

**PEMODELAN REGRESI *LEAST ABSOLUTE SHRINKAGE*
AND SELECTION OPERATOR DENGAN ALGORITMA *LEAST*
ANGLE REGRESSION DALAM MENGATASI
MULTIKOLINIERITAS PADA PENGELUARAN PER KAPITA
DI WONOSOBO**

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh
Gelar Sarjana Matematika
dalam Ilmu Matematika



Oleh : **DELLA SETYA RAHMA**
NIM : 2108046026

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO
SEMARANG
2025

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama : Della Setya Rahma
NIM : 2108046026
Jurusan/Program Studi : Matematika/ Matematika

menyatakan bahwa skripsi yang berjudul :

PEMODELAN REGRESI *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* DENGAN ALGORITMA *LEAST ANGLE REGRESSION* DALAM MENGATASI MULTIKOLINIERITAS PADA PENGELUARAN PER KAPITA DI WONOSOBO

secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 23 Juni 2025

Pembuat pernyataan,



Della Setya Rahma

NIM : 2108046026



KEMENTERIAN AGAMA R.I.
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka (Kampus II) Ngaliyan Semarang
Telp. 024-7601295 Fax. 7615387

PENGESAHAN

Naskah skripsi berikut ini :

Judul : **PEMODELAN REGRESI LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR DENGAN ALGORITMA LEAST ANGLE REGRESSION DALAM MENGATASI MULTIKOLINIERITAS PADA PENGELUARAN PER KAPITA DI WONOSOBO**

Penulis : Della Setya Rahma

NIM : 2108046026

Jurusan : Matematika

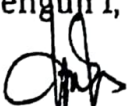
Telah diujikan dalam sidang *tugas akhir* oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Ilmu Matematika.

Semarang, 30 Juni 2025

DEWAN PENGUJI

Penguji I,

Penguji II,


Seftina Diah Miasary, M.Sc

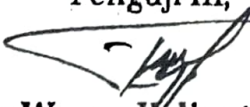

Zulaikha, M.Si.

NIP: 198709212019032010

NIP: 199204092019032027

Penguji III,

Penguji IV,

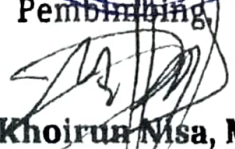

Agus Wayan Yulianto, M.Sc.


Yolanda Norasia, M.Si.

NIP: 198907162019031607

NIP: 199409232019032011

Pembimbing


Eva Khoirun Nisa, M.Si.

NIP: 198701022019032010

NOTA DINAS

Semarang, 17 Juni 2025

Yth. Ketua Program Studi Matematika
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum warahmatullahi wabarakatuh

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul : PEMODELAN REGRESI *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* DENGAN ALGORITMA *LEAST ANGLE REGRESSION* DALAM MENGATASI MULTIKOLINIERITAS PADA PENGELUARAN PER KAPITA DI WONOSOBO

Nama : Della Setya Rahma
NIM : 2108046026
Jurusan : Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqasyah.

Wassalamu'alaikum warahmatullahi wabarakatuh

Pembimbing,



Eva Khoirun Nisa, M.Si.

NIP : 19870102 201903 2 010

ABSTRAK

Pengeluaran per kapita digunakan sebagai indikator standar hidup yang dipengaruhi oleh pengetahuan dan peluang untuk menghasilkan barang dan jasa sebagai sumber pendapatan. Dalam penelitian ini, dilakukan pemodelan dari faktor-faktor yang diduga memengaruhi pengeluaran per kapita menggunakan regresi LASSO (*Least Absolute Shrinkage and Selection Operator*) dengan algoritma LAR (*Least Angle Regression*) karena pada faktor-faktor yang diduga memengaruhi pengeluaran per kapita terdeteksi multikolinieritas. LASSO menyusutkan koefisien regresi dari variabel bebas yang memiliki korelasi tinggi dengan variabel lain menjadi tepat pada nol atau mendekati nol. Berdasarkan hasil analisis, model terbaik LASSO diperoleh dengan menggunakan validasi silang pada langkah ke-2 (dimulai dari langkah ke-0) dengan nilai s (besarnya penalti/hukuman) = 0,5, menghasilkan *Mean Squared Error* (MSE) sebesar 0,2397145 dan koefisien determinasi R_k^2 sebesar 0,89. Nilai MSE yang kecil dan R_k^2 yang mendekati 1 menunjukkan bahwa model memiliki tingkat akurasi prediksi yang baik. Hasil akhir menunjukkan bahwa variabel Indeks Pembangunan Manusia dan Rata-rata Lama Sekolah berpengaruh terhadap pengeluaran per kapita di Kabupaten Wonosobo tahun 2015-2024.

Kata kunci : multikolinieritas, LASSO, LAR, pengeluaran per kapita

KATA PENGANTAR

Puji dan syukur kehadiran Tuhan Yang Maha Esa penulis panjatkan dengan segala kerendahan hati karena berkat dan rahmat-Nya penulis dapat menyelesaikan Skripsi yang berjudul “Pemodelan Regresi *Least Absolute Shrinkage And Selection Operator* Dengan Algoritma *Least Angle Regression* Dalam Mengatasi Multikolinieritas Pada Pengeluaran Per Kapita Di Wonosobo”. Sholawat serta salam senantiasa tercurahkan kepada junjungan Nabi Besar Muhammad SAW, suri teladan seluruh umat manusia, yang ajarannya senantiasa menjadi cahaya penerang dalam kehidupan dan syafa’atnya selalu dinantikan di hari akhir kelak.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika (S.Mat.) pada Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang. Proses penyusunan Skripsi ini tidak lepas dari doa, bantuan, bimbingan, motivasi dan peran dari banyak pihak. Sehingga penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak, terutama ucapan terima kasih penulis sampaikan kepada:

1. Allah SWT yang dengan kasih sayang-Nya telah memberikan kesehatan, kesabaran, serta kelancaran dalam setiap proses penulis hingga Skripsi ini dapat terselesaikan;
2. Bapak Prof. Nizar, M.Ag., selaku Rektor UIN Walisongo Semarang;
3. Bapak Prof. Dr. H. Musahadi, M.Ag., selaku Dekan Fakultas Sains dan Teknologi UIN Walisongo Semarang;

4. Ibu Any Muanalifah, M.Si, Ph.D., selaku Ketua Program Studi Matematika;
5. Ibu Eva Khoirun Nisa, M.Si., selaku Dosen Pembimbing yang telah meluangkan waktu, memberikan bimbingan, arahan dan masukan yang berharga dari awal hingga akhir penulisan Skripsi ini;
6. Teristimewa untuk kedua orang tua tercinta dan *support system* terbaik, Bapa Yatino dan Mama Fatimah, yang telah memberikan kasih sayang, nasihat, motivasi, serta memberikan dukungan dan doa yang tiada henti sejak penulis dilahirkan hingga saat ini. Semoga Allah SWT senantiasa melimpahkan kesehatan, kebahagiaan, dan keberkahan usia kepada Bapa dan Mama, serta membalas segala kebaikan dan pengorbanan yang telah diberikan dengan pahala yang berlipat ganda, karena telah memberikan figur orang tua terbaik bagi penulis;
7. Saudara tersayang, Mba Munafi'ah dan Adik Gian Aditya Ardhani, yang telah menjadi bagian dalam proses penyusunan Skripsi ini. Terima kasih atas canda tawa yang selalu menghibur penulis dikala sedih dan dukungan yang tidak ada habisnya;
8. Sahabat terbaik penulis sedari SMK hingga tua nanti, Sri Wulandari, Nurul Holifah, Sekar Widianti dan Aulia Eka Martanti yang selalu memberi semangat, canda tawa, serta menjadi tempat berbagi cerita. Terima kasih atas kebersamaan dan *support* yang telah diberikan. Semoga persahabatan ini tetap terjalin dengan baik hingga di masa

depan;

9. Teman sekaligus sahabat penulis, Nelis Sa'adah, Siti Hidayatun Nisyak, Ely Chabibah, dan Tiara Fitrianingtyas. Kita memang tidak begitu dekat sejak awal perkuliahan, tetapi melalui proses menyusun Skripsi ini, kita menjadi semakin akrab. Terima kasih karena telah menemani penulis, menunjukkan kepedulian, menjadi tempat berbagi cerita dan keluh kesah penulis, serta saling mendukung satu sama lain di tengah perjuangan menyelesaikan Skripsi ini;
10. Sahabat penulis sejak awal perkuliahan hingga saat ini, Indy Naurah 'Atiqah. Terima kasih karena telah membuat penulis tidak merasa sendirian di perantauan ini. Semoga kita dapat meraih kesuksesan bersama di masa depan;
11. Teman-teman seperjuangan Program Studi Matematika angkatan 2021, terutama Siti Nur Arafah, Ayu Imtiyaz Sari dan Aeni Nurjanah. Terima kasih karena telah memberikan dukungan dan semangat satu sama lain serta memberikan pengalaman yang berharga yang akan penulis kenang;
12. Teman-teman kos Mbah Pur, terutama Fitria Khoirunnisa dan Naili Nikmatul Magfiroh. Terima kasih telah menemani penulis, menghibur di saat penat, dan membuat penulis merasa betah selama tinggal di kos. Kebersamaan kalian menjadi salah satu bagian berharga dalam perjalanan ini;
13. Teman-teman Benten, yang telah menjadi teman sejak awal perkuliahan. Terima kasih atas kebersamaan, canda tawa, serta semangat yang diberikan selama masa kuliah. Semoga

silaturahmi ini tetap terjaga meski kita menempuh jalan masing-masing setelah lulus;

14. Semua Pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi hingga selesainya Skripsi ini.
15. Della Setya Rahma, ya! diri saya sendiri. Apresiasi sebesar-besarnya atas perjuangan yang telah dilakukan untuk menyelesaikan apa yang dimulai. Tidak mudah untuk dapat bertahan sampai di titik ini. Terima kasih untuk tetap hidup dan merayakan dirimu sendiri. *God, thank You for shaping me into a woman who can stand on her own. I'm proud of every achievement I've accomplished and finally, I've made it to this point.*

Semoga kebaikan semuanya menjadi amal ibadah yang diterima dan mendapat pahala yang berlimpah dari Allah SWT. Aamiin.

Penyusunan Skripsi ini bukanlah suatu hal yang mudah, namun dengan dukungan dari semua pihak yang telah disebutkan di atas, penulis dapat menyelesaikannya. Semoga Skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan menjadi sumbangan kecil bagi peningkatan kualitas hidup masyarakat, khususnya untuk mengetahui faktor apa saja yang memengaruhi Pengeluaran Per Kapita di Kabupaten Wonosobo. Atas segala kekurangan dan kelemahan dalam Skripsi ini penulis mengharapkan saran dan kritik yang membangun. Semoga karya tulis yang sederhana ini dapat menjadi bacaan yang bermanfaat dan dapat dikembangkan bagi peneliti-peneliti selanjutnya.

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN KEASLIAN	ii
PENGESAHAN	iv
NOTA DINAS	iv
ABSTRAK	v
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	6
C. Tujuan Penelitian	6
D. Manfaat Penelitian	7
E. Batasan Masalah	8
BAB II LANDASAN TEORI	9
A. Regresi Linier Berganda	9
B. Uji Multikolinieritas	12
C. Standarisasi Data	15
D. Regresi <i>Least Absolute Shrinkage and Selection Operator</i> (LASSO)	17
E. Algoritma <i>Least Angle Regression</i> (LAR)	20
F. Validasi Silang	27
G. Pengeluaran Per Kapita	28
H. Penelitian Terdahulu	30
BAB III METODE PENELITIAN	35
A. Jenis Penelitian	35
B. Sumber Data	35
C. Variabel Penelitian	36
D. Analisis Data	36

E.	Alur Penelitian	37
BAB IV HASIL DAN PEMBAHASAN		41
A.	Deskriptif Data	41
B.	Model Awal Menggunakan Regresi Linier Berganda	42
C.	Uji Multikolinieritas	43
D.	Standarisasi data	43
E.	Penanganan multikolinieritas menggunakan regresi LASSO dengan algoritma LAR	46
F.	Uji validasi silang	60
G.	Membentuk model terbaik LASSO dengan algoritma LAR	64
H.	Uji Koefisien Determinasi Model Regresi LASSO	67
I.	Membandingkan koefisien regresi linier berganda dengan regresi LASSO	68
J.	Membandingkan Nilai VIF dalam mendeteksi multikolinieritas menggunakan regresi Linier Berganda dan regresi LASSO	69
K.	Interpretasi hasil	70
BAB V KESIMPULAN DAN SARAN		73
A.	Kesimpulan	73
B.	Saran	74
DAFTAR PUSTAKA		75
Lampiran-lampiran		80

DAFTAR TABEL

Tabel	Judul	Halaman
Tabel 3.1	Variabel Penelitian	36
Tabel 4.1	Deskriptif Data	41
Tabel 4.2	Hasil Uji Multikolinieritas	43
Tabel 4.3	Standarisasi Data	46
Tabel 4.4	Nilai Variabel Aktif Tahap 1	49
Tabel 4.5	Nilai Koefisien LASSO pada Setiap Langkah	59
Tabel 4.6	Urutan Masuknya Variabel ke dalam Model LASSO	60
Tabel 4.7	Nilai Validasi Silang dengan <i>Mode Steps</i>	61
Tabel 4.8	Nilai s pada Setiap Langkah Algoritma LAR	63
Tabel 4.9	Koefisien Regresi LASSO	64
Tabel 4.10	Hasil Koefisien Determinasi Model LASSO	67
Tabel 4.11	Perbandingan Koefisien Regresi Linier Berganda dan Regresi LASSO	68
Tabel 4.12	Perbandingan nilai VIF pada Regresi Linier Berganda dan LASSO	69

DAFTAR GAMBAR

Gambar	Judul	Halaman
Gambar 3.1	Alur Penelitian	40
Gambar 4.1	Nilai Validasi Silang dengan <i>Mode Steps</i>	62
Gambar 4.2	Nilai <i>s</i> dengan <i>Mode Steps</i>	63

DAFTAR LAMPIRAN

	Halaman
Lampiran 1	Data Penelitian Pengeluaran Per Kapita Kabupaten Wonosobo Tahun 2015-2024 dan Faktor-Faktor yang Diduga Mempengaruhinya
Lampiran 2	Deskriptif data
Lampiran 3	Model Awal Regresi Linier Berganda dengan OLS
Lampiran 4	Uji Multikolinieritas
Lampiran 5	Standarisasi Data
Lampiran 6	Regresi LASSO dengan Algoritma LAR
Lampiran 7	Validasi Silang
Lampiran 8	Model Awal Regresi Linier Berganda dengan OLS yang Telah Distandarisasi
Lampiran 9	Model Terbaik Regresi LASSO dan Nilai Koefisien Determinasi
Lampiran 10	Nilai Intersep pada Model Regresi LASSO
Lampiran 11	Nilai VIF Setelah Menggunakan Regresi LASSO

BAB I

PENDAHULUAN

A. Latar Belakang

Multikolinieritas merupakan salah satu permasalahan umum yang sering muncul dalam analisis regresi linier berganda, yaitu ketika dua atau lebih variabel bebas terdapat hubungan yang kuat. Hal ini menyebabkan estimasi koefisien pada model regresi linier menjadi tidak stabil (Montgomery dkk., 2021). Menurut Gujarati & Porter (2009), multikolinieritas terjadi ketika variabel-variabel bebas dalam model regresi memiliki hubungan linier yang kuat, ditandai dengan koefisien korelasinya tinggi atau bahkan bernilai satu. Seharusnya, model regresi yang baik tidak terjadi korelasi pada variabel bebas (Ghozali, 2005). Menurut Widarjono (2018) multikolinieritas juga memiliki arti yang sama yaitu terjadi ketika variabel bebas dalam regresi linier berganda memiliki korelasi yang tinggi antar variabel bebas, sehingga dapat memengaruhi hasil estimasi model.

Pada beberapa penelitian, multikolinieritas kerap muncul sebagai salah satu masalah yang sering terjadi. Untuk mengatasi hal ini perlu adanya model statistik yang dapat mengatasinya, diantaranya yaitu Regresi Ridge, *Principal Component Regression* (PCR), *Least Absolute Shrinkage and Selection Operator* (LASSO), *Elastic Net*, dan beberapa model lainnya. Dari beberapa model yang dapat diterapkan sebagai solusi untuk mengatasi multikolinieritas, regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan model yang paling tepat karena berdasarkan penelitian Rahmawati & Suratman (2022) menjelaskan bahwa regresi LASSO

mampu mengatasi multikolinieritas dengan menyederhanakan model melalui pemilihan variabel bebas yang berpengaruh terhadap variabel terikat. Artinya, LASSO dapat menghilangkan variabel yang tidak berpengaruh dengan cara menyusutkan koefisiennya menjadi nol. Berbeda dengan *Ridge* atau *Elastic Net* yang cenderung mempertahankan semua variabel dalam model, meskipun pengaruhnya kecil. Ketika dibandingkan dengan PCR, regresi LASSO juga terbukti lebih efektif karena mampu mengatasi multikolinieritas secara langsung melalui proses seleksi variabel, yaitu dengan mengeliminasi variabel yang tidak berpengaruh tanpa harus mengubah seluruh variabel menjadi komponen utama (*principal components*). Sedangkan, PCR menangani multikolinieritas dengan mengubah variabel asli menjadi beberapa komponen utama yang tidak saling berkorelasi, kemudian menggunakan komponen-komponen tersebut dalam model regresi. Meskipun efektif, model ini tidak menggunakan variabel asli secara langsung, sehingga menyulitkan dalam menginterpretasikan pengaruh masing-masing variabel bebas terhadap variabel terikat (Nasution & Pane, 2024).

Selain itu, pada penelitian Tibshirani (1996) menjelaskan bahwa regresi LASSO merupakan suatu model estimasi yang meminimumkan jumlah kuadrat *error* guna menyeleksi variabel dengan menambahkan penalti (hukuman) untuk menyederhanakan model. Penalti ini dapat mengakibatkan beberapa koefisien menjadi nol sehingga menghasilkan model yang lebih sederhana, lebih mudah diinterpretasikan, dan akurasi prediksinya tetap optimal meskipun jumlah pengamatan terbatas (Melkumova & Shatskikh, 2017).

Regresi LASSO juga mempunyai beberapa algoritma yang dapat

diterapkan untuk menyelesaikan permasalahan optimasi, seperti *Least Angle Regression* (LAR), *General Linear Model Net* (GLMNET), *Coordinate Descent*, *Gradient Descent*, *Dual Coordinate Descent*, dan lain sebagainya. Algoritma yang diterapkan dalam penelitian ini adalah *Least Angle Regression* (LAR) karena dapat memudahkan dalam memahami urutan variabel yang masuk ke dalam model dan mudah diimplementasikan secara manual (Hastie dkk., 2009). Lain halnya pada penelitian Bayu (2023), LAR bekerja lebih cepat dan efisien dalam mencari solusi LASSO. Ketika dibandingkan dengan GLMNET, Algoritma LAR terbukti lebih unggul dalam mengatasi multikolinieritas karena mampu menghasilkan model dengan tingkat akurasi prediksi yang lebih baik, yang ditunjukkan oleh nilai koefisien determinasi yang lebih tinggi dan MSE yang lebih rendah (Bayu, 2023).

Regresi LASSO banyak diterapkan untuk mengetahui faktor-faktor yang memengaruhi variabel terikat namun mengalami multikolinieritas. Salah satu contoh penerapannya yaitu untuk mengetahui faktor-faktor yang diduga memengaruhi pengeluaran per kapita dalam konteks nasional atau global (Hasim, 2022). Meskipun banyak penelitian yang menggunakan regresi LASSO dalam konteks nasional atau global, penelitian yang berfokus pada daerah spesifik seperti Kabupaten Wonosobo masih sangat terbatas. Padahal, analisis pada daerah spesifik dapat memberikan wawasan baru tentang faktor-faktor yang memengaruhi pengeluaran per kapita. Kabupaten Wonosobo memiliki karakteristik geografis dan sosial ekonomi yang khas dibandingkan daerah lain di Provinsi Jawa Tengah. Sebagai daerah pegunungan dengan potensi di sektor pertanian dan pariwisata, Wonosobo menunjukkan pola pengeluaran per kapita yang

berbeda. Selain itu, tingkat pengeluaran per kapita di daerah ini tercatat lebih rendah dari rata-rata provinsi, sehingga penting untuk dianalisis lebih lanjut. Hasil dari penelitian ini diharapkan dapat menjadi dasar dalam perumusan kebijakan yang lebih tepat sasaran dan berkelanjutan guna meningkatkan kesejahteraan masyarakat setempat.

Kondisi pengeluaran per kapita di Kabupaten Wonosobo pada tahun 2015 tercatat Rp635.354 per bulan, kemudian meningkat menjadi Rp656.429 pada tahun 2016, Rp705.160 pada tahun 2017, dan Rp930.699 pada tahun 2018. Namun, di tahun 2019 mengalami penurunan menjadi Rp892.553 akibat kenaikan harga komoditas yang tidak diimbangi dengan peningkatan pendapatan dan inflasi yang tidak terkendali. Pada tahun 2020, pengeluaran per kapita kembali turun menjadi Rp833.995 akibat pandemi. Setelah pemulihan ekonomi, pengeluaran per kapita meningkat menjadi Rp923.840 pada tahun 2021, Rp987.422 pada tahun 2022, dan Rp1.080.574 pada tahun 2023. Namun, pada tahun 2024 pengeluaran per kapita kembali menurun karena meningkatnya tenaga kerja tanpa diiringi oleh cukupnya peluang kerja. Kondisi ini juga terkait dengan kualitas pendidikan tenaga kerja yang belum memadai. Pendidikan yang kurang relevan atau rendahnya tingkat keterampilan dan kompetensi lulusan menyebabkan banyak pekerja sulit bersaing di pasar kerja. Akibatnya, meskipun jumlah tenaga kerja bertambah, tingkat pengangguran juga meningkat karena lulusan pendidikan belum siap memenuhi kebutuhan pasar kerja. Hal ini menunjukkan bahwa pengeluaran per kapita di Kabupaten Wonosobo mengalami fluktuasi yang dipengaruhi oleh berbagai faktor ekonomi dan pendidikan (Badan Pusat Statistik, 2024).

Dengan adanya fluktuasi tersebut, maka di perlukan analisis lebih lanjut untuk mengetahui faktor-faktor yang memengaruhinya. Untuk itu, pengeluaran per kapita dijadikan sebagai variabel terikat, sedangkan variabel bebas mencakup Indeks Pembangunan Manusia (IPM), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Umur Harapan Hidup (UHH). Pemilihan variabel-variabel ini didasarkan pada informasi dari situs *website* BPS yang menunjukkan bahwa variabel-variabel tersebut diduga memengaruhi pengeluaran per kapita. Selain itu, penelitian sebelumnya yang dilakukan oleh Hasim (2022) juga mendukung bahwa variabel-variabel ini memiliki pengaruh signifikan. Oleh karena itu, penelitian ini menggunakan variabel-variabel tersebut untuk analisis lebih lanjut. Namun, setelah dilakukan pemeriksaan terhadap variabel-variabel bebas pada data pengeluaran per kapita Kabupaten Wonosobo tahun 2015–2024, ditemukan adanya multikolinieritas yang ditunjukkan oleh nilai VIF (*Variance Inflation Factor*) yang lebih dari 10. Kondisi ini menunjukkan bahwa beberapa variabel bebas memiliki korelasi yang cukup kuat satu sama lain, sehingga dapat memengaruhi keakuratan estimasi model. Oleh karena itu, untuk mengatasi permasalahan tersebut, penelitian ini menggunakan regresi LASSO dengan algoritma LAR.

Berdasarkan latar belakang yang telah diuraikan, maka peneliti mengambil judul **"Pemodelan Regresi *Least Absolute Shrinkage And Selection Operator* Dengan Algoritma *Least Angle Regression* Dalam Mengatasi Multikolinieritas Pada Pengeluaran Per Kapita Di Wonosobo"** dengan tujuan untuk membentuk model menggunakan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan algoritma *Least Angle*

Regression (LAR) dalam mengatasi multikolinieritas guna menganalisis faktor-faktor yang memengaruhi pengeluaran per kapita di Kabupaten Wonosobo. Proses analisis dilakukan dengan bantuan *software* RStudio dan sebagian langkah perhitungan juga dilakukan secara manual untuk memperjelas proses masuknya variabel dan pemodelan yang dihasilkan.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression* dalam mengatasi multikolinieritas pada pengeluaran per kapita di Kabupaten Wonosobo?
2. Apa saja faktor yang memengaruhi pengeluaran per kapita di Kabupaten Wonosobo menggunakan model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression* dalam mengatasi multikolinieritas?

C. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini sebagai berikut:

1. Memperoleh model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression* dalam mengatasi multikolinieritas pada pengeluaran per kapita di Kabupaten Wonosobo.

2. Mengetahui faktor-faktor yang memengaruhi pengeluaran per kapita di Kabupaten Wonosobo menggunakan model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression* dalam mengatasi multikolinieritas.

D. Manfaat Penelitian

Adapun manfaat penelitian ini sebagai berikut:

1. Manfaat bagi Peneliti

Mengenalkan konsep dasar dan pemaparan tentang model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression* dalam mengatasi multikolinieritas.

2. Manfaat bagi Jurusan

Menyediakan sumber tambahan dalam literatur dan referensi untuk pengajaran di bidang terkait.

3. Manfaat bagi Universitas

Menambah pemahaman Matematika Sains melalui penerapan model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression*.

4. Manfaat bagi Pembaca

Dapat memahami dan menguasai cara penerapan model regresi *Least Absolute Shrinkage and Selection Operator* dengan algoritma *Least Angle Regression* dalam mengatasi multikolinieritas pada analisis data.

5. Manfaat bagi Pemerintah Kabupaten Wonosobo

Memberikan kontribusi informasi yang mendukung kebijakan Pemerintah Kabupaten Wonosobo dalam menyusun kebijakan yang sesuai untuk meningkatkan kesejahteraan masyarakat, serta sebagai acuan dalam memberikan perhatian khusus terhadap faktor-faktor yang memengaruhi pengeluaran per kapita.

E. Batasan Masalah

Batasan masalah yang diterapkan dalam penelitian ini, adalah sebagai berikut:

1. Fokus utamanya adalah mengatasi multikolinieritas dalam pemodelan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan algoritma *Least Angle Regression* (LAR) dan tidak memperhatikan model lain.
2. Data yang digunakan dalam penelitian ini merupakan data sekunder yang sepenuhnya bersumber dari situs *website* BPS Kabupaten Wonosobo pada tahun 2015-2024 dengan variabel bebas yaitu faktor-faktor yang diduga memengaruhi pengeluaran per kapita di Kabupaten Wonosobo.

BAB II

LANDASAN TEORI

A. Regresi Linier Berganda

Regresi dapat dibagi menjadi dua berdasarkan jumlah variabel bebas yang digunakan, yaitu regresi linier sederhana dan berganda. Regresi linier berganda merupakan pengembangan dari regresi linier sederhana dengan menambahkan lebih dari satu variabel bebas. Pada kedua jenis regresi ini, istilah "linier" diasumsikan bahwa variabel terikat berhubungan secara langsung dengan hasil kombinasi linier dari variabel-variabel bebas (Tranmer dkk., 2020). Regresi linier berganda merupakan analisis statistik yang diterapkan untuk mengetahui adanya hubungan atau pengaruh antara lebih dari satu variabel bebas terhadap variabel terikat. Syarat-syarat regresi linier berganda, yaitu (Mubarak, 2021):

1. Jumlah sampel antara variabel terikat dan bebas yaitu sama.
2. Variabel terikat hanya ada satu yaitu variabel Y .
3. Nilai residual (selisih antara nilai aktual/sebenarnya dan nilai prediksi dari model) berdistribusi normal.
4. Seluruh asumsi klasik seperti normalitas, homogenitas, autokorelasi, dan multikolinieritas telah terpenuhi.

Ketika jumlah variabel bebas bertambah, bentuk dasar dari persamaan regresi linier berganda yang mencakup dua atau lebih

variabel bebas dituliskan sebagai berikut (Rencher, 2008):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.1)$$

Keterangan:

- Y_i : Variabel terikat untuk observasi ke- i .
- β_0 : Konstanta dalam model regresi.
- β_k : Koefisien regresi dari variabel bebas ke- k .
- X_{ik} : Nilai dari variabel bebas ke- k untuk observasi ke- i .
- ε_i : *Error* untuk observasi ke- i .
- $i : 1, 2, \dots, n$; n adalah banyaknya observasi.
- k : Jumlah variabel bebas dalam model regresi.

Maka model regresi linier berganda dapat ditulis dalam bentuk matriks sebagai berikut:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Estimasi parameter pada regresi linier berganda umumnya dilakukan dengan metode *Ordinary Least Squares* (OLS). Tujuan utama dari metode ini adalah untuk memperoleh nilai parameter β yang meminimalkan jumlah kuadrat *error* antara nilai observasi ($\mathbf{Y}$) dan nilai prediksi ($\mathbf{X}\beta$) (Hastie dkk., 2016). Oleh karena itu,

pendekatan OLS berfokus pada minimisasi fungsi kuadrat *error*, yang dapat dinyatakan dalam notasi matriks sebagai berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

Keterangan:

- $\mathbf{Y}$: Vektor variabel terikat berukuran $(n \times 1)$.
- $\mathbf{X}$: Matriks variabel bebas berukuran $(n \times (k + 1))$.
- $\boldsymbol{\beta}$: Vektor parameter regresi berukuran $((k + 1) \times 1)$.
- $\boldsymbol{\varepsilon}$: Vektor *error* yang berukuran $(n \times 1)$.

Jumlah kuadrat *error* dapat dinyatakan dalam notasi matriks sebagai $\sum \varepsilon_i^2 = \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$ dengan $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ sehingga diperoleh sebagai berikut (Aprianto dkk., 2020):

$$\begin{aligned} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} - (\mathbf{X}\boldsymbol{\beta})^\top \mathbf{Y} + (\mathbf{X}\boldsymbol{\beta})^\top (\mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^\top \mathbf{Y} - (\mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta})^\top - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \quad (2.3) \\ &= \mathbf{Y}^\top \mathbf{Y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}^\top \mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Untuk memastikan bahwa jumlah kuadrat *error* berada pada titik minimum, dilakukan proses penurunan terhadap parameter $\boldsymbol{\beta}$, kemudian hasil turunannya disamakan dengan nol. Hal ini didasarkan pada prinsip dasar kalkulus bahwa nilai minimum atau maksimum suatu fungsi dicapai ketika turunan pertamanya bernilai nol. Dalam regresi linier, pendekatan ini digunakan untuk menentukan nilai parameter $\boldsymbol{\beta}$ yang meminimalkan jumlah kuadrat *error*. Titik ketika turunan bernilai nol disebut sebagai

titik stasioner, yang merupakan kandidat nilai minimum dari fungsi tersebut dan di peroleh sebagai berikut:

$$\begin{aligned}
 & \frac{\partial(\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon})}{\partial \boldsymbol{\beta}^\top} = 0 \\
 \Leftrightarrow & \frac{\partial(\mathbf{Y}^\top \mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} = 0 \\
 \Leftrightarrow & -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} = 0 \\
 \Leftrightarrow & \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y} \\
 \Leftrightarrow & (\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{X})\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} \\
 \Leftrightarrow & \mathbf{I}\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} \\
 \Leftrightarrow & \boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} \tag{2.4}
 \end{aligned}$$

Maka diperoleh estimasi parameter OLS yaitu:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \tag{2.5}$$

B. Uji Multikolinieritas

Uji multikolinieritas bertujuan guna menunjukkan ada tidaknya hubungan linier antara satu atau lebih variabel bebas pada model regresi yang akan digunakan (Kadir, 2015). Ragnar Frisch merupakan tokoh yang pertama kali mengenalkan istilah multikolinieritas, yang secara luas diartikan sebagai kondisi dimana variabel bebas $(X_1, X_2, \dots, X_k)$ memiliki hubungan linier yang kuat (Mubarak, 2021). Ketika multikolinieritas terjadi dalam regresi, hal ini dapat menyebabkan berbagai permasalahan dalam analisis data. Berikut adalah dampak negatif dari multikolinieritas (Kumari, 2008):

1. Parameter yang diestimasi menjadi tidak signifikan.
2. Estimasi standar *error* untuk parameter akan besar.
3. Sulit untuk mengukur sejauh mana masing-masing variabel bebas dalam memengaruhi variabel terikat.
4. Menyebabkan peneliti mengeluarkan variabel yang penting dari model karena kesulitan dalam menilai apakah variabel tersebut memiliki pengaruh yang signifikan atau tidak terhadap variabel terikat.

Dengan demikian, untuk menghindari dampak negatif tersebut, penting untuk mendeteksi adanya multikolinieritas dalam model regresi melalui pengamatan terhadap nilai *Variance Inflation Factor* (VIF). Berikut adalah rumus VIF (Nisa & Maulina, 2024):

$$VIF_k = \frac{1}{1 - R_k^2} \quad (2.6)$$

dengan R_k^2 merupakan nilai besarnya koefisien determinasi ketika variabel bebas ke- k diregresikan terhadap variabel bebas lainnya. Koefisien determinasi berfungsi agar dapat menilai seberapa besar performa model regresi dalam menjelaskan variasi (perbedaan nilai data) dari hasil observasi. Semakin tinggi nilai R_k^2 atau semakin dekat nilainya dengan 1, maka semakin baik kemampuan variabel bebas dalam memberikan penjelasan hubungan dengan variabel terikat yang dianalisis. Berikut adalah karakteristik dari koefisien determinasi:

1. Nilai R_k^2 selalu positif, karena berasal dari jumlah kuadrat yang memiliki rumus (Mubarak, 2021):

$$R_k^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (2.7)$$

Keterangan:

- R_k^2 : Koefisien determinasi.
- SSR (*Sum of Squares for Regression*) : Jumlah kuadrat regresi (variasi yang dijelaskan model).
- SST (*Total Sum of Squares*) : Jumlah kuadrat total (variasi keseluruhan dalam data).
- SSE (*Sum of Squares for Error*) : Jumlah kuadrat *error* (variasi yang tidak dijelaskan model).

dengan:

$$\begin{aligned} SSR &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \\ SSE &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ SST &= SSE + SSR \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \end{aligned} \quad (2.8)$$

Keterangan:

- Y_i : Nilai aktual untuk observasi ke- i .
- $\hat{Y}_i$: Nilai prediksi untuk observasi ke- i .
- $\bar{Y}$: Rata-rata dari seluruh observasi Y_i .
- n : Jumlah data dalam dataset.

2. Nilai $0 \leq R_k^2 \leq 1$.

- Jika $R_k^2 = 0$, berarti tidak ada hubungan antara X dan Y atau model tidak dapat menjelaskan Y dengan baik.
- Jika $R_k^2 = 1$, model yang terbentuk dapat menjelaskan Y dengan baik.

Untuk memeriksa ada tidaknya multikolinieritas dalam model, dapat digunakan nilai *Variance Inflation Factor* (VIF). Apabila nilai VIF lebih dari 10, maka dapat disimpulkan terdapat multikolinieritas antar variabel bebas (Mubarak, 2021).

C. Standarisasi Data

Standarisasi data adalah proses yang digunakan untuk menyetarakan variabel yang memiliki satuan atau skala dengan pengukuran berbeda, agar seluruh variabel berada pada skala yang sama dan dapat dianalisis secara setara tanpa ada yang lebih menonjol hanya karena perbedaan satuan (Wahidah dkk., 2023). Dalam suatu dataset, sering terdapat variabel dengan satuan atau skala pengukuran yang berbeda-beda. Misalnya, satu variabel diukur dalam jutaan rupiah, sedangkan variabel lainnya diukur dalam persen atau tahun. Perbedaan skala ini dapat menyebabkan ketidakseimbangan dalam analisis karena variabel dengan nilai yang lebih besar cenderung mendominasi atau memberikan pengaruh lebih besar terhadap model, bukan karena memang lebih penting, melainkan karena skala angkanya lebih tinggi. Untuk menghindari bias semacam ini, maka dilakukan standarisasi data. Salah satu metode yang digunakan untuk standarisasi data adalah *Z-Score*, yang dapat dihitung menggunakan rumus berikut (Wahidah dkk., 2023):

1. Menghitung Rata-rata

Berikut adalah rumus rata-rata yang digunakan (Wahidah dkk., 2023):

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad (2.9)$$

Keterangan:

- $\bar{Y}$: Rata-rata sampel variabel Y .
- Y_i : Nilai untuk observasi ke- i .
- n : Banyaknya jumlah observasi.

2. Menghitung Standar Deviasi atau Simpangan Baku

Standar deviasi merupakan ukuran statistik yang menunjukkan tingkat penyebaran data terhadap nilai rata-rata (mean). Rumus standar deviasi (Wahidah dkk., 2023):

$$S = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}} \quad (2.10)$$

Keterangan:

- S : Standar deviasi.
- Y_i : Nilai untuk observasi ke- i .
- $\bar{Y}$: Rata-rata sampel variabel Y .

3. Menghitung Z-Score

Z-Score merupakan nilai baku atau nilai standar dari suatu data. Berikut rumus *Z-Score* (Wahidah dkk., 2023):

$$z_i = \frac{Y_i - \bar{Y}}{S} \quad (2.11)$$

Keterangan:

- z_i : Nilai *Z-Score* untuk observasi ke- i .
- Y_i : Nilai untuk observasi ke- i .
- $\bar{Y}$: rata-rata sampel variabel Y .
- S : Standar deviasi.

D. Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO)

Tibshirani merupakan tokoh yang memperkenalkan regresi LASSO untuk pertama kalinya pada tahun 1996 untuk mengatasi masalah multikolinieritas, yaitu ketika kondisi variabel bebas saling berhubungan. Model regresi ini bekerja dengan menyusutkan (*shrinkage*) nilai koefisien regresi LASSO menjadi bernilai nol atau mendekati nol, sehingga LASSO secara otomatis dapat menyeleksi variabel bebas yang tidak memiliki hubungan signifikan dengan variabel terikat (Tibshirani, 1996). Dengan kemampuannya dalam seleksi variabel, model ini dapat diterapkan pada data berdimensi besar dengan tujuan untuk meningkatkan akurasi atau memperbaiki estimasi pada regresi linier berganda (Tibshirani, 1996)

Model regresi linier berganda secara umum dinyatakan pada Persamaan (2.1) untuk mengestimasi parameter β , metode yang digunakan yaitu *Ordinary Least Squares* (OLS), karena OLS bertujuan untuk meminimalkan jumlah kuadrat *error* antara nilai observasi dengan nilai yang diprediksi (Hastie dkk., 2016).

Jumlah kuadrat *error* (SSE) dalam regresi linier dapat dinyatakan dalam Persamaan (2.9) dan telah dinyatakan dalam

bentuk notasi matriks pada Persamaan (2.3). Selanjutnya, bentuk tersebut dikembangkan dan diturunkan untuk memperoleh estimasi parameter regresi dengan OLS seperti yang ditunjukkan pada Persamaan (2.4). Namun, ketika jumlah variabel bebas k besar, terutama jika $k > n$ (jumlah variabel bebas lebih banyak dari jumlah observasi), metode OLS tidak dapat menghasilkan estimasi parameter yang sesuai. Dalam kondisi tersebut, estimasi parameter menjadi tidak stabil dan tidak dapat diandalkan karena model mengalami multikolinieritas (Hastie dkk., 2016). Contohnya terlihat pada dataset mikroarray kanker kolorektal yang digunakan dalam penelitian Alon dkk., (1999), yang juga dijadikan studi kasus oleh Tibshirani (1996). Dataset tersebut terdiri dari sekitar $k = 2000$ gen (variabel bebas) dan hanya $n = 62$ jumlah observasi pasien. Dalam kondisi seperti ini, OLS gagal diterapkan, sehingga LASSO menjadi solusi yang lebih tepat karena kemampuannya dalam melakukan seleksi variabel dan estimasi parameter secara keseluruhan pada data berdimensi tinggi.

Oleh karena itu, regresi LASSO digunakan sebagai solusi untuk mengatasi permasalahan tersebut. Berikut adalah nilai estimasi regresi LASSO (Tibshirani, 1996):

$$\hat{\beta}^{\text{lasso}} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{k=1}^p \beta_k X_{ik} \right)^2 \right\} \quad (2.12)$$

dengan syarat penalti $\sum_{k=1}^p |\beta_k| \leq t$. Simbol $\arg$ merupakan singkatan dari “*argument of*”, yaitu nilai input (β) yang menyebabkan suatu fungsi mencapai nilai minimum atau maksimum (nilai ekstrem).

Keterangan:

- Y_i : Variabel terikat untuk observasi ke- i .
- β_0 : Konstanta dalam model regresi.
- β_k : Koefisien regresi dari variabel bebas ke- k .
- X_{ik} : Nilai variabel bebas ke- k untuk observasi ke- i .
- $i : 1, 2, \dots, n$; n adalah banyaknya observasi.
- $k : 1, 2, \dots, p$; p adalah banyaknya variabel bebas.

Menurut Tibshirani (1996), t merupakan parameter *tuning* yang digunakan untuk mengontrol besarnya penyusutan (*shrinkage*) pada koefisien regresi LASSO. Parameter t berperan sebagai batas total penyusutan koefisien dengan memberikan kendala agar jumlah nilai absolut dari semua koefisien regresi tidak melebihi t , dengan syarat: $t \geq 0$. Dengan demikian:

1. Jika nilai $t < t_0$, maka koefisien OLS akan menyusut mendekati nol atau tepat pada nol. Hal ini merupakan sifat yang diinginkan dari regresi LASSO, karena memungkinkan model untuk melakukan seleksi variabel dengan mengeliminasi variabel-variabel yang kurang relevan terhadap variabel terikat.
2. Jika nilai $t \geq t_0$, maka koefisien regresi LASSO memberikan hasil yang sama dengan koefisien OLS.

Nilai t dapat dituliskan sebagai $t = \sum_{k=1}^p |\hat{\beta}_k|$, sedangkan nilai t_0 yaitu $t_0 = \sum_{k=1}^p |\hat{\beta}_k^0|$ dimana $\hat{\beta}_k^0$ adalah koefisien regresi dengan menggunakan OLS.

Kemudian, koefisien dalam regresi LASSO ditetapkan berdasarkan parameter *tuning* yang telah dinormalisasi yaitu s :

$$s = \frac{t}{t_0} = \frac{\sum_{k=1}^p |\hat{\beta}_k|}{\sum_{k=1}^p |\hat{\beta}_k^0|} = \frac{\sum_{k=1}^p |\hat{\beta}_k^{\text{lasso}}|}{\sum_{k=1}^p |\hat{\beta}_k^{\text{ols}}|} \quad (2.13)$$

Nilai s digunakan untuk menunjukkan besarnya penalti (hukuman) yang diterapkan pada model. Nilai s yang optimal biasanya didapatkan dari proses validasi silang (Dewi, 2010).

E. Algoritma *Least Angle Regression* (LAR)

Algoritma *Least Angle Regression* (LAR) merupakan metode analitik yang dikembangkan oleh Efron dkk, (2004) untuk menentukan urutan masuknya variabel ke dalam model secara bertahap, berdasarkan korelasi terbesar dengan residual. Sementara itu, rumus regresi LASSO sulit diselesaikan secara analitik karena mengandung $|\beta_k|$ yang tidak diferensiabel. Oleh karena itu, LAR yang dimodifikasi digunakan sebagai pendekatan numerik untuk menghitung koefisien regresi LASSO. Dengan mengikuti jalur solusi secara linier, algoritma ini mampu menghasilkan nilai koefisien regresi yang identik dengan solusi LASSO. Dengan demikian, LASSO dan LAR memiliki keterkaitan erat dalam penyelesaian model regresi.

Menurut Efron dkk, (2004) algoritma LAR melakukan estimasi $\hat{\mu} = X\hat{\beta}$ secara bertahap dengan menambahkan satu variabel bebas ke dalam model pada setiap langkah. Sebelumnya, semua koefisien regresi diatur menjadi nol, yaitu $\beta_k = 0, \forall k$. Karena dalam regresi linier, prediksi diberikan oleh $\hat{\mu} = X\hat{\beta}$, maka ketika semua koefisien bernilai nol, hasil prediksi juga menjadi nol, yaitu

$\hat{\mu} = X(0) = 0$. Oleh karena itu, algoritma ini memulai proses dengan nilai awal $\hat{\mu}_0 = 0$ dan secara bertahap menambahkan variabel ke dalam model. Dengan seleksi variabel secara bertahap, algoritma ini dapat dimodifikasi untuk memperoleh estimasi koefisien LASSO ($\hat{\beta}^{\text{lasso}}$) sebagai berikut (Efron dkk., 2004):

1. Mencari suatu vektor yang sebanding dengan vektor korelasi antara variabel bebas dan residual dari setiap variabel bebas, yaitu:

$$\hat{c} = X^{\top}(Y - \hat{\mu}) \quad (2.14)$$

Keterangan:

- $\hat{c}$: Vektor korelasi.
- $X^{\top}$: *Transpose* dari matriks X .
- Y : Vektor variabel terikat.
- $\hat{\mu}$: Vektor prediksi model.

2. Menentukan nilai korelasi terbesar dari vektor korelasi, melalui:

$$\hat{C} = \max \{|\hat{c}_k|\} \quad (2.15)$$

sehingga diperoleh $s_k = \text{sign} \{\hat{c}_k\}$ untuk $k \in A$. Dengan fungsi tanda (*sign*) didefinisikan sebagai:

$$\text{sign} \{\hat{c}_k\} = \begin{cases} +1, & \hat{c}_k > 0 \\ 0, & \hat{c}_k = 0 \\ -1, & \hat{c}_k < 0 \end{cases}$$

Karena nilai mutlak $|\hat{c}_k|$ selalu bernilai nol atau positif,

akibatnya tanda absolut dari s_k adalah:

$$|s_k| = 1, \quad \text{untuk semua } \hat{c}_k \neq 0.$$

Artinya, meskipun $\hat{c}_k$ negatif, nilai mutlaknya tetap positif, tetapi arah perubahan koefisien tetap mengikuti tanda s_k .
Keterangan:

- $\hat{C}$: Nilai korelasi absolut terbesar.
 - s_k : Tanda dari koefisien (β_k) .
 - k : Indeks yang menunjukkan variabel bebas ke- k .
 - A : Kumpulan indeks dari variabel aktif yang dipilih dalam model.
3. Menentukan $\mathbf{X}_A$, dimana A adalah kumpulan indeks dari variabel aktif. Variabel aktif yaitu variabel bebas yang saat ini sedang digunakan dalam model karena memiliki nilai korelasi terbesar. Variabel-variabel tersebut dikumpulkan ke dalam sebuah submatriks yang disebut $\mathbf{X}_A$, didefinisikan matriks sebagai:

$$\mathbf{X}_A = (\dots s_k \mathbf{X}_k \dots)_{k \in A}$$

Submatriks $\mathbf{X}_A$ dibentuk dari kumpulan variabel bebas aktif, yaitu kolom-kolom dari matriks $\mathbf{X}$ yang indeksnya termasuk dalam himpunan A , dimana masing-masing kolom tersebut telah dikalikan dengan tanda s_k . Tanda s_k menunjukkan arah dari koefisien regresi, yang nilainya dapat berupa $+1$ atau -1 , tergantung hubungannya positif atau negatif. Kemudian,

hitung:

$$G_A = X_A^\top X_A \quad \text{dan} \quad A_A = (\mathbf{1}_A^\top G_A^{-1} \mathbf{1}_A)^{-\frac{1}{2}} \quad (2.16)$$

dengan $\mathbf{1}_A$ adalah vektor yang elemennya semua 1 dan panjangnya sama dengan jumlah variabel yang dipilih dalam himpunan variabel aktif A .

4. Menghitung nilai vektor *equiangular*, vektor *equiangular* merupakan vektor satuan yang membentuk sudut yang sama dengan setiap variabel aktif yang sedang digunakan dalam model saat ini. Vektor ini digunakan untuk menentukan arah perubahan koefisien regresi saat menambahkan variabel ke dalam model, dengan cara memastikan bahwa sudut antara kolom-kolom pada X_A memiliki besar yang sama dan membentuk sudut kurang dari 90° . Nilai vektor *equiangular* dicari dengan rumus sebagai berikut:

$$u_A = X_A w_A \quad (2.17)$$

dimana:

$$w_A = A_A G_A^{-1} \mathbf{1}_A \quad (2.18)$$

Keterangan:

- u_A : Vektor *equiangular*.
- X_A : Submatriks variabel aktif.
- w_A : Vektor bobot *equiangular*.

5. Menghitung vektor *inner product*, vektor *inner product* digunakan untuk mengukur korelasi antara residual dan variabel-variabel bebas aktif selama proses LAR. Selama

proses LAR, vektor residual dihitung pada setiap langkah untuk menentukan arah perubahan koefisien regresi. Dapat dihitung dengan rumus sebagai berikut:

$$\mathbf{a} \equiv \mathbf{X}^\top \mathbf{u}_A \quad (2.19)$$

Keterangan:

- $\mathbf{a}$: Hasil vektor *inner product*.
- $\mathbf{X}^\top$: *Transpose* dari matriks $\mathbf{X}$.
- $\mathbf{u}_A$: Vektor *equiangular*.

6. Menghitung $\hat{\mu}_{A+}$ dengan menggunakan rumus sebagai berikut:

$$\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma} \mathbf{u}_A \quad \text{dan} \quad A_+ = A - \{k\} \quad (2.20)$$

Keterangan:

- $\hat{\mu}_{A+}$: Vektor prediksi model setelah diperbarui.
- $\hat{\mu}_A$: Vektor prediksi model sebelum diperbarui.
- $\hat{\gamma}$: Langkah optimal.
- $\mathbf{u}_A$: Vektor *equiangular*.
- A_+ : Kumpulan indeks dari variabel aktif setelah diperbarui.
- A : Kumpulan indeks dari variabel aktif.
- $\{k\}$: Variabel k yang dikeluarkan dari himpunan variabel aktif.

Jadi $\hat{\gamma}$ dapat diperoleh dengan persamaan sebagai berikut:

$$\hat{\gamma} = \min_{k \in A^c}^+ \left\{ \frac{\hat{C} - \hat{c}_k}{A_A - \mathbf{a}_k}, \frac{\hat{C} + \hat{c}_k}{A_A + \mathbf{a}_k} \right\} \quad (2.21)$$

Untuk $\min_{k \in A^c}^+$ menunjukkan bahwa nilai positif terkecil yang dipilih dari γ_k , dihitung untuk setiap variabel k yang belum termasuk dalam himpunan aktif A . Himpunan dari A berisi indeks variabel aktif yang telah masuk ke dalam model, sedangkan A^c berisi indeks variabel yang belum masuk ke dalam model. Jadi, fokus perhitungan hanya pada indeks $k \in A^c$, karena hanya variabel-variabel tersebut yang sedang dievaluasi untuk kemungkinan masuk ke dalam model. Hal utama yang ingin dicapai dari proses ini adalah menentukan variabel baru yang paling memenuhi syarat untuk bergabung ke dalam model.

7. Algoritma LAR untuk LASSO mengharuskan tanda dari koordinat bukan nol $\hat{\beta}_k$ sama dengan tanda $\hat{c}_k$ dengan rumus berikut:

$$\text{sign}(\hat{\beta}_k) = \text{sign}(\hat{c}_k) = s_k \quad (2.22)$$

Jika $\hat{c}_k > 0$, maka $\hat{\beta}_k$ harus bernilai positif, sedangkan jika $\hat{c}_k < 0$, maka $\hat{\beta}_k$ harus bernilai negatif. Ketentuan ini bertujuan untuk memastikan bahwa setiap variabel yang masuk ke dalam model (variabel aktif) memiliki koefisien dengan tanda yang sama seperti korelasinya terhadap residual, yaitu $\hat{c}_k$. Dengan demikian, koefisien dari variabel aktif tidak boleh berubah tanda selama proses berlangsung agar kestabilan model tetap terjaga.

8. Langkah ini digunakan ketika terdapat potensi perubahan tanda pada salah satu koefisien variabel aktif pada langkah ke tujuh. Hal ini terjadi jika nilai $\tilde{\gamma}$, yaitu langkah dimana suatu koefisien $\beta_k(\gamma)$ menjadi nol, lebih kecil dari langkah optimal $\hat{\gamma}$ yang digunakan untuk menambahkan variabel baru ke

dalam model. Nilai $\tilde{\gamma}$ dihitung menggunakan rumus:

$$\tilde{\gamma} = \min_{\gamma_k > 0} \left\{ \frac{-\hat{\beta}_k}{d_k} \right\} \quad (2.23)$$

dengan $d_k = s_k \mathbf{w}_{Ak}$, dimana d_k merupakan elemen ke- k dari vektor arah perubahan koefisien $\mathbf{d}$. Elemen tersebut menunjukkan perubahan koefisien hanya untuk variabel ke- k , dengan $k \in A$. Jika $\tilde{\gamma} < \hat{\gamma}$, maka salah satu koefisien akan menjadi nol dan berpotensi berubah tanda. Karena perubahan tanda tidak diperbolehkan dalam LASSO, proses harus dihentikan. Setelah proses dihentikan, model yang diperoleh dianggap sebagai model akhir dari regresi LASSO. Pada langkah ini, variabel yang koefisiennya menjadi nol akan dihapus dari himpunan variabel aktif dan tidak lagi diperhitungkan dalam perhitungan vektor *equiangular* berikutnya. Namun, variabel tersebut masih dapat dimasukkan kembali dalam perhitungan LAR pada langkah-langkah selanjutnya apabila kembali memenuhi syarat untuk masuk ke dalam model.

9. Menghitung calon koefisien LASSO, merupakan nilai koefisien regresi yang diperoleh pada setiap langkah dalam proses algoritma LAR. Calon koefisien LASSO dapat dihitung dengan rumus:

$$\beta_k^{(t+1)} = \beta_k^{(t)} + \hat{\gamma} \times d_k \quad (2.24)$$

Keterangan:

- $\beta_k^{(t)}$: Koefisien regresi LASSO untuk variabel ke- k pada

langkah ke- t .

- $\beta_k^{(t+1)}$: Koefisien regresi LASSO untuk variabel ke- k setelah diperbarui pada langkah ke- $t + 1$.
- $\hat{\gamma}$: Langkah optimal.
- d_k : Komponen arah perubahan koefisien untuk variabel ke- k .

10. Proses ini terus diulang dengan memilih dan memperbarui variabel secara bertahap hingga semua variabel yang relevan telah dimasukkan ke dalam model atau tidak ada lagi variabel yang bisa dipilih. Dengan demikian, langkah pemilihan variabel dalam model telah selesai.

F. Validasi Silang

Validasi silang atau yang biasa disebut dengan *cross validation* merupakan metode yang digunakan untuk mengestimasi nilai *Mean Squared Error* (MSE) guna meningkatkan ketepatan dalam pemilihan model yang terbaik. MSE merupakan ukuran rata-rata dari kuadrat selisih antara hasil prediksi model dan nilai observasi pada data untuk menilai seberapa besar kesalahan prediksi yang dihasilkan oleh model (Amansyah dkk., 2024). Berikut adalah estimasi validasi silang (James dkk., 2013):

$$CV = \frac{1}{n} \sum_{i=1}^n MSE_i \quad (2.25)$$

dengan rumus MSE (Kutner dkk., 2005) sebagai berikut:

$$MSE_i = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - 1} \quad (2.26)$$

Keterangan:

- CV : *Cross Validation*.
- n : Jumlah total observasi.
- MSE_i : *Mean Squared Error* untuk observasi ke- i .
- Y_i : Nilai aktual untuk observasi ke- i .
- $\hat{Y}_i$: Nilai prediksi dari model untuk observasi ke- i .

Salah satu metode yang sering digunakan dalam validasi silang yaitu *K-fold*. *K-fold* adalah metode validasi silang yang membagi data menjadi c bagian (*fold*) yang kira-kira sama besar. Proses validasi silang *K-fold* akan menghasilkan sebanyak c perkiraan kesalahan prediksi, yaitu $MSE_1, MSE_2, \dots, MSE_c$. Pemilihan jumlah *fold* (c) yang umum digunakan dalam *K-fold cross-validation* biasanya adalah $c = 5$ atau $c = 10$ (Hastie dkk., 2009).

Dalam regresi LASSO, model terbaik diperoleh berdasarkan nilai *Mean Squared Error* (MSE) minimum, karena nilai MSE yang mendekati nol menunjukkan bahwa model memiliki tingkat akurasi prediksi yang baik dan menentukan parameter s yang optimal melalui teknik validasi silang. Teknik ini bertujuan untuk memilih nilai s yang paling optimal dalam meningkatkan akurasi prediksi (Jamco dkk., 2023).

G. Pengeluaran Per Kapita

Pengeluaran per kapita mencerminkan seberapa besar rata-rata pengeluaran setiap orang dan digunakan sebagai ukuran standar hidup. Besarnya pengeluaran ini dipengaruhi oleh

tingkat pengetahuan dan kesempatan yang dimiliki seseorang untuk menghasilkan barang atau jasa, yang pada akhirnya menjadi sumber penghasilan (Mahrany, 2012). Secara matematis, pengeluaran per kapita dapat dituliskan sebagai berikut:

$$\text{Pengeluaran per kapita} = \frac{\text{Total pengeluaran}}{\text{Jumlah penduduk}} \quad (2.27)$$

Keterangan:

- Total pengeluaran adalah jumlah keseluruhan uang yang dihabiskan oleh populasi dalam jangka waktu tertentu.
- Jumlah penduduk adalah total jumlah individu yang termasuk dalam populasi tersebut.

Pengeluaran per kapita ditentukan oleh berbagai faktor, sehingga penting untuk memahami faktor-faktor tersebut bagi pembuat kebijakan dan ekonomi untuk membuat keputusan yang dapat meningkatkan kesejahteraan ekonomi masyarakat. Berikut ini merupakan beberapa faktor yang diduga berpengaruh terhadap pengeluaran per kapita (Badan Pusat Statistik, 2024):

1. Indeks Pembangunan Manusia (IPM)

IPM berfungsi sebagai indikator yang menggambarkan tingkat pencapaian masyarakat dalam menikmati hasil pembangunan di bidang pendapatan, kesehatan, pendidikan, serta aspek lainnya. Selain itu, IPM juga digunakan untuk menilai keberhasilan dalam meningkatkan taraf tingkat kesejahteraan masyarakat atau penduduk.

2. Rata-rata Lama Sekolah (RLS)

RLS merupakan rata-rata masa belajar yang telah dijalani oleh penduduk usia 15 tahun ke atas hingga tingkat pendidikan terakhir. Indikator ini digunakan untuk menilai tingkat dan kualitas pendidikan masyarakat di suatu daerah. Rendahnya RLS biasanya dipengaruhi oleh tingginya angka putus sekolah, yang salah satunya disebabkan oleh ketidakmampuan membayar biaya pendidikan.

3. Harapan Lama Sekolah (HLS)

HLS merupakan perkiraan periode waktu dalam menempuh pendidikan oleh seorang anak pada usia tertentu di masa depan. HLS penting untuk menunjukkan sejauh mana sistem pendidikan memberi kesempatan belajar di berbagai jenjang dan mencerminkan seberapa lama pendidikan yang diharapkan bisa dicapai oleh setiap anak.

4. Umur Harapan Hidup (UHH)

UHH menggambarkan rata-rata jumlah tahun yang diperkirakan masih dapat dijalani oleh seseorang. UHH juga dapat diartikan sebagai estimasi rata-rata usia tambahan yang akan dijalani oleh seseorang setelah mencapai usia tertentu, misalnya usia ke x .

H. Penelitian Terdahulu

Peneliti akan mengkaji sejumlah penelitian sebelumnya yang memiliki keterkaitan dengan topik penelitian. Hal ini bertujuan sebagai referensi dan bahan perbandingan guna memperoleh informasi yang relevan dengan judul penelitian. Adapun beberapa

penelitian sebelumnya yang berhubungan dengan penelitian ini antara lain sebagai berikut:

1. Penelitian yang dilakukan oleh Andana, dkk (2017) dengan judul “Model Regresi Menggunakan *Least Absolute Shrinkage And Selection Operator* (LASSO) Pada Data Banyaknya Gizi Buruk Kabupaten/Kota Di Jawa Tengah”. Penelitian ini bertujuan untuk mengatasi permasalahan multikolinieritas pada data yang berkaitan dengan faktor-faktor penyebab gizi buruk. Regresi LASSO digunakan untuk mereduksi atau mengecilkan nilai koefisien dari variabel bebas yang memiliki tingkat korelasi tinggi dihitung menggunakan algoritma LAR karena lebih efisien. Berdasarkan analisis yang telah dilakukan, diperoleh pada tahap kedua saat nilai s (besarnya penalti/hukuman) = 0,02 dengan nilai MSE (*Mean Squared Error*) sebesar 0,82977. Dapat disimpulkan bahwa variabel bayi (0-6 bulan) yang diberikan ASI eksklusif, rumah tangga yang menerapkan perilaku hidup bersih dan sehat, bayi yang menerima imunisasi Hepatitis B, bayi yang mendapatkan imunisasi DPT-HB3, rumah dengan sanitasi yang memadai, serta rumah dengan air minum yang memenuhi syarat kesehatan ini berpengaruh terhadap gizi buruk pada bayi di Jawa Tengah pada 2014.
2. Penelitian yang dilakukan oleh Jamco, dkk (2023) dengan judul “Penanganan Multikolinearitas Pada Regresi Linier Berganda Menggunakan Regresi LASSO (Studi Kasus: Distribusi Presentase Produk Domestik Regional Bruto Di Provinsi Maluku Tahun 1999-2021)”. Menyatakan bahwa pertumbuhan ekonomi menjadi salah satu indikator

utama dalam mengevaluasi atau menilai keadaan suatu perekonomian suatu negara atau wilayah. Dalam penelitian ini, pertumbuhan ekonomi diukur menggunakan Produk Domestik Regional Bruto (PDRB), yang rentan terhadap masalah asumsi klasik seperti multikolinieritas. Untuk mengatasi permasalahan tersebut, digunakan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO), karena dapat menyusutkan koefisien hingga mendekati atau tepat nol sekaligus melakukan seleksi variabel bebas. Berdasarkan analisis *Variance Inflation Factor* (VIF), ditemukan adanya multikolinieritas pada data pertumbuhan ekonomi Provinsi Maluku 1999-2021 jika menggunakan *Ordinary Least Square* (OLS). Namun, variabel bebas tersebut dapat dikecilkan dengan LASSO bahkan koefisiennya bisa sampai nol, menandakan bahwa variabel tersebut tidak memberikan dampak yang signifikansi terhadap model.

3. Penelitian yang dilakukan oleh Pardede, dkk (2022) dengan judul “Penerapan Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk Mengidentifikasi Variabel yang Berpengaruh terhadap Kejadian Stunting di Indonesia”. Menyatakan bahwa pada data stunting di Indonesia mengalami multikolinieritas sehingga menggunakan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan algoritma *Least Angle Regression* (LAR) untuk mengatasinya. LASSO dapat menyusutkan koefisien variabel yang berkorelasi tinggi hingga nol. Hasil analisisnya menunjukkan bahwa variabel ASI eksklusif (X_1), konsumsi protein (X_2), imunisasi DPT-HB (X_5), tinggi badan ibu (X_8), dan diare (X_9) berpengaruh terhadap stunting di Indonesia

tahun 2018.

4. Penelitian yang dilakukan oleh Rachmawati & Sari (2021) dengan judul "*LASSO Regression for Daily Rainfall Modeling at Citeko Station, Bogor, Indonesia*". Penelitian ini memodelkan intensitas curah hujan harian di stasiun pengamatan Citeko, Bogor dengan 16 variabel bebas yang diasumsikan berdistribusi normal. Hasil analisisnya menunjukkan bahwa regresi LASSO berhasil menyusutkan variabel bebas menjadi 9 variabel dan hasilnya lebih memuaskan dibandingkan regresi linier berganda. Dengan nilai AIC masing-masing 5901,20 untuk LASSO dan 5911,43 untuk regresi linier berganda. Selain itu, koefisien determinasi regresi LASSO lebih tinggi, yaitu 22,73%, dibandingkan 21,41% pada regresi linier berganda. Arah angin yang paling sering bertiup ke utara ditemukan sebagai variabel dengan pengaruh terbesar terhadap intensitas curah hujan.
5. Penelitian yang dilakukan oleh Kato & Uemura (2012) dengan judul "*Period Analysis using the Least Absolute Shrinkage and Selection Operator (LASSO)*". Menyatakan bahwa regresi LASSO dipilih karena model dapat mengestimasi hingga nol. Kemudian diimplementasikan menggunakan algoritma *Least Angle Regression* (LAR) dengan bantuan R. Penelitian ini, menguji dataset buatan untuk mengevaluasi efektivitas dalam mendeteksi sinyal periodik. Hasilnya dapat secara akurat memperkirakan periode meskipun masih terdapat noise/gangguan. Kemudian dilakukan analisis yang menunjukkan bahwa regresi LASSO mampu mengatasi hal tersebut, dan untuk

mengoptimalkanya penelitian ini menggunakan validasi silang untuk memilih parameter yang paling tepat, sehingga model dapat bekerja dengan baik.

BAB III

METODE PENELITIAN

A. Jenis Penelitian

Jenis penelitian yang digunakan yaitu kuantitatif. Penelitian kuantitatif merupakan penelitian yang menggunakan angka-angka untuk mengumpulkan dan menganalisis data. Dalam penelitian ini, data diolah dengan statistik untuk mendapatkan hasil yang jelas (Siyoto & Sodik, 2015). Selain itu, menurut Abubakar (2021) penelitian kuantitatif merupakan jenis penelitian yang mengumpulkan data dalam bentuk angka atau data kualitatif yang telah dikonversi menjadi data numerik.

B. Sumber Data

Sumber data yang digunakan dalam penelitian ini berasal dari data sekunder. Data sekunder merupakan informasi yang dikumpulkan dan disajikan oleh pihak lain dengan menggunakan metode yang berbeda, serta diperoleh secara tidak langsung melalui lembaga atau sumber lain yang memiliki keterkaitan dengan topik penelitian (Hidayatullah, 2015). Penelitian ini menggunakan data sekunder secara keseluruhan, yang diperoleh dari instansi resmi yaitu Badan Pusat Statistik (BPS) Kabupaten Wonosobo. Data tersebut telah dipublikasikan dan dapat diakses melalui link <https://wonosobokab.bps.go.id/>. Data yang digunakan dalam penelitian ini mencakup pengeluaran per kapita di Kabupaten Wonosobo beserta faktor-faktor yang diduga memengaruhinya yang telah terlampir pada Lampiran 1.

C. Variabel Penelitian

Variabel penelitian merupakan sebuah atribut atau karakteristik dari seseorang atau objek yang berbeda-beda, variabel ini dipilih oleh peneliti untuk diteliti, lalu digunakan sebagai dasar untuk menarik kesimpulan yang bersifat umum. Hasil dari penelitian ini dapat memberi manfaat secara teori, praktik atau metode dalam bidang yang diteliti (Abubakar, 2021). Pada penelitian ini menggunakan dua variabel yaitu variabel terikat (Y) dan variabel bebas (X). Adapun variabel-variabel yang digunakan adalah sebagai berikut:

Tabel 3.1. Variabel Penelitian

Variabel	Simbol	Satuan
Pengeluaran Per Kapita	Y	Rupiah
Indeks Pembangunan Manusia	X_1	Persen
Rata-rata Lama Sekolah	X_2	Tahun
Harapan Lama Sekolah	X_3	Tahun
Umur Harapan Hidup	X_4	Tahun

D. Analisis Data

Data pada penelitian ini dianalisis menggunakan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan algoritma *Least Angle Regression* (LAR). Adapun tahapan analisis data yang akan dilakukan yaitu sebagai berikut:

1. Memasukkan data.
2. Membentuk model awal menggunakan regresi linier berganda.
3. Melakukan uji multikolinieritas dan mengidentifikasinya, jika tidak terjadi multikolinieritas dilanjutkan dengan uji normalitas, homoskedastisitas dan autokorelasi, serta uji signifikansi parameter (simultan dan parsial). Selanjutnya, uji koefisien determinasi dan menginterpretasikan hasilnya. Namun, jika terjadi multikolinieritas, maka dilanjutkan dengan standarisasi data.
4. Membentuk model menggunakan regresi LASSO dengan algoritma LAR.
5. Melakukan uji validasi silang.
6. Membentuk model terbaik LASSO dengan algoritma LAR.
7. Menguji koefisien determinasi berdasarkan model terbaik.
8. Membandingkan model awal yang menggunakan regresi linier berganda dengan model terbaik LASSO.
9. Membandingkan nilai VIF sebelum dan setelah menggunakan regresi LASSO.
10. Menginterpretasikan hasil yang telah diperoleh.

E. Alur Penelitian

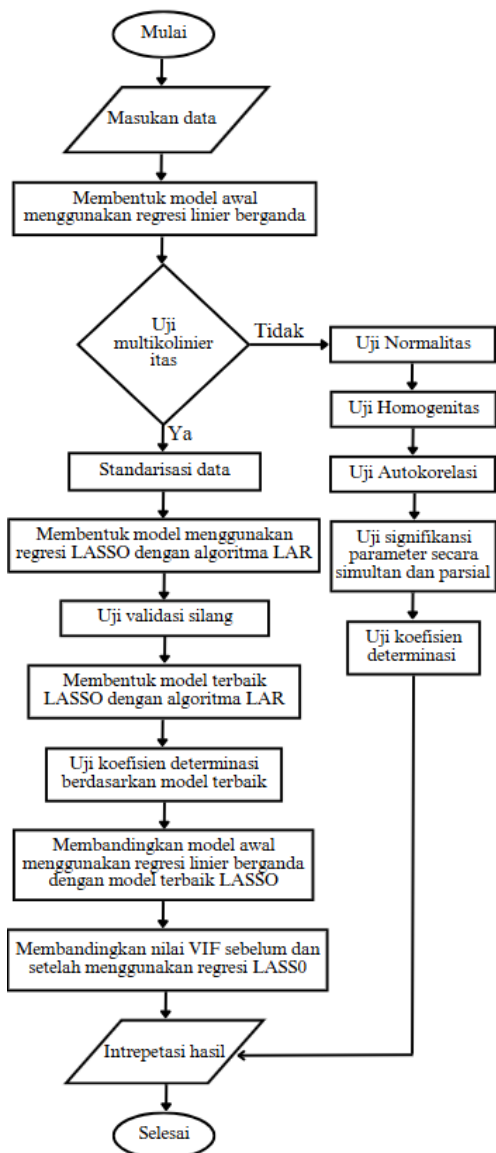
Alur penelitian ini diawali dengan memasukkan data dari sumber yang telah diperoleh. Data yang digunakan mencakup

pengeluaran per kapita sebagai variabel terikat dan variabel bebas yaitu faktor-faktor yang di duga mempengaruhinya. Kemudian membentuk model awal menggunakan regresi linier berganda, dilanjutkan uji multikolinieritas. Uji multikolinearitas diidentifikasi dengan melihat nilai *Variance Inflation Factor* (VIF). Jika seluruh nilai VIF kurang dari 10, maka dapat disimpulkan bahwa tidak terjadi multikolinieritas. Dengan demikian, analisis dapat dilanjutkan dengan uji normalitas, homoskedastisitas dan autokorelasi. Kemudian ke tahap pengujian selanjutnya yaitu, signifikansi parameter (baik secara simultan maupun parsial), serta pengujian koefisien determinasi. Setelah seluruh tahapan tersebut dilakukan, lalu interpretasikan hasil dan proses analisis dapat dianggap selesai. Namun, jika ada nilai VIF yang lebih dari 10, maka disimpulkan terjadi multikolinieritas. Untuk mengatasinya, dilakukan standarisasi data terlebih dahulu, kemudian gunakan regresi LASSO dengan algoritma *Least Angle Regression* (LAR).

Selanjutnya, dilakukan validasi silang untuk memperoleh nilai *Mean Squared Error* (MSE) dan s . Nilai MSE digunakan untuk membentuk model terbaik dan nilai s digunakan untuk menunjukan besarnya nilai penalti (hukuman) dari regresi LASSO dengan algoritma LAR. Setelah model terbaik diperoleh, lalu uji koefisien determinasi R_k^2 berdasarkan model terbaik untuk mengetahui seberapa baik variabel bebas dalam menjelaskan hubungan terhadap variabel terikat. Kemudian, lakukan perbandingan antara model awal yang menggunakan regresi linier berganda dengan model terbaik regresi LASSO, serta bandingkan juga nilai VIF sebelum dan setelah menggunakan regresi LASSO. Langkah terakhir adalah melakukan interpretasi terhadap hasil yang diperoleh, sehingga seluruh proses analisis

dinyatakan selesai.

Berdasarkan alur penelitian yang telah diuraikan, untuk memberikan gambaran yang lebih jelas mengenai proses yang dilakukan, berikut disajikan diagram alir penelitian:



Gambar 3.1. Alur Penelitian

BAB IV

HASIL DAN PEMBAHASAN

A. Deskriptif Data

Deskriptif data merupakan teknik yang digunakan untuk menganalisis data dengan menyajikan atau menggambarkan data yang telah dikumpulkan, tanpa membuat generalisasi atau kesimpulan yang bersifat umum (Sugiyono, 2024). Analisis deskriptif data dalam penelitian ini dapat dilihat pada Lampiran 2. Adapun ringkasan hasil analisis deskriptif data disajikan pada Tabel berikut:

Tabel 4.1. Deskriptif Data

Statistik	Y	X_1	X_2	X_3	X_4
Minimum	635354	65,7	6,11	11,43	71,02
Maksimum	1080574	69,82	6,9	11,81	72,4
Mean	862607,3	67,959	6,646	11,711	71,692
Median	908196,5	68,245	6,78	11,745	71,71
Range	445220	4,12	0,79	0,38	1,38
Standar Deviasi	151263,1	1,3378	0,3037	0,1102	0,4579

Dengan keterangan sebagai berikut:

- Y : Pengeluaran Per Kapita
- X_1 : Indeks Pembangunan Manusia (IPM)
- X_2 : Rata-rata Lama Sekolah (RLS)
- X_3 : Harapan Lama Sekolah (HLS)

- X_4 : Umur Harapan Hidup (UHH)

B. Model Awal Menggunakan Regresi Linier Berganda

Berdasarkan output pada Lampiran 3, dilakukan analisis menggunakan regresi linier berganda dengan *Ordinary Least Squares* (OLS) untuk mengidentifikasi faktor-faktor yang diduga memengaruhi pengeluaran per kapita di Kabupaten Wonosobo. Adapun persamaan model tersebut adalah sebagai berikut:

$$\hat{Y} = 3042554 + 158835X_1 + 103685X_2 + 239478X_3 - 229857X_4$$

Model ini menunjukkan bahwa nilai prediksi pengeluaran per kapita yang dipengaruhi oleh Indeks Pembangunan Manusia (IPM) sebagai X_1 , Rata-rata Lama Sekolah (RLS) sebagai X_2 , dan Harapan Lama Sekolah (HLS) sebagai X_3 memiliki pengaruh positif terhadap pengeluaran per kapita. Sebaliknya, variabel Umur Harapan Hidup (UHH) sebagai X_4 memberikan pengaruh negatif. Artinya, setiap peningkatan satu satuan pada X_1 , X_2 , dan X_3 diperkirakan akan meningkatkan pengeluaran per kapita masing-masing sebesar 158.835, 103.685, dan 239.478 satuan. Sementara itu, peningkatan satu satuan pada X_4 justru diperkirakan akan menurunkan pengeluaran per kapita sebesar 229.857 satuan. Nilai konstanta sebesar 3.042.554 menunjukkan prediksi pengeluaran per kapita saat seluruh variabel bebas bernilai nol. Secara keseluruhan, model ini mengindikasikan bahwa peningkatan IPM, RLS, dan HLS berkontribusi terhadap kenaikan pengeluaran per kapita, sedangkan peningkatan UHH cenderung menurunkannya.

C. Uji Multikolinieritas

Dilakukan untuk membuktikan ada tidaknya hubungan linier antara satu atau lebih variabel bebas pada model regresi yang akan digunakan. Pada penelitian ini, analisis mengenai multikolinieritas dilakukan dengan memanfaatkan nilai *Variance Inflation Factor* (VIF) dan hasilnya dapat dilihat pada Lampiran 4.

Tabel 4.2. Hasil Uji Multikolinieritas

Variabel	VIF
X_1	31,246585
X_2	9,073431
X_3	4,734037
X_4	26,075121

Berdasarkan tabel 4.2, dapat disimpulkan bahwa **variabel X_2 dan X_3 tidak terdapat multikolinieritas** karena nilai VIF kurang dari 10 sedangkan pada variabel X_1 dan X_4 **terdapat multikolinieritas** karena nilai VIF lebih dari 10. Oleh karena itu, dilanjutkan dengan menstandarisasi data kemudian menggunakan regresi LASSO untuk mengatasi permasalahan multikolinieritas tersebut.

D. Standarisasi data

Standarisasi data merupakan bagian proses pemusatan dan penskalaan nilai. Data yang akan distandarisasi yaitu pengeluaran per kapita (Y), Indeks pembangunan manusia (X_1), Rata-rata lama sekolah (X_2), Harapan lama sekolah (X_3) dan Umur harapan hidup (X_4) dengan perhitungan sebagai berikut:

Langkah 1: Menghitung Rata-rata untuk data pengeluaran per kapita (Y)

Rumus rata-rata:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

Substitusi nilai:

$$\begin{aligned}\bar{Y} &= \frac{980047 + 1080574 + 987422 + 923840 + 833995 + \\ &\quad 892553 + 930699 + 705160 + 656429 + 635354}{10} \\ &= \frac{8626073}{10} \\ &= 862607,3\end{aligned}$$

Langkah 2: Menghitung Standar Deviasi (S_Y)

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}$$

Hitung selisih kuadrat:

$$\begin{aligned}(980047 - 862607,3)^2 &= 13792083136,09 \\ (1080574 - 862607,3)^2 &= 47509482308,89 \\ (987422 - 862607,3)^2 &= 15578709336,09 \\ (923840 - 862607,3)^2 &= 3749443549,29 \\ (833995 - 862607,3)^2 &= 818663711,29 \\ (892553 - 862607,3)^2 &= 896744948,49 \\ (930699 - 862607,3)^2 &= 4636479608,89 \\ (705160 - 862607,3)^2 &= 24789652277,29 \\ (656429 - 862607,3)^2 &= 42509491390,89 \\ (635354 - 862607,3)^2 &= 51644062360,89\end{aligned}$$

Jumlah total selisih kuadrat:

$$\sum (Y_i - \bar{Y})^2 = 205924812628,1$$

Hitung standar deviasi:

$$\begin{aligned} S_Y &= \sqrt{\frac{205924812628,1}{10 - 1}} \\ &= 151263,14 \end{aligned}$$

Langkah 3: Menghitung Z-Score untuk $Y_1 = 980047$

Rumus Z-Score:

$$z_1 = \frac{Y_1 - \bar{Y}}{S_Y}$$

Substitusi nilai:

$$\begin{aligned} z_1 &= \frac{980047 - 862607,3}{151263,14} \\ &= \frac{117439,7}{151263,14} \\ &= 0,776393 \end{aligned}$$

Jadi, nilai Z-Score untuk Y_1 adalah 0,776393.

Untuk perhitungan lengkap dapat dilihat pada Lampiran 5 dan kemudian disajikan pada tabel di bawah ini:

Tabel 4.3. Standarisasi Data

Observasi	Y	X_1	X_2	X_3	X_4
1	0,7764	1,4687	0,8365	0,8984	1,5465
2	1,4410	1,0997	0,8036	0,8077	1,0441
3	0,8252	0,7061	0,7707	0,6262	0,7820
4	0,4049	0,3289	0,5731	0,4445	0,5418
5	-0,1892	0,1567	0,5401	0,3540	0,2796
6	0,1980	0,1977	0,0791	0,2632	-0,2010
7	0,4502	-0,1796	0,3425	-0,1906	-0,5068
8	-1,0409	-0,9341	-0,4479	-0,2814	-0,8563
9	-1,3631	-1,5081	-1,7325	-0,3721	-1,1621
10	-1,5024	-1,3359	-1,7652	-2,5500	-1,4679

E. Penanganan multikolinieritas menggunakan regresi LASSO dengan algoritma LAR

1. Mencari vektor yang sebanding dengan vektor korelasi antara variabel bebas dan residual dari setiap variabel bebas pada seleksi pertama diperoleh nilai yaitu:

$$\begin{aligned}
 \hat{c} &= \mathbf{X}^\top (\mathbf{Y} - \hat{\mu}) \\
 &= \mathbf{X}^\top (\mathbf{Y} - \mathbf{0}) \\
 &= \mathbf{X}^\top \mathbf{Y}
 \end{aligned}$$

Karena $\mathbf{X}^\top$ adalah *transpose* dari $\mathbf{X}$, maka dihitung perkalian dot product antara setiap kolom $\mathbf{X}_j$ dan vektor $\mathbf{Y}$

sebagai berikut:

$$\hat{c}_k = \mathbf{X}_k^\top \mathbf{Y} = \sum_{i=1}^n \mathbf{X}_{ik} \mathbf{Y}_i$$

$$\begin{aligned} \hat{c}_1 &= (1,4687 \times 0,7764) + (1,0997 \times 1,4410) + \dots \\ &\quad + (-1,6888 \times -1,3559) \end{aligned}$$

$$\begin{aligned} \hat{c}_2 &= (0,8365 \times 0,7764) + (0,8036 \times 1,4410) + \dots \\ &\quad + (-1,7652 \times -1,5024) \end{aligned}$$

$$\begin{aligned} \hat{c}_3 &= (0,8984 \times 0,7764) + (0,8076 \times 1,4410) + \dots \\ &\quad + (-2,5500 \times -1,5024) \end{aligned}$$

$$\begin{aligned} \hat{c}_4 &= (1,5465 \times 0,7764) + (1,0441 \times 1,4410) + \dots \\ &\quad + (-1,4678 \times -1,5024) \end{aligned}$$

Hasil perhitungan vektor korelasi awal adalah sebagai berikut dan dapat dilihat pada Lampiran 6 (Poin A):

$$\hat{c} = \begin{bmatrix} 8,4040 \\ 8,2222 \\ 7,0884 \\ 7,9294 \end{bmatrix}$$

2. Menentukan nilai korelasi terbesar dari vektor korelasi

$$\hat{C} = \max \{|\hat{c}_k|\}$$

$$\max \{8,4040, 8,2222, 7,0884, 7,9294\} = 8,4040$$

Jadi, nilai korelasi terbesar adalah $\hat{C} = 8,4040$. Kemudian, tanda s_k ditentukan sebagai:

$$\begin{aligned} s_k &= \text{sign}\{|\hat{c}_k|\} \\ &= \text{sign}(8,4040) \\ &= 1 \end{aligned}$$

Karena $\hat{c}_k$ bernilai positif, maka nilai mutlaknya adalah:

$$|\hat{c}_k| = 8,4040 \quad (\text{tetap positif})$$

Dengan demikian, arah perubahan koefisien regresi β_k akan mengikuti tanda dari $s_k = 1$, sehingga β_k akan bertambah (bergerak ke arah positif) pada langkah tersebut dan telah terlampir pada Lampiran 6 (Poin B).

3. Menentukan X_A

Dari langkah (2), nilai korelasi terbesar adalah 8,4040 sehingga dapat disimpulkan bahwa variabel pertama yang masuk ke model adalah X_1 , maka indeks variabel aktif yaitu himpunan $A = \{1\}$ dengan nilai X_A berisi data dari variabel X_1 yang telah di standarisasi dan dapat dilihat pada Lampiran 6 (Poin C):

Tabel 4.4. Nilai Variabel Aktif Tahap 1

Observasi	Nilai X_1
1	1,4687
2	1,0997
3	0,7061
4	0,3289
5	0,1567
6	0,1977
7	-0,1796
8	-0,9341
9	-1,5081
10	-1,3359

Kemudian untuk langkah selanjutnya yaitu menentukan nilai G_A dan A_A sebagai berikut dan dapat dilihat pada Lampiran 6 (Poin C):

a) Menentukan G_A

$$\begin{aligned}
 G_A &= \mathbf{X}_A^\top \mathbf{X}_A = \sum \mathbf{X}_A^2 \\
 G_A &= (1,4687)^2 + (1,0997)^2 + (0,7061)^2 \\
 &\quad + (0,3289)^2 + (0,1567)^2 + (0,1977)^2 \\
 &\quad + (-0,1796)^2 + (-0,9341)^2 + (-1,5081)^2 \\
 &\quad + (-1,3359)^2 \\
 &= 2,1571 + 1,2094 + 0,4986 \\
 &\quad + 0,1082 + 0,0246 + 0,0391 \\
 &\quad + 0,0323 + 0,8726 + 2,2744 + 1,7847 \\
 &= 9,00
 \end{aligned}$$

Jadi, $G_A = 9,00$

b) Menentukan A_A

$$A_A = (\mathbf{1}_A^\top G_A^{-1} \mathbf{1}_A)^{-\frac{1}{2}}$$

Karena hanya terdapat satu variabel aktif, maka matriks G_A merupakan skalar dengan nilai 9,00. Invers dari skalar G_A dapat dihitung sebagai berikut:

$$G_A^{-1} = \frac{1}{9,00} = 0,1111$$

Vektor $\mathbf{1}_A$ merupakan vektor dengan elemen satu sebanyak jumlah variabel aktif. Karena pada langkah ini hanya terdapat satu variabel aktif, maka vektor $\mathbf{1}_A$ dan transposenya dapat dituliskan sebagai berikut:

$$\mathbf{1}_A = \begin{bmatrix} 1 \end{bmatrix} \quad \text{dan} \quad \mathbf{1}_A^\top = \begin{bmatrix} 1 \end{bmatrix}$$

Dengan demikian, hasil perkalian $\mathbf{1}_A^\top G_A^{-1} \mathbf{1}_A$ dapat dihitung sebagai:

$$\mathbf{1}_A^\top G_A^{-1} \mathbf{1}_A = 1 \times 0,1111 \times 1 = 0,1111$$

Selanjutnya, hitung nilai A_A :

$$\begin{aligned} A_A &= (\mathbf{1}_A^\top G_A^{-1} \mathbf{1}_A)^{-\frac{1}{2}} \\ &= (0,1111)^{-\frac{1}{2}} \\ &= \frac{1}{\sqrt{0,1111}} \\ &= 3,00015 \approx 3,00 \end{aligned}$$

Jadi, nilai $A_A = 3,00$.

4. Menghitung nilai vektor *equiangular*

Digunakan untuk menentukan arah pergerakan koefisien regresi secara adil dan seimbang terhadap semua variabel aktif, agar perubahan nilai koefisien tetap konsisten untuk setiap variabel yang sedang digunakan dalam model. Dihitung dengan rumus di bawah ini:

$$u_A = X_A w_A$$

dimana:

$$w_A = A_A G_A^{-1} \mathbf{1}_A$$

Hitung w_A :

$$\begin{aligned} w_A &= A_A \times G_A^{-1} \times \mathbf{1}_A \\ &= 3,00 \times 0,1111 \\ &= 0,3333 \end{aligned}$$

Karena $\mathbf{1}_A$ adalah vektor yang elemennya semua 1, maka setiap elemen w_A adalah 0,3333.

Kemudian, hitung $u_A = X_A \times w_A$ dan hasilnya dapat dilihat juga pada Lampiran 6 (Poin D):

$$\mathbf{u}_A = \begin{bmatrix} 1,4687 \times 0,3333 \\ 1,0997 \times 0,3333 \\ 0,7061 \times 0,3333 \\ 0,3289 \times 0,3333 \\ 0,1567 \times 0,3333 \\ 0,1977 \times 0,3333 \\ -0,1796 \times 0,3333 \\ -0,9341 \times 0,3333 \\ -1,5081 \times 0,3333 \\ -1,3359 \times 0,3333 \end{bmatrix} = \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix}$$

5. Menghitung nilai vektor *inner product*

$$\begin{aligned} \mathbf{a} &\equiv \mathbf{X}^\top \mathbf{u}_A \\ &= \begin{bmatrix} 1,4687 & 1,0997 & \cdots & -1,3359 \end{bmatrix} \times \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix} \\ &= 0,7191 + 0,4031 + 0,1662 + 0,0361 + 0,0082 \\ &\quad + 0,0130 + 0,0108 + 0,2908 + 0,7581 + 0,5949 \\ &= 3,0002 \approx 3,00 \end{aligned}$$

Jadi, nilai vektor *inner product* untuk variabel aktif yaitu 3.00 dan dapat dilihat pada Lampiran 6 (Poin E).

6. Menghitung $\hat{\gamma}$

$$\hat{\gamma} = \min_{k \in A^c}^+ \left\{ \frac{\hat{C} - \hat{c}_k}{A_A - \mathbf{a}_k}, \frac{\hat{C} + \hat{c}_k}{A_A + \mathbf{a}_k} \right\}$$

Untuk nilai $\mathbf{a}_k$ belum diketahui hasilnya secara langsung karena $\mathbf{a}_k$ merupakan hasil nilai vektor *inner product* dari variabel bebas yang belum aktif. Pada langkah ini, variabel bebas yang belum aktif adalah X_2 , X_3 , dan X_4 . Oleh karena itu, perlu dilakukan perhitungan nilai $\mathbf{a}_k$ untuk masing-masing variabel tersebut:

- Perhitungan $\mathbf{X}_2^\top \mathbf{u}_A$:

$$\begin{aligned} \mathbf{X}_2^\top \mathbf{u}_A &= \begin{bmatrix} 0,8365 & 0,8036 & \cdots & -1,7652 \end{bmatrix} \times \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix} \\ &= 0,4095 + 0,2946 + 0,1815 + 0,0629 + 0,0283 \\ &\quad + 0,0053 - 0,0205 + 0,1395 + 0,8708 + 0,7861 \\ &= 2,758 \end{aligned}$$

- Perhitungan $\mathbf{X}_3^\top \mathbf{u}_A$:

$$\begin{aligned}
 \mathbf{X}_3^\top \mathbf{u}_A &= \begin{bmatrix} 0,8984 & 0,8077 & \cdots & -2,5500 \end{bmatrix} \times \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix} \\
 &= 0,4399 + 0,2962 + 0,1475 + 0,0488 + 0,0186 \\
 &\quad + 0,0174 + 0,0114 + 0,0877 + 0,1871 + 1,1356 \\
 &= 2,389
 \end{aligned}$$

- Perhitungan $\mathbf{X}_4^\top \mathbf{u}_A$:

$$\begin{aligned} \mathbf{X}_4^\top \mathbf{u}_A &= \begin{bmatrix} 1,5465 & 1,0441 & \dots & -1,4679 \end{bmatrix} \times \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix} \\ &= 0,7572 + 0,3828 + 0,1841 + 0,0595 + 0,0147 \\ &\quad - 0,0133 + 0,0304 + 0,2667 + 0,5841 + 0,6537 \\ &= 2,919 \end{aligned}$$

Kemudian, diperoleh untuk $\min_{k \in A^c}^+$ yaitu:

- Untuk $k = 2$ dengan $c_2 = 8,2222$:

$$\begin{aligned} \frac{8,4040 - 8,2222}{3,00 - 2,758} &= \frac{0,1818}{0,242} = 0,75 \\ \frac{8,4040 + 8,2222}{3,00 + 2,758} &= \frac{16,6262}{5,758} = 2,89 \end{aligned}$$

- Untuk $k = 3$ dengan $c_3 = 7,0884$:

$$\begin{aligned} \frac{8,4040 - 7,0884}{3,00 - 2,389} &= \frac{1,3156}{0,611} = 2,16 \\ \frac{8,4040 + 7,0884}{3,00 + 2,389} &= \frac{15,4924}{5,389} = 2,88 \end{aligned}$$

- Untuk $k = 4$ dengan $c_4 = 7,9294$:

$$\frac{8,4040 - 7,9294}{3,00 - 2,919} = \frac{0,4746}{0,081} = 5,86$$

$$\frac{8,4040 + 7,9294}{3,00 + 2,919} = \frac{16,3334}{5,919} = 2,76$$

Dengan demikian, nilai $\hat{\gamma}$ diperoleh dengan mengambil nilai minimum positif dari hasil pembagian di atas:

$$\hat{\gamma} = \min_{k \in A^c}^+ \{0,75; 2,89; 2,16; 2,88; 5,86; 2,76\} = 0,75$$

Jadi, nilai $\min_{k \in A^c}^+$ yaitu 0,75 dan hasilnya dapat dilihat di Lampiran 6 (Poin F).

7. Menghitung $\hat{\mu}_{A+}$

Selanjutnya dilakukan perhitungan vektor prediksi $\hat{\mu}_{A+}$ dengan menggunakan rumus:

$$\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma}u_A$$

Karena ini merupakan langkah pertama dari proses seleksi variabel dengan algoritma LAR, maka nilai $\hat{\mu}_A$ masih berupa vektor nol. Dengan demikian, perhitungan $\hat{\mu}_{A+}$ hanya bergantung pada hasil perkalian antara $\hat{\gamma}$ dengan vektor *equiangular* u_A . Telah diketahui dari hasil sebelumnya bahwa nilai $\hat{\gamma} = 0,75$ dan vektor *equiangular*. Maka vektor prediksi dapat di hitung sebagai berikut dan telah terlampir pada Lampiran 6 (Poin G):

$$\mathbf{u}_A = \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix}$$

Maka diperoleh:

$$\hat{\mu}_{A+} = 0,75 \times \mathbf{u}_A = \begin{bmatrix} 0,3672 \\ 0,2749 \\ 0,1765 \\ 0,0822 \\ 0,0392 \\ 0,0494 \\ -0,0448 \\ -0,2335 \\ -0,3769 \\ -0,3339 \end{bmatrix}$$

Nilai $\hat{\mu}_{A+}$ ini akan digunakan pada langkah kedua dalam proses seleksi variabel dengan algoritma LAR.

8. Menghitung nilai calon koefisien LASSO

Nilai calon koefisien regresi LASSO dihitung menggunakan rumus:

$$\begin{aligned}
 \beta_k^{(t+1)} &= \beta_k^{(t)} + \hat{\gamma} \times d_k \\
 &= \beta_k^{(t)} + \hat{\gamma} \times s_k \mathbf{w}_{Ak} \\
 &= 0 + 0,75 \times 1 \times 0,3333 \\
 &= 0,2499
 \end{aligned}$$

Jadi, nilai calon koefisien regresi LASSO pada langkah pertama yaitu $\beta_k^{(t+1)} = 0,2499$ dan dapat dilihat pada Lampiran 6 (Poin H).

9. Melakukan pengecekan apakah $\text{sign}(\hat{\beta}_k) = \text{sign}(\hat{c}_k) = s_k$. Pada seleksi variabel pertama diperoleh $\text{sign}(\hat{\beta}_1) = \text{sign}(\hat{c}_1) = s_1 = 1$, karena $\hat{\beta}_1$ dan $\hat{c}_1$ memiliki tanda yang sama dengan s_1 maka dapat melanjutkan seleksi variabel kedua.

10. Langkah yang sama dilakukan secara berulang pada setiap proses seleksi hingga seluruh variabel bebas terseleksi.

Berikut adalah tabel semua nilai koefisien LASSO dan dapat dilihat juga ada Lampiran 6 (Poin I):

Tabel 4.5. Nilai Koefisien LASSO pada Setiap Langkah

Langkah	$\beta_1 (X_1)$	$\beta_2 (X_2)$	$\beta_3 (X_3)$	$\beta_4 (X_4)$
0	0,00000000	0,00000000	0,00000000	0,00000000
1	0,2498419	0,00000000	0,00000000	0,00000000
2	0,5865269	0,3366850	0,00000000	0,00000000
3	0,5973010	0,3408444	0,01239943	0,00000000
4	1,2804893	0,2081409	0,17446217	-0,6956909

Tabel 4.5 menunjukkan perkembangan nilai koefisien regresi LASSO pada setiap langkah seleksi variabel menggunakan algoritma *Least Angle Regression* (LAR) dengan proses perhitungannya dilakukan melalui lima langkah. Pada langkah awal (langkah ke-0), seluruh koefisien bernilai nol, yang berarti belum ada variabel bebas yang dimasukkan ke dalam model.

Memasuki langkah ke-1, variabel X_1 menjadi variabel pertama yang terpilih karena memiliki korelasi paling tinggi dengan variabel terikat. Hal ini dapat dilihat dari nilai koefisien $\beta_1 = 0,2498$, sedangkan koefisien variabel lainnya masih bernilai nol.

Pada langkah ke-2, variabel X_2 ditambahkan ke dalam model dengan nilai koefisien sebesar $\beta_2 = 0,3367$ dan koefisien β_1 meningkat menjadi $\beta_1 = 0,5865$. Ini menunjukkan bahwa X_1 dan X_2 memiliki pengaruh bersama dari kedua variabel tersebut terhadap model.

Langkah ke-3 menunjukkan masuknya variabel X_3 ke dalam model dengan nilai koefisien yang masih kecil, yaitu $\beta_3 = 0,0124$, sementara β_1 dan β_2 masih mengalami sedikit peningkatan. Hal ini menunjukkan bahwa X_3 mulai memberikan kontribusi meskipun belum dominan.

Pada langkah terakhir, seluruh variabel telah masuk ke dalam

model. Terlihat bahwa $\beta_4 = -0,6957$, yang berarti variabel X_4 memberikan kontribusi negatif terhadap variabel terikat. Dengan kata lain, ketika nilai X_4 meningkat, maka nilai prediksi dari variabel terikat cenderung menurun, dengan asumsi variabel lain tetap konstan.

Urutan masuk variabel ke dalam model LASSO ditentukan berdasarkan saat pertama kali nilai koefisien variabel tersebut berubah dari nol menjadi bukan nol pada setiap langkah algoritma. Berdasarkan Tabel 4.5, pada Langkah 1, variabel X_1 menjadi variabel pertama yang masuk ke dalam model karena nilai koefisiennya berubah dari 0 menjadi 0,2498. Selanjutnya, pada Langkah 2, variabel X_2 masuk ke dalam model dengan koefisien sebesar 0,3366. Pada Langkah 3, variabel X_3 masuk dengan nilai koefisien sebesar 0,0123. Terakhir, pada Langkah 4, variabel X_4 masuk dengan koefisien sebesar $-0,6956$.

Dengan demikian, urutan masuk variabel ke dalam model LASSO adalah sebagai berikut:

Tabel 4.6. Urutan Masuknya Variabel ke dalam Model LASSO

Langkah ke-	Variabel yang Masuk
1	X_1
2	X_2
3	X_3
4	X_4

F. Uji validasi silang

Dalam pemilihan model terbaik pada regresi LASSO, dilakukan proses validasi silang k -fold dengan menggunakan *mode steps*.

Mode steps merupakan proses evaluasi nilai validasi silang pada setiap langkah, yaitu saat koefisien variabel mulai dimasukkan ke dalam model. Berdasarkan output pada Lampiran 7, dalam penelitian ini digunakan metode *K-fold* dengan jumlah bagian $c = 10$ untuk mengevaluasi performa model dan memilih parameter s optimal. Nilai validasi silang dengan *mode steps* diperoleh sebagaimana ditampilkan pada Tabel 4.7 berikut:

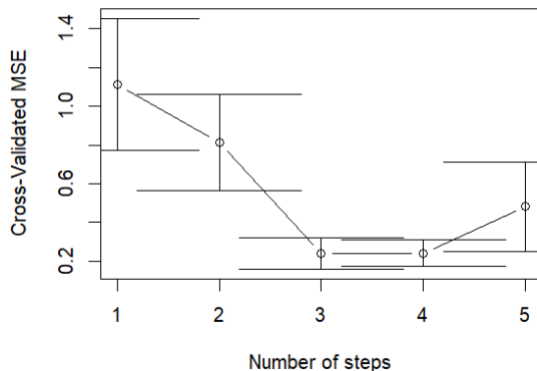
Tabel 4.7. Nilai Validasi Silang dengan *Mode Steps*

Langkah	Nilai Validasi Silang
0	1,1111111
1	0,8134334
2	0,2397145
3	0,2425746
4	0,4819177

Validasi silang dimulai dari langkah ke-0 karena ingin mengevaluasi kinerja model sebelum variabel aktif di masukan. Pada tahap ini, model hanya memprediksi nilai variabel terikat berdasarkan rata-rata data. Meskipun variabel mulai masuk pada langkah ke-1, langkah ke-0 tetap dihitung sebagai acuan awal dalam menilai peningkatan akurasi model dan digunakan sebagai pembanding awal untuk mengevaluasi seberapa besar penurunan nilai validasi silang setelah variabel mulai dimasukkan ke dalam model pada langkah-langkah selanjutnya.

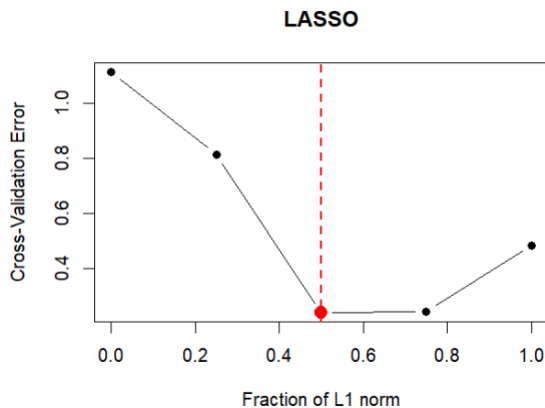
Dalam pemilihan model terbaik menggunakan *mode steps*, diperoleh nilai validasi silang minimum sebesar 0,2397145 yang merupakan nilai *mean squared error* (MSE) minimum dari proses validasi silang. Oleh karena itu, model pada langkah ke-2 (dimulai

dari langkah ke-0) dipilih sebagai model terbaik. Hal ini juga dapat dilihat pada Gambar 4.1, di mana pada tahap ke-3 (dimulai dari langkah ke-1) terlihat titik terendah yang menandai perubahan tren nilai validasi silang, dari menurun menjadi meningkat.



Gambar 4.1. Nilai Validasi Silang dengan *Mode Steps*

Berdasarkan Gambar 4.1 dapat dilihat bahwa terdapat 5 tahapan yang di hasilkan oleh *mode steps*. Ini menunjukkan, nilai s yang dihasilkan juga sebanyak 5 nilai. Berdasarkan output pada Lampiran 7, nilai s optimal yang diperoleh yaitu 0, 5.



Gambar 4.2. Nilai s dengan *Mode Steps*

Selanjutnya, nilai s yang diperoleh dari *mode steps* akan dibandingkan dengan nilai s yang dihasilkan oleh algoritma LAR. Nilai s dari algoritma LAR ditampilkan pada tabel berikut:

Tabel 4.8. Nilai s pada Setiap Langkah Algoritma LAR

Langkah	$\sum_{k=i}^p \hat{\beta}_k^{\text{lasso}} $	$\sum_{k=1}^p \hat{\beta}_k^{\text{ols}} $	$s = \frac{\sum_{k=i}^p \hat{\beta}_k^{\text{lasso}} }{\sum_{k=1}^p \hat{\beta}_k^{\text{ols}} }$
0	0,0000	2,3583	0,0000
1	0,2498	2,3583	0,1060
2	0,9231	2,3583	0,3915
3	0,9504	2,3583	0,4031
4	2,3585	2,3583	1,0000

Nilai $\sum_{k=i}^p |\hat{\beta}_k^{\text{lasso}}|$ diperoleh dari penjumlahan koefisien pada Tabel 4.5, sedangkan $\sum_{k=1}^p |\hat{\beta}_k^{\text{ols}}|$ diperoleh dari penjumlahan hasil OLS yang telah distandarisasi, sebagaimana terlampir pada Lampiran 8. Tabel 4.8 menginformasikan bahwa nilai s yang

diperoleh pada langkah ke-3 (dimulai dari langkah ke-0), yaitu sebesar 0,4031. Nilai ini lebih mendekati nilai s yang dihasilkan melalui *mode steps*, yaitu sebesar 0,5.

Berdasarkan hasil validasi silang pada Tabel 4.7 dengan menggunakan *mode steps*, model terbaik LASSO diperoleh pada langkah ke-2 (dimulai dari langkah ke-0). Model ini dipilih karena menghasilkan nilai *mean squared error* (MSE) paling minimum dibandingkan dengan langkah-langkah lainnya. Oleh karena itu, model pada langkah ke-2 (dimulai dari langkah ke-0) dianggap sebagai model yang paling optimal dalam memprediksi data, karena memberikan tingkat kesalahan prediksi yang paling minimum.

G. Membentuk model terbaik LASSO dengan algoritma LAR

Berdasarkan output pada Lampiran 9, model terbaik LASSO diperoleh pada langkah ke-2 (dimulai dari langkah ke-0), yang menghasilkan koefisien regresi LASSO sebagaimana ditunjukkan pada Tabel 4.9. Dalam tabel tersebut terlihat bahwa model regresi LASSO menyusutkan nilai koefisien dari regresi linier berganda hingga menjadi nol, sehingga secara otomatis melakukan seleksi terhadap variabel-variabel bebas yang akan dimasukkan ke dalam model.

Tabel 4.9. Koefisien Regresi LASSO

Variabel	Koefisien
X_1	0,5865269
X_2	0,3366850
X_3	0,0000000
X_4	0,0000000

Berdasarkan Tabel 4.9, dapat disimpulkan bahwa variabel bebas yang memiliki pengaruh signifikan terhadap variabel terikat dalam analisis faktor-faktor yang memengaruhi pengeluaran per kapita di Kabupaten Wonosobo tahun 2015–2024 adalah X_1 (Indeks Pembangunan Manusia) dan X_2 (Rata-rata Lama Sekolah), dengan nilai koefisien regresi masing-masing sebesar 0,5865269 dan 0,3366850. Sementara itu, variabel X_3 (Harapan Lama Sekolah) dan X_4 (Usia Harapan Hidup) memiliki koefisien regresi sebesar 0, yang berarti tidak dipilih oleh model karena dianggap tidak memberikan pengaruh signifikan terhadap pengeluaran per kapita. Oleh karena itu, penggunaan regresi LASSO mampu menyederhanakan model dengan hanya mempertahankan variabel yang signifikan.

Nilai intersep β_0 dapat dilihat pada Lampiran 10 (Poin B), yang dihasilkan dari model regresi LASSO sebesar $-2,4277 \times 10^{-15}$. Nilai intersep β_0 tersebut dapat dihitung secara manual dengan menggunakan rumus umum dalam regresi linier, yaitu:

$$\beta_0 = \bar{Y} - \sum_{i=1}^n \beta_k \times \bar{X}_k$$

$$\beta_0 = \bar{Y} - (\beta_1 \times \bar{X}_1 + \beta_2 \times \bar{X}_2 + \beta_3 \times \bar{X}_3 + \beta_4 \times \bar{X}_4)$$

Dengan menggunakan nilai-nilai yang telah diperoleh sebelumnya dan dapat dilihat secara rinci pada Lampiran 10 (Poin A), perhitungan intersep β_0 dilakukan berdasarkan data yang telah distandarisasi. Nilai dari rata-rata dan koefisien regresi tersebut digunakan dalam rumus perhitungan β_0 untuk memastikan konsistensi antara hasil perhitungan manual dan output dari

software R.

$$\begin{aligned}\bar{Y} &= -2.802879 \times 10^{-16} \\ \bar{X}_1 &= -3.502732 \times 10^{-15} \\ \bar{X}_2 &= 2.761463 \times 10^{-16} \\ \bar{X}_3 &= -3.219647 \times 10^{-15} \\ \bar{X}_4 &= 9.331403 \times 10^{-15} \\ \beta_1 &= 0,5865269 \\ \beta_2 &= 0,3366850 \\ \beta_3 &= 0 \\ \beta_4 &= 0\end{aligned}$$

Hitung setiap komponen:

$$\beta_1 \times \bar{X}_1 = 0.5865269 \times 3.502732 \times 10^{-15} = 2.054447 \times 10^{-15}$$

$$\beta_2 \times \bar{X}_2 = 0.3366850 \times 2.761463 \times 10^{-16} = 9,297432 \times 10^{-17}$$

$$\beta_3 \times \bar{X}_3 = 0 \times (-3.219647 \times 10^{-15}) = 0$$

$$\beta_4 \times \bar{X}_4 = 0 \times 9.331403 \times 10^{-15} = 0$$

Jumlahkan hasil:

$$\sum_{k=1}^4 \beta_k \times \bar{X}_k = 2.054447 \times 10^{-15} + 9,297432 \times 10^{-17} + 0 + 0 = 2.14744 \times 10^{-15}$$

Hitung intersep:

$$\begin{aligned}\beta_0 &= \bar{Y} - \sum_{k=1}^4 \beta_k \times \bar{X}_k \\ &= -2.802879 \times 10^{-16} - 2.14744 \times 10^{-15} \\ &= -2.42773 \times 10^{-15}\end{aligned}$$

Berikut model persamaan regresi LASSO dengan algoritma LAR:

$$\hat{Y} = -2,4277 \times 10^{-15} + 0,5866X_1 + 0,3367X_2$$

Berdasarkan model regresi LASSO yang diperoleh, koefisien X_1 (Indeks Pembangunan Manusia) sebesar 0,5866 menunjukkan bahwa setiap peningkatan satu satuan pada IPM akan meningkatkan pengeluaran per kapita sebesar 0,5866 satuan. Demikian pula, koefisien X_2 (Rata-rata Lama Sekolah) sebesar 0,3367 menunjukkan bahwa setiap kenaikan satu satuan pada RLS akan meningkatkan pengeluaran per kapita sebesar 0,3367 satuan. Nilai intersep sebesar $-2,4277 \times 10^{-15}$ sangat mendekati nol, yang wajar terjadi karena data telah distandarisasi sebelum dilakukan pemodelan.

H. Uji Koefisien Determinasi Model Regresi LASSO

Berdasarkan output pada Lampiran 10 (Poin B), efektivitas dari model regresi LASSO dengan algoritma LAR menghasilkan nilai koefisien determinasi sebesar $R_k^2 = 0,8902$, yang mengindikasikan bahwa sekitar 89% variasi pada variabel terikat dapat dijelaskan oleh dua variabel bebas yang tetap ada dalam model. Sementara itu, variabel lainnya dipengaruhi oleh faktor-faktor di luar model yang tidak turut dianalisis dalam penelitian ini.

Tabel 4.10. Hasil Koefisien Determinasi Model LASSO

Model	R_k^2
LASSO	0,8902

I. Membandingkan koefisien regresi linier berganda dengan regresi LASSO

Salah satu perbedaan utama antara model Regresi Linier Berganda dan Regresi LASSO terletak pada nilai koefisien regresi yang dihasilkan. Model Regresi Linier Berganda menyertakan semua variabel bebas tanpa melakukan proses seleksi, sementara Regresi LASSO memiliki kemampuan untuk melakukan seleksi variabel dengan menyusutkan sebagian koefisien menjadi nol. Untuk menunjukkan perbedaan tersebut, berikut disajikan perbandingan nilai koefisien dari kedua model pada Tabel 4.11.

Tabel 4.11. Perbandingan Koefisien Regresi Linier Berganda dan Regresi LASSO

Variabel	Regresi Linier Berganda	Regresi LASSO
X_1	158835	0,5866
X_2	103685	0,3367
X_3	239478	0
X_4	-229857	0

Pada Tabel 4.11, terlihat bahwa pada model Regresi Linier Berganda, seluruh variabel bebas (X_1 hingga X_4) memiliki koefisien yang relatif besar, baik bernilai positif maupun negatif, yang menunjukkan bahwa semua variabel dimasukkan ke dalam model tanpa proses seleksi. Namun, pada model Regresi LASSO, hanya dua variabel yang memiliki koefisien tidak nol, yaitu X_1 (Indeks Pembangunan Manusia) dengan nilai koefisien sebesar 0,5866, dan X_2 (Rata-rata Lama Sekolah) dengan nilai koefisien sebesar 0,3367. Sementara itu, variabel X_3 (Harapan Lama Sekolah) dan X_4 (Umur Harapan Hidup) mengalami penyusutan

koefisien sampai menjadi nol, yang berarti kedua variabel tersebut dieliminasi dari model karena dianggap tidak memberikan kontribusi signifikan terhadap variabel terikat.

Hal ini menunjukkan bahwa regresi LASSO melakukan penyusutan nilai koefisien-koefisien regresi ke arah nol. Tujuan dari penyusutan ini adalah untuk meningkatkan kesederhanaan dan interpretabilitas model. Dengan demikian, LASSO tidak hanya menghasilkan model yang lebih sederhana, tetapi juga membantu dalam seleksi variabel yang benar-benar relevan terhadap prediksi pengeluaran per kapita.

J. Membandingkan Nilai VIF dalam mendeteksi multikolinieritas menggunakan regresi Linier Berganda dan regresi LASSO

Untuk melihat bagaimana regresi LASSO dapat mengatasi permasalahan multikolinearitas, maka dilakukan perbandingan nilai *Variance Inflation Factor* (VIF) antara model Regresi Linier Berganda dan model Regresi LASSO. Berikut adalah perbandingan nilai VIF dari kedua model ditampilkan pada Tabel 4.12 berikut.

Tabel 4.12. Perbandingan nilai VIF pada Regresi Linier Berganda dan LASSO

Variabel	Regresi Linier Berganda	Regresi LASSO
X_1	31,246585	6,4428
X_2	9,073431	6,4428
X_3	4,734037	0
X_4	26,075121	0

Pada Tabel 4.12, terlihat bahwa nilai VIF dalam regresi linier

berganda untuk variabel X_1 (Indeks Pembangunan Manusia) dan X_4 (Umur Harapan Hidup) relatif tinggi. Nilai-nilai ini mengindikasikan adanya multikolinieritas yang cukup signifikan antar variabel bebas, karena nilai VIF di atas 10 jadi dianggap mengalami multikolinieritas.

Sebaliknya, setelah dilakukan analisis menggunakan regresi LASSO, nilai VIF mengalami penurunan yang cukup drastis. Nilai VIF untuk variabel X_3 (Harapan Lama Sekolah) dan X_4 (Umur Harapan Hidup) menyusut menjadi nol karena kedua variabel tersebut dieliminasi dari model akibat koefisiennya yang menyusut menjadi nol. Sementara itu, nilai VIF untuk variabel X_1 (Indeks Pembangunan Manusia) dan X_2 (Rata-rata Lama Sekolah) menurun menjadi 6,4428 yang dapat dilihat pada Lampiran 11. Nilai ini berada di bawah batas multikolinieritas. Jadi, regresi LASSO tidak hanya menyederhanakan model dengan mengeliminasi variabel yang kurang relevan, tetapi juga efektif dalam mengurangi tingkat multikolinieritas antar variabel bebas yang tersisa.

K. Interpretasi hasil

Regresi LASSO dengan algoritma LAR diterapkan pada data pengeluaran per kapita Kabupaten Wonosobo sebagai variabel terikat (Y), dengan variabel bebas berupa Indeks Pembangunan Manusia (X_1), Rata-rata Lama Sekolah (X_2), Harapan Lama Sekolah (X_3), dan Umur Harapan Hidup (X_4). Dengan memanfaatkan *software* RStudio, kita dapat mengembangkan pemodelan tersebut untuk memperoleh estimasi parameter, mengatasi permasalahan multikolinieritas antar variabel bebas,

serta menyederhanakan model dengan mengeliminasi variabel yang tidak signifikan. Selain itu, Algoritma LAR juga melakukan proses seleksi variabel secara bertahap hingga diperoleh model terbaik berdasarkan kriteria validasi silang, sehingga menghasilkan model regresi yang lebih stabil dan tepat dalam memprediksi pengeluaran per kapita.

Berdasarkan output pada Tabel 4.9, regresi LASSO menghasilkan model terbaik dengan hanya menyisakan dua variabel bebas yang signifikan, yaitu X_1 (Indeks Pembangunan Manusia) dan X_2 (Rata-rata Lama Sekolah). Model regresi LASSO yang diperoleh adalah:

$$\hat{Y} = -2,4277 \times 10^{-15} + 0,5866X_1 + 0,3367X_2$$

Berdasarkan model yang diperoleh, nilai prediksi pengeluaran per kapita menghasilkan nilai intersep sebesar $-2,4277 \times 10^{-15}$ yang sangat mendekati nol. Hal ini wajar terjadi karena data telah distandarisasi sebelum dilakukan pemodelan. Koefisien X_1 (Indeks Pembangunan Manusia) sebesar 0,5866 menunjukkan bahwa setiap peningkatan satu satuan pada IPM akan meningkatkan pengeluaran per kapita sebesar 0,5866 satuan dan koefisien X_2 (Rata-rata Lama Sekolah) sebesar 0,3367 menunjukkan bahwa setiap kenaikan satu satuan pada RLS akan meningkatkan pengeluaran per kapita sebesar 0,3367 satuan. Dengan demikian, X_1 dan X_2 memberikan kontribusi positif terhadap pengeluaran per kapita di Kabupaten Wonosobo.

Berdasarkan hasil prediksi yang disajikan pada Lampiran 7, diperoleh nilai validasi silang untuk parameter $s = 0,5$. Pada nilai tersebut, diperoleh nilai *Mean Squared Error* (MSE) sebesar 0,2397145 dan koefisien determinasi R_k^2 sebesar 0,89. Hal ini

menunjukkan bahwa model mampu menjelaskan sekitar 89% variasi data, dengan *error* prediksi yang relatif kecil, sehingga model dapat dikatakan memiliki kinerja prediksi yang cukup baik. Karena nilai MSE tergolong kecil dan nilai R_k^2 mendekati 1, maka dapat disimpulkan bahwa bahwa model memiliki tingkat akurasi prediksi yang baik.

Secara sosial, tingginya nilai Indeks Pembangunan Manusia (IPM) dan Rata-rata Lama Sekolah (RLS) mencerminkan peningkatan kualitas hidup masyarakat. Ketika kualitas hidup membaik, masyarakat tidak hanya berfokus pada kebutuhan dasar, tetapi juga mulai memperhatikan pendidikan, kesehatan, dan kesejahteraan jangka panjang. Pendidikan yang lebih tinggi, tercermin dari rata-rata lama sekolah yang mendorong masyarakat memiliki pemahaman yang lebih luas tentang pentingnya kualitas hidup, sehingga pola konsumsi mereka pun menjadi lebih kompleks. Hal ini berdampak pada meningkatnya pengeluaran per kapita karena masyarakat cenderung membelanjakan uangnya untuk hal-hal yang menunjang kesejahteraan jangka panjang. Dengan demikian, semakin tinggi indeks pembangunan manusia dan rata-rata lama sekolah, semakin besar kemungkinan masyarakat untuk melakukan pengeluaran yang lebih besar dan beragam, sehingga keduanya berpengaruh terhadap besarnya pengeluaran per kapita (Badan Pusat Statistik, 2024)

BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil analisis data yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Model terbaik regresi LASSO pada data pengeluaran per kapita Kabupaten Wonosobo tahun 2015–2024 diperoleh dengan menggunakan algoritma LAR pada saat validasi silang langkah ke-2 (dimulai dari langkah ke-0) menghasilkan nilai *Mean Squared Error* (MSE) sebesar 0,2397145 yang merupakan nilai MSE minimum dan diperoleh pada nilai parameter penalti $s = 0,5$ adalah sebagai berikut:

$$\hat{Y} = -2,4277 \times 10^{-15} + 0,5866X_1 + 0,3367X_2$$

Model ini menghasilkan nilai koefisien determinasi sebesar $R_k^2 = 0,8902$, yang menunjukkan bahwa model mampu menjelaskan sekitar 89% variasi pengeluaran per kapita berdasarkan variabel X_1 dan X_2 yang dipilih oleh regresi LASSO.

2. Dari model terbaik regresi LASSO, diperoleh bahwa variabel Indeks Pembangunan Manusia (IPM) dan Rata-rata Lama Sekolah (RLS) berpengaruh positif terhadap pengeluaran per kapita Kabupaten Wonosobo tahun 2015–2024. Nilai intersep sebesar $-2,4277 \times 10^{-15}$ yang mendekati nol dapat diterima karena data telah distandarisasi sebelum

pemodelan. Koefisien regresi dari variabel IPM sebesar 0,5866 dan RLS sebesar 0,3367 menunjukkan bahwa peningkatan masing-masing variabel akan meningkatkan pengeluaran per kapita, sehingga keduanya memberikan kontribusi positif terhadap model.

B. Saran

Berdasarkan hasil pembahasan penelitian dan kesimpulan yang diperoleh, maka dapat diberikan saran-saran sebagai berikut:

1. Bagi Pemerintah Kabupaten Wonosobo

Diharapkan pengendalian terhadap faktor-faktor yang memengaruhi pengeluaran per kapita dapat ditingkatkan melalui pemeriksaan data yang lebih baik, karena hal ini dapat mengurangi risiko kesalahan perhitungan dan mencegah keterlambatan dalam perbaikan.

2. Bagi Peneliti Selanjutnya

Variabel yang digunakan dalam penelitian ini terbatas, yaitu hanya sebanyak lima variabel. Oleh karena itu, disarankan kepada peneliti selanjutnya untuk menambahkan variabel-variabel lain yang berkaitan dengan pengeluaran per kapita, agar dapat memberikan pemahaman yang lebih luas mengenai faktor-faktor yang memengaruhi pengeluaran per kapita.

DAFTAR PUSTAKA

- Abubakar, R. (2021). *Metodologi Penelitian*. SUKA-Press UIN Sunan Kalijaga.
- Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., & Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, 96(12), 6745–6750.
- Amansyah, I., Indra, J., Nurlaelasari, E., & Juwita, A. R. (2024). Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia. *Innovative: Journal Of Social Science Research*, 4(4), 1199–1216.
- 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.
- Aprianto, A., Debataraja, N. N., & Imro'ah, N. (2020). Metode Cochran-Orcutt untuk Mengatasi Autokorelasi pada Estimasi Parameter Ordinary Least Squares. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 9(1).
- Badan Pusat Statistik. (2024). *Badan Pusat Statistik*. Diakses pada 17 Desember 2024, dari <http://www.bps.go.id/>.
- Bayu, F. E. (2023). *Penerapan regresi LASSO menggunakan algoritma LARS dan GLMNET dalam memprediksi model*

- saham ASEAN* (Doctoral dissertation, Universitas Negeri Malang).
- Dewi, Y. S. (2010). OLS, LASSO dan PLS pada Data Mengandung Multikolinearitas. *Jurnal Ilmu Dasar*, 11(1), 83-91.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). *Least Angle Regression The Annals of Statistics*, 32(2), 407-499.
- Ghozali, I. (2005). *Aplikasi Analisis Multivariate Dengan Program SPSS*. Universitas Diponegoro.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). New York: McGraw-Hill.
- Hasim, J. A. (2022). *Metode Regresi Least Absolute Shrinkage And Selection Operator (LASSO) Dan Elastic Net Untuk Data Yang Mengandung Multikolinearitas*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2016). *Statistical learning with sparsity: The Lasso and generalizations*. Chapman and Hall/CRC.
- Hidayatullah, S. (2015). *Cara Mudah Menguasai Statistik Deskriptif*. Salemba Teknika.
- Jamco, J. C., Kondolembang, F., & Van Delsen, M. S. N. (2023). Penanganan multikolinearitas pada regresi linier berganda menggunakan regresi Lasso (studi kasus: distribusi presentase produk domestik regional bruto di Provinsi

- Maluku tahun 1999-2021). *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, 2(02), 145–154.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer.
- Kadir (2015). *Statistika Terapan Konsep, Contoh Dan Analisis Data Dengan Program SPSS, Lisrel Dalam Penelitian*. Jakarta: PT. RajaGrafindo Persada.
- Kato, T., & Uemura, M. (2012). Period analysis using the least absolute shrinkage and selection operator (LASSO). *Publications of the Astronomical Society of Japan*, 64(6).
- Kumari, S. S. (2008). *Multicollinearity: Estimation and elimination*. *Journal of Contemporary Research in Management*, 3(1), 87-95.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill.
- Mahrany, Y. (2012). *Ilmu Ekonomi: Pengaruh Indikator Komposit Indeks Pembangunan Manusia Terhadap Pertumbuhan Ekonomi Di Sulawesi Selatan*. Makassar: Fakultas Ekonomi dan Bisnis, UNHAS.
- Mubarak, R. (2021). *PENGANTAR EKONOMETRIKA Edisi Pertama*.
- Melkumova, L., & Shatskikh, S. (2017). Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering*, 201, 746-755.

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Nasution, E. F., & Pane, R. (2024). Perbandingan Regresi LASSO dan Principal Component Regression dalam Mengatasi Masalah Multikolinearitas. *MES: Journal of Mathematics Education and Science*, 10(1), 120-128.
- Nisa, E. K. and Maulina, R. (2024). The Comparison of Inverse Gaussian and Gamma Regression: Application on Stunting Data in Jepara. *Jurnal Matematika, Statistika dan Komputasi*, 21(1), 334–344.
- 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.
- Rahmawati, F., & Suratman, R. Y. (2022). Performa Regresi Ridge dan Regresi Lasso pada Data dengan Multikolinearitas. *Leibniz: Jurnal Matematika*, 2(2), 1-10.
- Rencher, A. C. (2008). *Linear Models in Statistics*. John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 138-139.
- Rachmawati, R. F. N., & Sari, A. C. (2021). Lasso regression for daily rainfall modeling at Citeko Station, Bogor, Indonesia. *Procedia Computer Science*, 179, 383–390.
- Siyoto, S., & Sodik, M. A. (2015). *Dasar metodologi penelitian*. Literasi Media Publishing.

- Sugiyono. (2024). *Metode Penelitian Kuantitatif*. Bandung: Alfabeta.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.
- Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020) *Multiple Linear Regression*(2nd ed.). Manchester: Cathie Marsh Institute Working Paper 2020-01.
- Wahidah, N., Juwita, O., & Arifin, F. N. (2023). *Pengelompokkan Daerah Rawan Bencana di Kabupaten Jember Menggunakan Metode K-Means Clustering*.INFORMAL: Informatics Journal, 8(1), 22-29.
- Widarjono, A. (2018). *Ekonometrika: Teori dan Aplikasi*. Edisi 4, UPP STIM YKPN.

Lampiran 1. Data Penelitian Pengeluaran Per Kapita Kabupaten Wonosobo Tahun 2015-2024 dan Faktor-Faktor yang Diduga Mempengaruhinya

Tahun	Y (Rupiah)	X_1 (Persen)	X_2 (tahun)	X_3 (tahun)	X_4 (tahun)
2024	980047	69,82	6,90	11,81	72,40
2023	1080574	69,37	6,89	11,80	72,17
2022	987422	68,89	6,88	11,78	72,05
2021	923840	68,43	6,82	11,76	71,94
2020	833995	68,22	6,81	11,75	71,82
2019	892553	68,27	6,67	11,74	71,60
2018	930699	67,81	6,75	11,69	71,46
2017	705160	66,89	6,51	11,68	71,30
2016	656429	66,19	6,12	11,67	71,16
2015	635354	66,40	6,11	11,43	71,02

Sumber: Badan Pusat Statistik Kabupaten Wonosobo

Lampiran 2. Deskriptif data

```

> Y <- c(980047, 1080574, 987422, 923840, 833995, 892553, 930699, 705160, 656429, 635354)
> X1 <- c(69.82, 69.37, 68.89, 68.43, 68.22, 68.27, 67.81, 66.89, 66.19, 66.40)
> X2 <- c(6.90, 6.89, 6.88, 6.82, 6.81, 6.67, 6.75, 6.51, 6.12, 6.11)
> X3 <- c(11.81, 11.80, 11.78, 11.76, 11.75, 11.74, 11.69, 11.68, 11.67, 11.43)
> X4 <- c(72.40, 72.17, 72.05, 71.94, 71.82, 71.60, 71.46, 71.30, 71.16, 71.02)
> data <- data.frame(Y, X1, X2, X3, X4)
> library(psych)
> describe(data)

```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Y	1	10	862607.30	151263.13	908196.50	863768.12	113735.44	635354.00	1080574.00	445220.00	-0.29	-1.51	47833.60
X1	2	10	68.03	1.22	68.25	68.03	1.31	66.19	69.82	3.63	-0.17	-1.45	0.39
X2	3	10	6.65	0.30	6.78	6.68	0.16	6.11	6.90	0.79	-0.88	-0.98	0.10
X3	4	10	11.71	0.11	11.75	11.73	0.08	11.43	11.81	0.38	-1.50	1.36	0.03
X4	5	10	71.69	0.46	71.71	71.69	0.56	71.02	72.40	1.38	0.00	-1.56	0.14

Lampiran 3. Model Awal Regresi Linier Berganda dengan OLS

```
> # Hitung regresi linier berganda  
> Regresi <- lm(Y ~ X1 + X2 + X3 + X4, data = df)  
> Regresi
```

Call:

```
lm(formula = Y ~ X1 + X2 + X3 + X4, data = df)
```

Coefficients:

(Intercept)	X1	X2	X3	X4
3042554	158835	103685	239478	-229857

Lampiran 4. Uji Multikolinieritas

```
> library(car)
> # Hitung VIF untuk model regresi
> vif(Regresi)
```

X1	X2	X3	X4
31.246585	9.073431	4.734037	26.075121

Lampiran 5. Standarisasi Data

```
> df_std <- as.data.frame(scale(df))
> X <- as.matrix(df_std[, -1]) # X1 sampai X4
> Y <- as.matrix(df_std$Y)     # Y
> df_std
```

	Y	X1	X2	X3	X4
1	0.7763934	1.4686968	0.83648429	0.8983927	1.5464726
2	1.4409771	1.0996775	0.80355184	0.8076460	1.0440874
3	0.8251495	0.7060569	0.77061939	0.6261525	0.7819734
4	0.4048092	0.3288372	0.57302467	0.4446590	0.5417023
5	-0.1891558	0.1566282	0.54009222	0.3539123	0.2795883
6	0.1979709	0.1976303	0.07903789	0.2631656	-0.2009541
7	0.4501540	-0.1795894	0.34249750	-0.1905682	-0.5067537
8	-1.0408835	-0.9340288	-0.44788135	-0.2813149	-0.8562391
9	-1.3630440	-1.5080589	-1.73224699	-0.3720616	-1.1620387
10	-1.5023707	-1.3358499	-1.76517945	-2.5499834	-1.4678384

Lampiran 6. Regresi LASSO dengan Algoritma LAR

A. Vektor Korelasi

```
> # Pisahkan X dan Y
> X <- as.matrix(df_std[, -1]) # Semua kolom kecuali kolom pertama (Y)
> Y <- as.matrix(df_std$Y)     # Kolom pertama adalah Y
> # Hitung c_hat = X'Y (karena mu = 0 pada awal)
> c_hat <- t(X) %*% Y
> # Tampilkan hasil
> c_hat
      [,1]
X1 8.403978
X2 8.222120
X3 7.088320
X4 7.929315
```

B. Nilai Korelasi Absolut Terbesar

```
> # Cari nilai absolut terbesar dari c_hat
> C_hat <- max(abs(c_hat))
> # Tentukan tanda s_j untuk setiap j berdasarkan c_hat
> s_j <- sign(c_hat)
> # Tampilkan hasil
> C_hat
[1] 8.403978
> s_j
      [,1]
X1      1
X2      1
X3      1
X4      1
```

C. Menentukan X_A , G_A dan A_A

```
> # Variabel aktif dan data yang sesuai
> X_A_name <- colnames(X)[max_c_hat_index]
> X_A_data <- X[, max_c_hat_index]
> cat("Variabel aktif:", X_A_name, "\n")
Variabel aktif: X1
> cat(X_A_data, sep = "\n")
1.468697
1.099678
0.7060569
0.3288372
0.1566282
0.1976303
-0.1795894
-0.9340288
-1.508059
-1.33585
> # Hitung G_A = X_A' X_A
> G_A <- t(as.matrix(X_A_data)) %% as.matrix(X_A_data)
> # Hitung invers dari G_A
> G_A_inv <- solve(G_A)
> # Buat vektor 1_A sebanyak jumlah variabel aktif (saat ini baru 1 variabel, jadi 1x1)
> one_A <- matrix(1, nrow = 1, ncol = 1)
> # Hitung A_A = (1_A' G_A_inv 1_A)^(-1/2)
> A_A <- as.numeric((t(one_A) %% G_A_inv %% one_A)^(-0.5))
> # Tampilkan hasil
> cat("G_A:\n")
G_A:
> print(G_A)
      [,1]
[1,]      9
> cat("\nG_A_inv:\n")

G_A_inv:
> print(G_A_inv)
      [,1]
[1,] 0.1111111
> cat("\nA_A:\n")

A_A:
> print(A_A)
[1] 3
```

D. Menghitung Nilai Vektor *Equiangular*

```
> # Hitung w_A
> one_A <- matrix(1, nrow = ncol(G_A), ncol = 1) # Matriks 1_A dengan ukuran yang sesuai
> w_A <- A_A * G_A_inv %*% one_A
> # Tampilkan hasil w_A
> cat("\n w_A:\n")

w_A:
> print(w_A)
      [,1]
[1,] 0.3333333
> # Hitung u_A
> u_A <- X_A_data %*% w_A
> # Tampilkan hasil u_A
> cat("\n u_A:\n")

u_A:
> print(u_A)
      [,1]
[1,] 0.48956560
[2,] 0.36655917
[3,] 0.23535231
[4,] 0.10961240
[5,] 0.05220940
[6,] 0.06587678
[7,] -0.05986313
[8,] -0.31134295
[9,] -0.50268629
[10,] -0.44528329
```

E. Menghitung *Inner Product*

```
> # Hitung a = X_A' * u_A
> a <- t(X_A_data) %*% u_A
> # Tampilkan hasil a
> cat("\n a (X_A' * u_A):\n")

a (X_A' * u_A):
> print(a)
      [,1]
[1,] 3
```

F. Menghitung Nilai $\hat{\gamma}$

```
> # Hitung a_j untuk variabel yang belum aktif
> remaining_indices <- setdiff(1:ncol(X), max_c_hat_index)
> a_j <- apply(X[, remaining_indices], 2, function(xj) as.numeric(t(xj) %*% u_A))
> print("Nilai a_j (Xj' * u_A) untuk variabel yang belum aktif:")
[1] "Nilai a_j (Xj' * u_A) untuk variabel yang belum aktif:"
> print(a_j)
      X2      X3      X4
2.757369 2.389282 2.919260
> # Ambil nilai c_j untuk variabel yang belum aktif dari c_hat
> c_j_remain <- c_hat[remaining_indices]
> # Hitung gamma_1 dan gamma_2
> gamma_1 <- (C_hat - c_j_remain) / (A_A - a_j)
> gamma_2 <- (C_hat + c_j_remain) / (A_A + a_j)
> # Tampilkan gamma_1 dan gamma_2
> print("Nilai gamma_1:")
[1] "Nilai gamma_1:"
> print(gamma_1)
      X2      X3      X4
0.7495257 2.1542800 5.8788943
> print("Nilai gamma_2:")
[1] "Nilai gamma_2:"
> print(gamma_2)
      X2      X3      X4
2.887794 2.874650 2.759347
> # Gabungkan semua nilai gamma dan ambil yang positif terkecil
> gamma_all <- c(gamma_1, gamma_2)
> gamma_all_pos <- gamma_all[gamma_all > 0]
> gamma_j <- min(gamma_all_pos)
> cat("Nilai gamma yang dipilih (minimum positif):", gamma_j, "\n")
Nilai gamma yang dipilih (minimum positif): 0.7495257
```

G. Menghitung $\hat{\mu}_A$

```
> #Hitung mu baru
> # Karena mu_hat_A = 0 pada langkah pertama, maka:
> mu_A <- rep(0, nrow(X)) # Vektor nol sepanjang jumlah observasi
> # Hitung mu_Aplus
> mu_Aplus <- mu_A + as.numeric(gamma_j) * u_A
> # Tampilkan hasil
> cat("\nHat mu_Aplus:\n")
```

```
Hat mu_Aplus:
> print(mu_Aplus)
      [,1]
[1,] 0.36694200
[2,] 0.27474552
[3,] 0.17640260
[4,] 0.08215731
[5,] 0.03913228
[6,] 0.04937634
[7,] -0.04486895
[8,] -0.23335954
[9,] -0.37677629
[10,] -0.33375127
```

H. Nilai calon koefisien LASSO

```
> # Hitung koefisien beta untuk langkah ini (y * w_A)
> s_j <- 1 # Hanya satu nilai s_j diambil, yaitu 1
> beta <- gamma_j * s_j * w_A
> cat("Koefisien beta pada langkah pertama:\n")
Koefisien beta pada langkah pertama:
> print(beta)
      [,1]
[1,] 0.2498419
```

I. Nilai Semua Koefisien pada Regresi LASSO

```
> # 3. Jalankan algoritma LARS dengan metode LASSO
> library(lars)
> lasso_model <- lars(X, Y, type = "lasso")
> # 4. Tampilkan koefisien LASSO pada setiap langkah
> cat("Koefisien LASSO pada setiap langkah algoritma LARS:\n")
Koefisien LASSO pada setiap langkah algoritma LARS:
> print(lasso_model$beta)
```

	X1	X2	X3	X4
0	0.0000000	0.0000000	0.00000000	0.0000000
1	0.2498419	0.0000000	0.00000000	0.0000000
2	0.5865269	0.3366850	0.00000000	0.0000000
3	0.5973010	0.3408444	0.01239943	0.0000000
4	1.2804893	0.2081409	0.17446217	-0.6956909

```
attr(,"scaled:scale")
[1] 3 3 3 3
```

Lampiran 7. Validasi Silang

```
> # 5. Validasi silang (Cross-validation)
> set.seed(123)
> cv_step <- cv.lars(X, Y, K = 10, type = "lasso", mode = "step", plot.it = TRUE)
> # 6. Tentukan langkah terbaik
> fractions <- seq(0, 1, length.out = length(cv_step$cv))
> best_step <- which.min(cv_step$cv)
> best_fraction <- fractions[best_step]
> best_cv_error <- cv_step$cv[best_step]
> best_coef <- lasso_model$beta[best_step, ]
> # 7. Tampilkan hasil sesuai urutan yang diinginkan
> cat("\nLangkah terbaik adalah langkah ke-", best_step, "\n")

Langkah terbaik adalah langkah ke- 3
> cat("Nilai s (fraction) pada langkah terbaik:", round(best_fraction, 4), "\n")
Nilai s (fraction) pada langkah terbaik: 0.5
> cat("Nilai CV error (MSE) pada setiap langkah:\n")
Nilai CV error (MSE) pada setiap langkah:
> print(cv_step$cv)
[1] 1.1111111 0.8134334 0.2397145 0.2425746 0.4819177
> cat("Cross-Validation Error (MSE) pada langkah terbaik:", best_cv_error, "\n")
Cross-Validation Error (MSE) pada langkah terbaik: 0.2397145
```

Lampiran 8. Model Awal Regresi Linier Berganda dengan OLS yang Telah Distandarisasi

```
> Regresi <- lm(Y ~ X1 + X2 + X3 + X4, data = df_std)  
> Regresi
```

Call:

```
lm(formula = Y ~ X1 + X2 + X3 + X4, data = df_std)
```

Coefficients:

(Intercept)	X1	X2	X3	X4
2.240e-15	1.280e+00	2.081e-01	1.745e-01	-6.957e-01

Lampiran 9. Model Terbaik Regresi LASSO dan Nilai Koefisien Determinasi

```
> # 9. Tampilkan koefisien terbaik (pada langkah terbaik)
> cat("\nKoefisien LASSO pada langkah terbaik:\n")

Koefisien LASSO pada langkah terbaik:
> print(best_coef)
      X1      X2      X3      X4
0.5865269 0.3366850 0.0000000 0.0000000
> # 10. Hitung R-squared ( $R^2$ ) dari model terbaik
> # Prediksi Y dengan X dan koefisien terbaik
> Y_hat <- X %%% best_coef # Y prediksi
> # Hitung SSres dan SStot pada data terstandarisasi
> SSres <- sum((Y - Y_hat)^2)
> SStot <- sum((Y - mean(Y))^2) # mean(Y) = 0 karena distandarisasi
> R2 <- 1 - (SSres / SStot)
> cat("\nNilai R-squared ( $R^2$ ) dari model terbaik:", round(R2, 4), "\n")

Nilai R-squared ( $R^2$ ) dari model terbaik: 0.8902
~ |
```

Lampiran 10. Nilai Intersep pada Model Regresi LASSO

A. Nilai rata-rata Y , X_1 , X_2 , X_3 , dan X_4

```
> # Hitung rata-rata dari data standar
> Y_bar <- mean(df_std$Y)
> X1_bar <- mean(df_std$X1)
> X2_bar <- mean(df_std$X2)
> X3_bar <- mean(df_std$X3)
> X4_bar <- mean(df_std$X4)
> # Tampilkan hasil
> cat("Rata-rata Y :", Y_bar, "\n")
Rata-rata Y : -2.802879e-16
> cat("Rata-rata X1 :", X1_bar, "\n")
Rata-rata X1 : 3.502732e-15
> cat("Rata-rata X2 :", X2_bar, "\n")
Rata-rata X2 : 2.761463e-16
> cat("Rata-rata X3 :", X3_bar, "\n")
Rata-rata X3 : -3.219647e-15
> cat("Rata-rata X4 :", X4_bar, "\n")
Rata-rata X4 : 9.331403e-15
```

B. Nilai Intersep pada Model Regresi LASSO

```
> # Ambil koefisien pada langkah terbaik
> best_step <- 3
> best_coef <- lasso_model$beta[best_step, ]
> # Karena semua X dan Y sudah distandarisasi, maka:
> mean_Y_std <- mean(df_std$Y)
> mean_X_std <- colMeans(df_std[, -1])
> # Hitung beta_0 (intersep pada data standar)
> beta0_std <- mean_Y_std - sum(best_coef * mean_X_std)
> cat("Intersep ( $\beta_0$ ) dengan data standar:", beta0_std, "\n")
Intersep ( $\beta_0$ ) dengan data standar: -2.427729e-15
```

Lampiran 11. Nilai VIF Setelah Menggunakan Regresi LASSO

```
> # 11. Periksa multikolinearitas dengan VIF hanya pada variabel terpilih
> library(car)
> # Ambil nama variabel yang digunakan (koefisien ≠ 0)
> selected_vars <- names(best_coef[best_coef != 0])
> # Subset data standar hanya pada variabel terpilih
> X_selected <- df_std[, selected_vars]
> # Buat model OLS untuk menghitung VIF (Y sebagai fungsi dari variabel terpilih)
> ols_model <- lm(Y ~ ., data = cbind(Y = df_std$Y, X_selected))
> # Hitung VIF
> vif_values <- vif(ols_model)
> # Tampilkan hasil VIF
> cat("\nNilai VIF untuk variabel yang dipilih oleh LASSO:\n")
```

Nilai VIF untuk variabel yang dipilih oleh LASSO:

```
> print(vif_values)
      X1      X2
6.442763 6.442763
```

DAFTAR RIWAYAT HIDUP

A. Identitas Diri

1. Nama Lengkap : Della Setya Rahma
2. Tempat & Tgl. Lahir : Wonosobo, 26 Juni 2003
3. Alamat Rumah : Kemitug RT 02 RW 02,
Tirip, Kec. Wadaslintang,
Kab. Wonosobo, Jawa Tengah
4. HP : 083112077873
5. E-mail : 2108046026@student.walisongo.ac.id,
rahmadella033gmail.com

B. Riwayat Pendidikan

1. Mi Al-Fatah Kemitug (2009-2015)
2. MTs Al-Fatah Kemitug (2015-2018)
3. SMKN 1 Wadaslintang (2018-2021)
4. UIN Walisongo Semarang (2021-2025)

Semarang, 20 Juni 2025

Della Setya Rahma

NIM: 2108046026