai keunggulan relatif masing-masing metode dalam menangani multikolinearitas dan kompleksitas model. Selanjutnya, penerapan metode regresi LASSO dengan algoritma LARS dan kriteria AIC dapat diperluas pada bidang kajian lain, khususnya data berdimensi tinggi dalam bidang sosial, ekonomi, kesehatan, maupun kebijakan publik, agar kontribusi metodologis penelitian menjadi lebih luas. Sebagai pelengkap analisis, penelitian lanjutan juga disarankan untuk menggunakan ukuran evaluasi berbasis prediksi, seperti *Mean Squared Error* (MSE) atau *Root Mean Squared Error* (RMSE), guna menilai kinerja prediktif model secara lebih kuantitatif.

DAFTAR PUSTAKA

- Akaike, H. (1987). Factor analysis and aic. *Psychometrika*, 52:317–332, <https://api.semanticscholar.org/CorpusID:47952861>.
- Andana, A. P., Safitri, D., & Rusgiyono, A. (2017). Model regresi menggunakan least absolute shrinkage and selection operator (lasso) pada data banyaknya gizi buruk kabupaten/kota di Jawa Tengah. *Jurnal Gaussian*, 6(1):21–30.
- Bjerkholt, O. (2018). Ragnar Frisch (1895-1973). *Memorandum (Institute of Pacific Relations, American Council)*, <https://api.semanticscholar.org/CorpusID:159243350>.
- Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference : A Practical Information-Theoretic Approach*, volume 2. Springer.
- Efron, B., Hastie, T. J., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32:407–499, <https://api.semanticscholar.org/CorpusID:121570279>.
- Fathurahman, M. (2010). Pemilihan model regresi terbaik menggunakan Akaike's information criterion. *J. EKSPONENSIAL*, 1(2):26–33.
- Fitria, E. R. & Fatchur, R. (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.
- Gujarati, D. N., Wibi Hardani, S., Saat, S., & Julius A Mulyadi, S. (2006). Dasar-dasar ekonometrika jilid 1.

- Madiatmoko, G. (2024). The application of the classical assumption test in multiple linear regression analysis (a case study of the preparation of the allometric equations of young makila). *JTAM (Jurnal Teori dan Aplikasi Matematika)*.
- 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*.
- Purba, S. A. (2021). Estimasi parameter data berdistribusi normal menggunakan maksimum likelihood berdasarkan newton raphson. *Jurnal Sains Dasar*, <https://api.semanticscholar.org/CorpusID:247426512>.
- 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.
- Rosyadi, M. Z. et al. (2018). Penerapan metode regresi ridge untuk mengatasi masalah multikolinearitas pada kasus indeks pembangunan manusia di provinsi jawa tengah.
- Sartika, I., Debataraja, N. N., & Imro'ah, N. (2020). Analisis regresi dengan metode least absolute shrinkage and selection operator (lasso) dalam mengatasi mul-

tikolinearitas. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 9(1).

Tarigan, P. B. (2023). Analisis asumsi klasik terhadap faktor-faktor produktivitas kerja tenaga pemanen di perkebunan kelapa sawit. *Agriprimatech*, 7.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

