

SKRIPSI

**OPTIMISASI ANALISIS REGRESI LASSO DENGAN
ALGORITMA LARS MENGGUNAKAN BAYESIAN
INFORMATION CRITERION (BIC)**

**STUDI KASUS: PENDUGAAN DATA INDEKS KETAHANAN
PANGAN TAHUN 2021**



**STATE ISLAMIC UNIVERSITY
ZULAEKA RITASARI
SUNAN KALIJAGA
20106010042
YOGYAKARTA**

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

2025

**OPTIMISASI ANALISIS REGRESI LASSO DENGAN
ALGORITMA LARS MENGGUNAKAN BAYESIAN
INFORMATION CRITERION (BIC)**
**STUDI KASUS: PENDUGAAN DATA INDEKS KETAHANAN
PANGAN TAHUN 2021**

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



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 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 : Zulaeka Ritasari
NIM : 20106010042
Judul Skripsi : Optimisasi Analisis Regresi LASSO dengan Algoritma LARS
Menggunakan Bayesian Information Criterion (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 mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 21 Januari 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-209/Un.02/DST/PP.00.9/01/2025

Tugas Akhir dengan judul : Optimisasi Analisis Regresi LASSO dengan Algoritma LARS menggunakan Bayesian Information Criterion (BIC) Studi Kasus : Pendugaan Data Indeks Ketahanan Pangan 2021

yang dipersiapkan dan disusun oleh:

Nama : ZULAEKA RITASARI
Nomor Induk Mahasiswa : 20106010042
Telah diujikan pada : Jumat, 24 Januari 2025
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: 679b16a24404f



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

Valid ID: 67970a80f3f66



Pengaji II
Muchammad Abrori, S.Si., M.Kom
SIGNED

Valid ID: 67970a866e261



Yogyakarta, 24 Januari 2025
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 679b3b453bd54

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Zulaeka Ritasari
NIM : 20106010042
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 Januari 2025



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

PENULIS MEMPERSEMBAHKAN KARYA SEDERHANA INI UNTUK:

KEDUA AYAH DAN MAMA DARI PENULIS, ADIK, KELUARGA BESAR,

SAHABAT, SERTA DIRI SENDIRI YANG TELAH MEMBERSAMAI DAN

MEMBERIKAN DUKUNGAN SELAMA PENELITIANINI.

KAMPUS UIN SUNAN KALIJAGA, IBU DAN BAPAK DOSEN, SERTA

CIVITAS AKADEMIKA YANG MEMBERIKAN PROSES, PENGALAMAN, SERTA

KENANGANNYA SELAMA PERJALANAN DALAM MENIMBA ILMU DI KAMPUS

TERCINTA INI.

"MATEMATIKA ANGKATAN 2020 TERCINTA"



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

"Beauty, Brain, and Behavior"

"Selesaikan pendidikanmu, bagaimana pun keadannya. Hatimu boleh patah, matamu boleh basah, kalau lagi stres ibadahnya ditingkatkan lagi, satu yang kamu ingat "jangan menyerah"."



PRAKATA

Alhamdulillahi rabbil 'alamin, segala puji syukur kepada Allah SWT atas limpahan rahmat, nikmat, berkat, dan hidayahnya berupa kekuatan, pertolongan, kesabaran, serta kelancaran. Sholawat salam semoga selalu tercurahkan kepada Nabi Agung Muhammad SAW dan para sahabatnya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Optimisasi Analisis Regresi LASSO dengan Algoritma LARS Menggunakan *Bayesian Information Criterion* (BIC) pada Data Indeks Ketahanan Pangan Tahun 2021".

Pada penelitian tugas akhir ini, penulis mengucapkan terima kasih kepada dosen pembimbing skripsi, yaitu Ibu Sri Utami Zuliana, S.Si.,M.Sc.,Ph.D., atas bimbingan, saran, dan pendampingannya selama proses penggeraan dari awal hingga akhir. Selain itu, penulis juga ingin mengucapkan terima kasih dan apresiasi kepada:

1. Rektor UIN Sunan Kalijaga Yogyakarta yaitu Bapak Prof. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D.
2. Dekan Fakultas Sains dan Teknologi yaitu Ibu Dr. Khurul Wardati, M.Si.
3. Ketua Program Studi Matematika yaitu Ibu Dr. Ephra Diana Supandi, S.Si., M.Sc.
4. Dosen Pembimbing Akademik yaitu Ibu Sri Istiyarti Uswatun Chasanah, M.Si. yang banyak membantu penulis dalam melewati pasang-surutnya dunia perkuliahan mulai dari awal perkuliahan hingga penulis sampai di titik ini.

5. Bapak/Ibu dosen program studi Matematika dan staf Fakultas Sains dan Teknologi yang telah memberikan ilmu dan layanan terbaiknya kepada penulis dari awal perkuliahan hingga akhir.
6. Kedua orang tua penulis, Ayah, Mama, Feri, Safa, Bude Ninik, dan Pakde Bambang, serta keluarga besar dari Ayah dan Mama atas pemberian cinta kasih, doa, bimbingan, serta pendampingannya sampai penulis mencapai di titik akhir perkuliahan ini. Bagian keluarga yang tidak terlupa yaitu Pusi, kucing tersayang yang menjadi penghibur ketika suka maupun duka.
7. Keluarga Besar Matematika Angkatan 2020 yang bersama-sama empat tahun perkuliahan, walaupun di awal hingga pertengahan perkuliahan kita terhalang jarak, namun dukungan dan kenangan manis kalian akan selalu penulis kenang.
8. Teman seerbimbingan penulis yaitu Anisa, Linggar, Adel, Uswa, Fatma, dan Yumna yang selalu memberikan dukungan ketika dilanda jalan buntu, namun uluran tangan kalian selalu membantu penulis untuk segera lulus dan menyelesaikan tugas akhir ini.
9. Teman-teman KKN 111 Kenalan merupakan keluarga baru yang selalu berbagi cerita senang dan susahnya selama 45 hari. Tak lupa Bapak Agus dan Ibu Yani yang super baik hati sekali, super sabar, dan selalu kami repotkan. Serta Ibu Riris pembimbing tercinta yang mengawal keberlangsungan KKN kami.
10. Teman dekat penulis Aretha, Fatin, Mufida, Fitri, Hanum, Mba Diana, dan Indah yang merupakan pengurus harian PC IPPNU Kabupaten Sleman yang selalu mengejar Sari agar cepat menyelesaikan tugas akhir, love you girls.

11. Teman SMA yang tak kan pernah putus komunikasi dan silaturahmi Nafisa, Dea, dan Regita.
12. Teman perkuliahan penulis yaitu Rayi, Miftah, Yuyun, Maul, Zila, Sekar yang selalu menemani pasang surut dinamika organisasi dan perkuliahan, kalian emang terbaik.
13. Terakhir penulis ingin memberikan ucapan terima kasih dan doa kepada seluruh pihak, semoga selalu dilancarkan dalam melakukan segala urusan.

Penulis menyadari atas ketidaksempurnaan pada skripsi ini. Oleh karena itu, penulis menerima atas segala saran, kritik, maupun masukan atas skripsi ini. Harapannya skripsi ini dapat bermanfaat bagi penulis dan semua kalangan yang membutuhkan referensi pada penelitian yang serupa. Akhir kata semoga Allah SWT selalu melimpahkan rahmat dan bimbingannya. *Aamiin.*

Yogyakarta, 6 Januari 2025



DAFTAR ISI

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

2.1.1. Indeks Ketahanan Pangan	9
2.1.2. Variabel Penelitian	9
2.2. Tinjauan Statistik	12
2.2.1. Analisis Regresi Linear	13
2.2.2. Multikolinearitas	14
2.2.3. Regresi Terpenalti	15
2.2.4. Least Absolute Shrinkage and Selection Operator (LASSO)	16
2.2.5. Algoritma Least Angle Regression and Shrinkage (LARS)	17
2.2.6. Bayesian Information Criterion	18
III METODE PENELITIAN	19
3.1. Metode Penelitian	19
3.2. Jenis dan Sumber Literatur	19
3.3. Studi Kasus	19
3.4. Metode Pengolahan Data	20
3.5. Software yang Digunakan	21
3.6. Diagram Analisis Data	22
IV HASIL DAN PEMBAHASAN	23
4.1. Statistika Deskriptif	23
4.2. Standarisasi Data	24
4.3. Analisis Regresi Linear	25
4.4. Uji Hipotesis	26
4.4.1. Uji F	26
4.4.2. Uji T	27
4.5. Uji Asumsi Klasik	28
4.5.1. Uji Normalitas	28
4.5.2. Uji Autokorelasi	29

4.5.3. Uji Heterokedastisitas	29
4.5.4. Uji Multikolinearitas	30
4.6. Analisis Regresi LASSO dengan Algoritma LARS	33
4.7. Menentukan Nilai Lambda Optimal dengan BIC	35
4.8. Pemilihan Model Terbaik dengan BIC	36
V PENUTUP	40
5.1. Kesimpulan	40
5.2. Saran	43
DAFTAR PUSTAKA	44
LAMPIRAN	48
A Data Penelitian	48
B Standarisasi Data	52
C Nilai Lambda pada Model Regresi LASSO	56
D Nilai BIC pada Setiap Lambda	58
E <i>Source Code Software R</i>	60
F Tinjauan Pustaka	70
G Curriculum Vitae	72

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR TABEL

3.1 Variabel Penelitian	20
4.1 Estimasi Parameter Regresi Linear	26
4.2 Nilai Variance Inflation Factor (VIF)	31
4.3 Interpretasi Koefisien Korelasi	33
4.4 Nilai Koefisien LASSO pada Tahapan Seleksi menggunakan LARS	35
4.5 Model Regresi LASSO dengan BIC (λ)	37
4.6 Koefisien Model Regresi LASSO	38
6.1 Persamaan dan Perbedaan Penelitian Terdahulu dengan Penelitian Sekarang	70
6.2 Persamaan dan Perbedaan Penelitian Terdahulu dengan Penelitian Sekarang	71

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

4.1	Boxplot	23
4.2	Boxplot Standarisasi Data	25
4.3	Plot korelasi antar variabel	32
4.4	Plot Pendugaan Koefisien Taksiran Regresi LASSO dengan Algoritma LARS	34
4.5	Plot BIC terhadap Lambda	37



DAFTAR LAMBANG

Y_i	= pengamatan dari variabel respon ke-i
X_j	= pengamatan dari variabel prediktor ke-j
β_0	= konstanta
β_j	= parameter regresi ke-j
ε_i	= galat pengamatan ke-i
R_j^2	= nilai koefisien determinasi
$VIF_{(j)}$	= faktor perubahan variansi dalam variabel bebas ke-j
L_1	= regularisasi LASSO
L_2	= regularisasi ridge
$\hat{\beta}^{lasso}$	= koefisien regresi LASSO
t	= nilai parameter <i>tuning</i> dengan $t \geq 0$
λ	= parameter penyusutan
$L(Y X, \beta, \sigma^2)$	= fungsi maksimum <i>likelihood</i>
p	= banyaknya variabel prediktor β_i
N	= banyaknya pengamatan
r	= residual

INTISARI

Optimisasi Analisis Regresi LASSO dengan Algoritma LARS Menggunakan Bayesian Information Criterion (BIC)

Studi Kasus: Pendugaan Data Indeks Ketahanan Pangan Tahun 2021

Zulaeka Ritasari

20106010042

Korelasi antara variabel prediktor pada analisis regresi dapat menjadi masalah atau pelanggaran asumsi klasik yang disebut dengan multikolinearitas. Multikolinearitas mengakibatkan variansi yang besar, sehingga estimasi parameter regresi menjadi tidak efisien. Deteksi multikolinearitas dilakukan dengan memperhatikan nilai *variance inflation factor* (VIF) dan koefisien korelasinya. Masalah tersebut dapat diatasi dengan menggunakan Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO). LASSO akan menyusutkan koefisien regresi mendekati nol atau sampai dengan nol yang perhitungannya dilakukan dengan bantuan algoritma *Least Absolute Regression and Shrinkage* (LARS). Dalam studi kasus pada data Indeks Ketahanan Pangan, variabel prediktor yang memiliki korelasi tinggi akan masuk secara bertahap ke dalam model. Variabel yang masuk pertama kali adalah X_4 kemudian dilanjutkan $X_9, X_2, X_1, X_8, X_6, X_{11}, X_7, X_5, X_3$ sampai dengan X_{10} . Hasil pendugaan koefisien pada setiap tahapan tersebut akan diperoleh beberapa model regresi. Tahapan-tahapan tersebut menghasilkan beberapa model regresi yang kemudian diuji dengan pemilihan model terbaik *Bayesian Information Criterion* (BIC). Nilai BIC minimum (0.342) menunjukkan lambda optimal sebesar 0.059, sehingga diperoleh model terbaik yaitu pada tahapan ke-35. Hasil penelitian menunjukkan tujuh dari sebelas variabel prediktor memiliki pengaruh terhadap indeks ketahanan pangan tahun 2021.

Kata Kunci: Multikolinearitas, LASSO, LARS, BIC, Indeks Ketahanan Pangan.

ABSTRACT

Optimization of LASSO Regression Analysis With The LARS Algorithm Using Bayesian Information Criterion (BIC)

Case study: Estimating Global Food Security Index (GFSI) 2021

Zulaeka Ritasari

20106010042

Correlation between predictor variables in regression analysis can be a problem or violation of linear regression assumptions that called multicollinearity. Multicollinearity causes the variance to become large, so that the estimation of regression parameters becomes inefficient. Multicollinearity can be detection by considering the variance inflation factor (VIF) value and its correlation coefficient. This problem can be overcome by using Least Absolute Shrinkage and Selection Operator (LASSO) Regression. LASSO will shrink the regression coefficient close to zero or exactly zero, the calculation of which is done with the help of the Least Absolute Regression (LARS) algorithm. In the case study on the Food Security Index data, predictor variables that have high correlation will be gradually entered into the model. The first variable to enter is X_4 then continued by $X_9, X_2, X_1, X_8, X_6, X_{11}, X_7, X_5, X_3$ up to X_{10} . The results of the coefficient estimation at each stage will produce several regression models. These stages produce several regression models which are then tested by selecting the best Bayesian Information Criterion (BIC) model. The minimum BIC value (0.342) indicates an optimal lambda of 0.059, so the best model is obtained at the thirty-fifth stage. The results of the study show that seven of the eleven predictor variables have an influence on the food security index in 2021.

Keywords: Multicollinearity, LASSO, LARS, BIC, Food Security Index.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Analisis regresi merupakan metode statistik yang memodelkan suatu hubungan antara satu atau lebih variabel prediktor dengan variabel respon. Berdasarkan jumlah variabel prediktornya, analisis regresi dibagi menjadi dua yaitu regresi linear sederhana (*simple regression*) dan regresi linear berganda (*multiple linear regression*). Analisis regresi linear sederhana bertujuan memodelkan hubungan antara satu variabel prediktor (X) dan satu variabel respon (Y). Sedangkan, analisis regresi linear berganda digunakan untuk menganalisa hubungan antara satu variabel respon dengan dua atau lebih variabel prediktor.

Dalam analisis regresi linear, terdapat koefisien regresi β yang digunakan untuk mengukur seberapa besar perubahan pada Y untuk setiap satuan perubahan X . Metode yang digunakan untuk mengestimasi koefisien regresi tersebut adalah Ordinary Least Squares (OLS) atau Metode Kuadrat Terkecil (MKT). OLS memiliki sifat BLUE (Best, Linear, Unbiased Estimator) yang harus memenuhi asumsi-asumsi. Asumsi yang dimaksud yaitu normalitas, homoskedastisitas, non-autokorelasi, dan non-multikolinearitas. Namun dalam prosesnya, sering ditemukan adanya asumsi yang tidak terpenuhi, seperti masalah multikolinearitas. Multikolinearitas menandakan adanya hubungan antar variabel prediktor. Adanya multikolinearitas tersebut, mengakibatkan estimasi parameter regresi menjadi tidak efisien, karena model memiliki variansi yang besar, sehingga perlu untuk mengatasinya.

Beberapa metode berikut dapat digunakan untuk mengatasi masalah multikolinearitas, seperti mengeluarkan variabel prediktor yang memiliki korelasi tinggi terhadap variabel prediktor lainnya, menambahkan ukuran data pengamatan, melakukan transformasi pada variabel prediktor yang memiliki masalah multikolinearitas, serta regresi terpenalti (Ridge dan LASSO) .

Penelitian ini menerapkan metode regresi *Least Absolute Shrinkage And Selection Operator* (LASSO) dalam mengatasi multikolinearitas. Penggunaan regresi LASSO didasarkan pada cara kerjanya yang dapat dengan menyusutkan koefisien regresi pada variabel prediktor yang memiliki korelasi tinggi hingga mendekati nol atau bahkan sampai dengan nol (Jolliffe, 2002). Oleh karena itu, LASSO dapat menyeleksi variabel prediktor dan model yang lebih sederhana. Perhitungan pada LASSO akan lebih akurat dengan bantuan algoritma LARS. Menurut Hastie (2008), *Least Angle Regression and Shrinkage* (LARS) merupakan algoritma yang lebih efisien, karena LARS memiliki modifikasi yang mempermudah komputasi pada LASSO. Proses algoritma LARS nantinya akan mengarahkan variabel prediktor berkorelasi tinggi secara terus-menerus hingga menuju pada titik paling rendah, sehingga korelasinya terhadap galat akan menurun.

Model analisis regresi yang diperoleh, juga perlu dilakukan pemilihan model terbaik. Tujuannya untuk menguji kualitas estimasi model, interpretasi model terhadap hubungan antar variabel, menghindari adanya *overfitting* (model yang terlalu rumit) atau *underfitting* (model terlalu sederhana) sehingga kurang mewakili pola dari data. Terdapat beberapa metode dalam pemilihan model terbaik, seperti *Bayesian Information Criterion (BIC)*, *Akaike Information Criterion (AIC)*, *Adjusted R-squared*, *Cross-Validation (CV)*, *Mallow's Cp*, *Generalized Cross Validation (GCV)*, *Mean Squared Error (MSE)*. Penelitian ini menggunakan *Bayesian Informa-*

tion Criterion (BIC) sebagai kriteria pemilihan model terbaik dalam model analisis regresi yang pertama kali ditemukan oleh Schwarz (1978). Schwarz menemukan BIC atas pengembangan dari beberapa teorema *Bayesian*, sehingga dikenal dengan istilah *Schwarz's Bayesian Information Criterion* (SBIC) atau *Schwarz's Bayesian Criterion* (SBC). Penggunaan BIC dalam pemilihan model terbaik menggunakan dasar dari metode *Maximum Likelihood Estimation* (MLE) dan teorema *Bayesian*. BIC bekerja dengan memberikan penalti yang lebih besar pada model yang kompleks dan cocok untuk sampel data yang besar.

Penelitian ini menggunakan data Indeks Ketahanan Pangan Tahun 2021. Ketahanan pangan merupakan faktor penting dari suksesnya pembangunan pada suatu negara. Hal itu telah diatur dalam UU No. 18 Tahun 2012 tentang pangan yang berbunyi "Ketahanan Pangan adalah kondisi terpenuhinya pangan bagi negara sampai dengan perseorangan, yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, beragam, bergizi, merata, dan terjangkau serta tidak bertentangan dengan agama, keyakinan, dan budaya masyarakat, untuk dapat hidup sehat, aktif, dan produktif secara berkelanjutan". Dalam rangka mengetahui Indeks ketahanan Pangan, Badan Pangan Nasional mencakup tiga aspek ketahanan pangan, yaitu ketersediaan pangan, keterjangkauan pangan, dan pemanfaatan pangan. Berdasarkan ketiga aspek tersebut, penelitian ini mengambil sebelas indikator dari turunannya yang akan diteliti pengaruhnya terhadap Indeks Ketahanan Pangan. Pada penelitian ini, ternyata ditemukan adanya masalah multikolinearitas pada data tersebut, dimana adanya hubungan yang kuat antar variabel prediktornya. Oleh karena itu, penelitian ini akan membahas solusi dari masalah multikolinearitas tersebut dengan analisis regresi LASSO menggunakan algoritma LARS dan pemilihan model terbaik BIC.

1.2. Rumusan Masalah

Berdasarkan uraian yang dijelaskan pada latar belakang diperoleh rumusan masalah pada penelitian ini, yaitu:

1. Bagaimana cara membentuk model analisis regresi LASSO dengan algoritma LARS menggunakan Bayesian Information Criterion (BIC)?
2. Apa saja faktor-faktor yang mempengaruhi Indeks Ketahanan Pangan tahun 2021?

1.3. Batasan Masalah

Pada penelitian ini, masalah yang dibatasi adalah sebagai berikut:

1. Penelitian ini menyelesaikan permasalahan multikolinearitas.
2. Metode statistiknya yaitu analisis regresi LASSO dengan algoritma LARS.
3. Pemilihan model terbaiknya berdasarkan nilai minimum *Bayesian Information Criterion* (BIC).
4. Data yang digunakan yaitu Indeks Ketahanan Pangan tahun 2021 pada 34 provinsi di Indonesia dengan indikator-indikator yang mempengaruhinya.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui langkah-langkah dari analisis regresi LASSO dengan algoritma LARS menggunakan pemilihan model terbaik *Bayesian Information Criterion* (BIC).

2. Membentuk model terhadap faktor-faktor yang mempengaruhi Indeks Ketahanan Pangan tahun 2021 yang terdapat masalah multikolinearitas.

1.5. Manfaat Penelitian

Manfaat dalam penelitian ini adalah:

1. Menambah referensi atau rujukan mengenai penerapan analisis regresi LAS-SO dalam menyelesaikan masalah multikolinearitas.
2. Pengembangan pada metode *Bayesian Information Criterion* (BIC) sebagai kriteria dalam memilih model regresi LASSO terbaik.
3. Menjadi dasar pertimbangan dalam meningkatkan kesejahteraan masyarakat melalui peningkatan Indeks Ketahanan Pangan dalam pemerintahan

1.6. Sistematika Penulisan

Sistematika penulisan penelitian ini yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan, dan tinjauan pustaka.

BAB II DASAR TEORI

Bab ini berisi dasar teori sebagai acuan dalam penelitian.

BAB III METODE PENELITIAN

Bab ini berisi langkah-langkah yang digunakan untuk penyelesaian masalah yang memuat metode penelitian, jenis dan literatur, studi kasus, metode pengolahan data, software yang digunakan, dan diagram analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dari analisis data yang diinterpretasikan secara detail.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran untuk penelitian-penelitian selanjutnya.

1.7. Tinjauan Pustaka

Pada penelitian yang dilakukan oleh Frans (2022), terdapat multikolinearitas pada variabel penanaman modal dalam negeri sebesar 12,570. Akibatnya, peduga koefisien regresi memiliki variansi yang besar. Solusinya menggunakan analisis regresi LASSO dengan algoritma LARS. Dalam prosesnya akan dimulai dari variabel prediktor dengan korelasi paling tinggi yang dimasukkan ke dalam model secara bertahap. Penelitian ini menggunakan *Cross Validation* dengan *mode fraction* dan *mode step* dalam pemilihan model terbaiknya. Nilai R-square dan MSE dari kedua metode tersebut menentukan model terbaik antara LASSO dan MKT. Hasil dari penelitiannya, analisis regresi LASSO menghasilkan model yang lebih baik daripada MKT. Berdasarkan hasil tersebut penelitian ini telah membuktikan bahwa LASSO dapat menghasilkan model yang lebih sederhana dan efisien untuk mengatasi permasalahan multikolinearitas.

Pada penelitian yang dilakukan oleh Januaviani (2019) menggunakan metode LASSO dengan algoritma LARS, serta proses *Mallows's Cp* sebagai pemilihan model terbaik. Data Deposit Devisa Indonesia menunjukkan permasalahan multikolinearitas yang sangat tinggi mencapai nilai VIF sebesar 419.722, 14.224, 369.514 berturut-turut pada variabel X_1 , X_2 , X_3 . Hal ini dianggap sebagai kelemahan dari analisis regresi karena tidak dapat dikatakan bersifat BLUE (Best Linear Unbiased Estimator) dan memiliki galat yang besar. Oleh karena itu digunakan metode LASSO yang dapat mengurangi multikolinearitas, sehingga meningkatkan keakuratan pada analisis regresi. Pada penelitian ini nilai *Mallows's Cp* yang paling kecil menunjukkan model terbaik pada tahapan ke-7 LASSO dengan algoritma LARS.

Pada penelitian yang dilakukan oleh Jamco (2023) menggunakan data faktor yang mempengaruhi PDRB di Provinsi Maluku dengan permasalahan multikolinearitas pada estimasi model OLS. Tujuan penelitian ini yaitu menentukan model regresi LASSO dengan algoritma LARS dalam mengatasai masalah tersebut. Pemilihan model terbaik LASSO di penelitian ini menggunakan kriteria *Cross Validation* diperoleh bahwa variabel pertanian, pertambangan dan penggalian, industri pengolahan, serta perdagangan hotel dan restoran memiliki pengaruh terhadap variabel PDRB di Provinsi Maluku.

Pada penelitian yang dilakukan oleh Mait (2021) dilakukan perbandingan kebaikan model pada metode regresi kuadrat terkecil (MKT) dan LASSO. Berdasarkan hasil uji menggunakan *Variance Inflation Factor* (VIF) diperoleh lima variabel yang mengandung masalah multikolinearitas. Hal tersebut berpengaruh dalam penggunaan MKT, sehingga hasil yang diperoleh tidak konsisten yang ditunjukkan dari nilai R^2 tinggi, akan tetapi memiliki pengaruh antara variabel prediktor dan variabel respon. Ketidakefektifan tersebut dinilai dari tingginya nilai RMSEP dan MAE pada metode MKT. Jika dibandingkan dengan LASSO, nilai RMSEP dan MAE cenderung mendekati 0 (nol), sehingga menghasilkan model regresi yang baik.

Pada penelitian yang dilakukan oleh Yanke (2022) mendapat penggunaan variabel prediktor yang banyak, sehingga menimbulkan masalah multikolinearitas. Penelitian ini menggunakan beberapa metode dalam penyelesaian masalah korelasi yang tinggi diantara variabel prediktornya. Metode pertama yaitu seleksi variabel prediktor dengan metode *backward*, *forward*, dan *stepwise*. Namun, hasil dari metode tersebut masih didapati adanya multikolinearitas. Selanjutnya, digunakan alternatif dengan Regresi PLS, Regresi Komponen Utama, Regresi Regularisasi

(LASSO, Ridge, dan Elastic Net). Pemilihan model terbaik penelitian ini didasarkan pada nilai *Cross Validation* yang paling kecil, sehingga dapat diperoleh model regresi LASSO dengan penanganan multikolinearitas terbaik.

Berdasarkan beberapa penelitian yang telah dipaparkan, dapat ditarik kesimpulan bahwasanya penerapan analisis regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan algoritma *Least Angle Regression* (LARS) efektif dalam menyelesaikan permasalahan multikolinearitas. Namun, didapatkan banyak penelitian menggunakan proses *Cross Validation* dan *Mallows's Cp* dalam pemilihan model terbaiknya. Dari beberapa penelitian belum banyak yang menggunakan *Bayesian Information Criterion* (BIC) sebagai metode pemilihan model terbaik. Oleh karena itu, penulis akan melakukan penelitian mengenai penyelesaikan masalah multikolinearitas pada data menggunakan metode regresi LASSO dengan algoritma LARS dan pemilihan model terbaik menggunakan *Bayesian Information Criterion* (BIC) pada data Indeks Ketahanan Pangan di Indonesia Tahun 2021.



BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dari analisis regresi LASSO dengan algoritma LARS pada data Indeks Ketahanan Pangan di Indonesia tahun 2021, diperoleh kesimpulan sebagai berikut:

1. Dalam data IKP terdapat permasalahan multikolinearitas yang dideteksi menggunakan nilai VIF > 10 , yaitu pada variabel Indeks Pembangunan Manusia (X_{10}) sebesar 13.331. Multikolinearitas berhasil diatasi menggunakan algoritma LARS pada regresi LASSO yang dilakukan secara bertahap dengan menetapkan koefisien awal semua variabel bernilai nol. Kemudian, variabel prediktor yang memiliki korelasi paling tinggi terhadap galat akan masuk yaitu X_4 . Variabel selanjutnya akan masuk secara bertahap yaitu X_9 , X_2 , X_1 , X_8 , X_6 , X_{11} , X_7 , X_5 , X_3 , dan X_{10} . Dari 12 tahapan tersebut, diperoleh beberapa model regresi yang akan dipilih berdasarkan nilai lambda optimal. Pemilihan lambda optimal diperoleh dengan menggunakan nilai BIC minimum. Dengan demikian, diperoleh model regresi LASSO terbaik dengan nilai koefisien yang cenderung menyusut ke arah nol atau bahkan nol.
2. Hasil penelitian ini diperoleh model terbaik regresi LASSO dengan algoritma LARS pada nilai lambda optimum (0.059) yang didasarkan pada nilai BIC

minimum (0.342) adalah:

$$\hat{Y} = -1.617 \times 10^{16} + 0.155X_1 - 0.204X_2 + 0.548X_4 - 0.073X_6 + 0.127X_8 + 0.188X_9 - 0.009X_{11} \quad (5.1)$$

Dari persamaan 5.1 menunjukkan variabel-variabel prediktor yang memiliki pengaruh signifikan terhadap Indeks Ketahanan Pangan. Interpretasi dari persamaan 5.1 model regresi LASSO yaitu:

- (a) Jika faktor lain dianggap konstan, ketika produktivitas tanaman padi (X_1) meningkat 1%, maka persentase indeks ketahanan pangan (IKP) akan meningkat sebesar 0.155. Hal ini menandakan produktivitas tanaman padi berkontribusi secara langsung terhadap ketersediaan pangan dengan jumlah beras yang meningkat, sehingga konsumsi masyarakat juga akan meningkat. Oleh karena itu, meningkatnya produktivitas padi tersebut mampu mengurangi kerawanan pangan dan meningkatkan nilai indeks ketahanan pangan.
- (b) Jika faktor lain dianggap konstan, ketika penduduk miskin (X_2) meningkat 1%, maka persentase indeks ketahanan pangan (IKP) akan menurun sebesar 0.204. Hal ini juga berarti jika penduduk miskin tidak mempunyai daya beli yang mencukupi dalam membiayai kebutuhan dasar hidup, seperti memperoleh akses makanan yang bergizi. Dalam hal tersebut, upaya dalam meningkatkan ketahanan pangan harus mempertimbangkan pada kondisi ekonomi masyarakat.
- (c) Jika faktor lain dianggap konstan, ketika rumah tangga dengan akses penerangan listrik (X_4) meningkat 1%, maka persentase indeks ketahanan pangan (IKP) akan meningkat sebesar 0.548. Tersedianya akses penerangan listrik akan menciptakan peluang terhadap akses pekerjaan, se-

hingga dapat mendorong pertumbuhan ekonomi dan kesejahteraan yang akan berdampak pada ketahanan pangan di daerah tersebut.

- (d) Jika faktor lain dianggap konstan, ketika rata-rata lama sekolah perempuan (X_6) meningkat 1%, maka persentase indeks ketahanan pangan (IKP) akan menurun sebesar 0.073. Hal tersebut bertolakbelakang secara teoritis, karena seiring dengan meningkatnya rata-rata lama sekolah pada perempuan mempengaruhi pada kesadaran dan pengetahuan terhadap makanan yang bergizi dan pengelolaan pangan yang efektif.
- (e) Jika faktor lain dianggap konstan, ketika kasus *stunting* (X_8) meningkat 1%, maka persentase indeks ketahanan pangan (IKP) akan meningkat sebesar 0.127. Hal tersebut bertolakbelakang secara teoritis, karena meningkatnya kasus stunting menandakan adanya permasalahan terhadap rendahnya akses makanan yang bergizi dan ketahanan pangan dalam suatu daerah. Namun, kasus stunting pada model ini memiliki pengaruh yang relatif kecil jika dibandingkan dengan variabel lainnya, sehingga kasus stunting tidak berpengaruh besar terhadap peningkatan indeks ketahanan pangan.
- (f) Jika faktor lain dianggap konstan, ketika angka harapan hidup (X_9) meningkat 1%, maka persentase indeks ketahanan pangan (IKP) akan meningkat sebesar 0.188. Tingginya angka harapan hidup menunjukkan adanya akses makanan dengan kualitas yang baik dan akses pelayanan kesehatan yang memadai bagi ibu hamil dan masyarakat. Kualitas hidup yang sehat dapat meningkatkan produktivitas kerja yang mendukung aktivitas ekonomi, termasuk produksi dan distribusi pangan.
- (g) Jika faktor lain dianggap konstan, ketika laju PDRB (X_{11}) meningkat

1%, maka persentase indeks ketahanan pangan (IKP) akan menurun sebesar 0.009. Hubungan ini bertolakbelakang secara teoritis karena umumnya peningkatan PDRB dapat menurunkan tingkat kemiskinan, sehingga akses terhadap pangan dapat meningkat. Namun hal ini tidak lepas dari fakta bahwa meningkatnya laju PDRB yang tidak merata akan menyebabkan ketimpangan ekonomi pada populasi tertentu yang tetap mengalami kerawanan pangan.

5.2. Saran

Saran untuk penelitian berikutnya yaitu dapat mengubah, menambahkan, atau bahkan membandingkan dengan metode pemilihan model terbaik selain BIC, seperti AIC, CGV, CV, dan lainnya. Hal tersebut akan mempengaruhi penelitian ke-depannya mengenai keakuratan pada masing-masing metode. Penelitian ini dapat digunakan sebagai landasan teori dalam pengolahan data-data lainnya yang mengandung permasalahan multikolinearitas.

DAFTAR PUSTAKA

- Ardiningrum & Umiyati (2021). Pengaruh indeks ketahanan pangan, pengeluaran pemerintah di bidang pendidikan dan kesehatan terhadap indeks pembangunan manusia pada 10 provinsi di pulau sumatera. *e-Jurnal Ekonomi Sumberdaya dan Lingkungan*.
- BPN (2022). *Indeks Ketahanan Pangan Tahun 2022*. Badan Pangan Nasional.
- Eilers & Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–102.
- Frans, Setyo Wira Rizki, D. K. (2022). Pemodelan pertumbuhan ekonomi kalimantan barat menggunakan pendekatan least absolute shrinkage and selection operator (lasso). *Buletin Ilmiah Math. Stat dan Terapannya (Bimaster)*, 11(1):111–120.
- Gujarati & Porter (2009). *Basic Econometrics Fifth Edition*. New York: McGraw-Hill.
- Gujarati, D. N. (2003). *Basic Econometrics Fourth Edition*. New York: McGraw-Hill.
- Hastie, Robert Tibshirani, J. F. (2008). *The Elements of Statistical Learning Data Mining, Inference, and Prediction Second Edition*. New York: Springer.
- Hidayah, L. (2021). *Pengaruh Nilai Tukar Petani (NTP), Luas Lahan Pertanian, Produk Domestik Regional Bruto dan Populasi terhadap Ketahanan Pangan di Indonesia*. PhD thesis, Universitas Negeri Malang.

- Jamco, F. Kondo Lembang, M. S. N. V. D. (2023). Penanganan multikolinieritas pada regresi linier berganda menggunakan regresi lasso. *Jurnal Matematika, Statistika dan Terapannya*, 2(2):145–154.
- Januaviani, Nurul Gusriani, K. J. S. S. A. T. B. (2019). The lasso (least absolute shrinkage and selection operator) method to predict indonesian foreign exchange deposit data. pages 2385–2392.
- Jolliffe, I. (2002). *Principal Component Analysis Second Edition*. New York: Springer.
- Khoirunissa, e. a. (2024). Analisis faktor-faktor penyebab inflasi di indonesia menggunakan regresi ridge, lasso, dan elastic-net. *Indonesian Journal of Applied Statistics*, 7(2):121–130.
- Mahalani & Rifai (2022). Least absolute shrinkage and selection operator (lasso) untuk mengatasi multikolinearitas pada model regresi linear berganda. *Bandung Conference Series: Statistics*, 2(2):119–125.
- Mait, Deiby Tineke Salaki, H. A. H. K. (2021). Kajian model prediksi metode least absolute shrinkage and selection operator (lasso) pada data mengandung multikolinearitas. *Jurnal Matematika dan Aplikasi*, 10(2):69–75.
- Ningsih, Intan, Y. (2023). Pemodelan tindak pidana kriminalitas di kota tangerang menggunakan metode regresi lasso. *Estimasi: Journal of Statistics and Its Application*, 4(1):64–77.
- Pardede, Sumargo, R. (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.

Prabowo, e. a. (2015). Pemodelan pertumbuhan ekonomi jawa tengah menggunakan pendekatan least absolute shrinkage and selection operator (lasso). *Jurnal Gaussian*, 4(4):855–864.

Rahmawati & Suratman (2022). Performa regresi ridge dan regresi lasso pada data dengan multikolinearitas. *Leibniz: Jurnal Matematika*, 2(2):1–10.

Robbani, e. a. (2019). Regresi least absolute shrinkage and selection operator (lasso) pada kasus inflasi di indonesia tahun 2014-2017. *Jurnal EurekaMatika*, 7(2):1–16.

Ruppert & Carroll (1997). Penalized regression splines. *Unpublished manuscript*.

R.W.Kennard, A. H. . (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Sartika, e. a. (2020). Analisis regresi dengan metode least absolute shrinkage and selection operator (lasso) dalam mengatasi multikolinearitas. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 9(1):31–38.

Schober, Boer, A. S. (2018). Correlation coefficients: Appropiriate use and interpretation. *Anesthesia Analgesia*, 126(5):1763–1768.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

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

Yanke, Nofrida Elly Zendrato, A. M. S. (2022). Penanganan masalah multikolinearitas pada pemodelan pertumbuhan ekonomi indonesia berdasarkan teori pertumbuhan.

buhan ekonomi endogenous. *Indonesian Journal of Statistics and Its Applications*, 6(2):228–244.

