

SKRIPSI

**OPTIMISASI ANALISIS REGRESI LASSO PADA KASUS
KRIMINALITAS DI INDONESIA TAHUN 2021 DENGAN
ALGORITMA LARS MENGGUNAKAN AIC**



LINGGAR UTAMI

20106010003

**STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA**

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2024

**OPTIMISASI ANALISIS REGRESI LASSO PADA KASUS
KRIMINALITAS DI INDONESIA TAHUN 2021 DENGAN
ALGORITMA LARS MENGGUNAKAN AIC**

Skripsi

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



diajukan oleh

LINGGAR UTAMI

20106010003

Kepada

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

2024



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir
Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

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

Nama : Linggar Utami
NIM : 20106010003
Judul Skripsi : Optimisasi Analisis Regresi LASSO Pada Kasus Kriminalitas di Indonesia Tahun 2021 dengan Algoritma LARS Menggunakan AIC

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 1 Maret 2024

Pembimbing

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

NIP. 19741003 200003 2 002



PENGESAHAN TUGAS AKHIR

Nomor : B-426/Un.02/DST/PP.00.9/03/2024

Tugas Akhir dengan judul : Optimisasi Analisis Regresi LASSO pada Kasus Kriminalitas di Indonesia Tahun 2021 menggunakan Algoritma LARS dengan AIC

yang dipersiapkan dan disusun oleh:

Nama : LINGGAR UTAMI
Nomor Induk Mahasiswa : 20106010003
Telah diujikan pada : Jumat, 08 Maret 2024
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

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

Valid ID: 65f26d08e8f6f



Penguji I

Lilih Deva Martias, M.Sc.
SIGNED

Valid ID: 65f26ad56f8fc



Penguji II

Pipit Pratiwi Rahayu, S.Si., M.Sc.
SIGNED

Valid ID: 65f13368b1d72



Yogyakarta, 08 Maret 2024
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 65f2aa79a3840

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Linggar Utami
NIM : 20106010003
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 29 Februari 2024


Linggar Utami

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

KARYA SEDERHANA INI PENULIS PERSEMBAHKAN UNTUK:

KEDUA ORANG TUA PENULIS, KAKAK PENULIS,
KELUARAGA BESAR PENULIS, YANG TELAH MEMBERIKAN DO'A, KASIH,
DAN CINTA NYA YANG TAK TERBATAS.
SAHABAT TERCINTA, KE-13 MEMBER SEVENTEEN, KUCING-KUCING
TERSAYANG DAN SEMUA ORANG-ORANG YANG PERNAH SAYA KENAL.
ALMAMATER KAMPUS TERCINTA YANG BANYAK MEMBERIKAN
WARNA DALAM HIDUP SELAMA 4 TAHUN UNIVERSITAS ISLAM NEGERI
SUNAN KALIJAGA YOGYAKARTA.

"MATEMATIKA ANGGKATAN 2020 TERCINTA"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

”Terjebak diruntinitas yang tidak kita sukai akan membunuh kita secara perlahan.”

(Fiersa Besari)

”Jangan benci hanya karena terabaikan, Jangan terobsesi hanya karena rasa nyaman.”

(Holing)

”Lebih baik mencoba ketika masih muda, walaupun gagal, masih bisa diperbaiki”

(Arsal Bahtiar)

”Hidup adalah pilihan dan setiap pilihan memiliki resikonya sendiri. Kita berhak memilih, dan wajib bertanggung jawab atas pilihan kita”

PRAKATA

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan karuniannya yang tak ternilai harganya berupa keimanan, kesabaran, kekuatan dan kelancaran. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Optimisasi Analisis Regresi LASSO pada kasus kriminalitas Di Indonesia Tahun 2021 Dengan Algoritma LARS menggunakan AIC"

Mengenai skripsi ini, penulis ingin mengucapkan terima kasih kepada pembimbing skripsi penulis, Ibu Sri Utami Zuliana, S.Si.,M.Sc.,Ph.D. Pertemuan rutin tiap satu kali sepekan memungkinkan penulis dalam mengenali progres pengerjaan skripsi ini. Bimbingan dan komentarnya mencegah penulis memilih cara yang lebih mudah. Selama menempuh perkuliahan ini penulis sangat berterima kasih kepada orang tua dan keluarga besar yang telah memberikan kasih sayang dan dukungan yang luar biasa. Tanpa mengurangi rasa hormat, penulis juga ingin menyampaikan ucapan terima kasih dan apresiasi yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Phil Al Makin, MA., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Bapak Muchammad Abrori, S.Si.,M.Kom., selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

4. Bapak Dr. Muhammad Wakhid Musthofa, S.Si., M.Si., selaku dosen pembimbing akademik yang telah banyak memberikan pengarahan kepada penulis dari awal perkuliahan hingga akhir.
5. Seluruh dosen program studi Matematika dan staf fakultas Sains dan Teknologi yang senantiasa memberikan ilmu dan layanan terbaik kepada penulis dari awal hingga akhir perkuliahan.
6. Kedua orang tua, kakak, dan keluarga besar penulis yang telah memberikan kasih sayang, doa dan dukungannya sehingga penulis sampai ke titik ini.
7. Keluarga Matematika angkatan 2020 yang telah kebersamai selama 4 tahun dibangu perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.
8. Teman-teman seperbimbingan Anisa, Adel, Fatma, Sari, Uswatun, Yumna yang telah bersama-sama selama bimbingan.
9. Teman-teman kuliah yang sudah mau menerima saya, mengizinkan saya ikut mengobrol dan makan bersama.
10. Teman-teman KKN 111 Ketelo terima kasih cerita indahya 45 hari selama KKN semoga kita kelak bisa kembali bersama menikmati waktu di Ketelo.
11. Teman-teman Novita Farasmawasti dan Erna Saraswati terima kasih sudah berada akasih banget udah bertahan disisi saya, sudah mau menjadi saudara perempuan saya, sudah bersama-sama ngelewatn masa sulit.
12. Untuk Ibu Yanuanita Widiyaningrum, M.Pd., guru tercinta semasa madrasah, terima kasih atas bimbingan dan doanya selama ini.

13. Buat member Seventeen yang saya cintai, terima kasih sudah hadir sebagai grup favorit saya yang sudah menghibur selama masa-masa sulit saya dan terima kasih sudah menciptakan karya-karya yang luar biasa.
14. Terakhir semua pihak yang penulis tidak bisa sebutkan satu-persatu, semoga dimudahkan segala urusannya.

Penulis sangat menyadari tidak ada yang sempurna di dunia ini. Oleh karena itu segala bentuk saran, kritik atas skripsi ini penulis menerimanya dengan senang hati. Penulis berharap skripsi ini dapat memberikan manfaat khususnya bagi penulis dan semua kalangan yang membutuhkan untuk dijadikan referensi dalam menulis. Akhir kata semoga Allah SWT melimpahkan segala rahmat dan hidayah-Nya kepada kita semua, *Aamiin*.

Yogyakarta, 22 Februari 2024

Linggar Utami

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASILAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMBANG	xv
INTISARI	xvii
ABSTRACTxviii
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	7
1.3. Batasan Masalah	7
1.4. Tujuan Penelitian	8
1.5. Manfaat Penelitian	8
1.6. Sistematika Penulisan	8
1.7. Tinjauan Pustaka	9
II DASAR TEORI	13
2.1. Analisi regresi	13

2.1.1.	Analisis Regresi Linear sederhana	13
2.1.2.	Analisis Regresi Linear Berganda	14
2.2.	<i>Maximum Likelihood Estimation</i> (MLE)	15
2.3.	Multikolinearitas	16
2.4.	Regresi Terpenalti	19
2.4.1.	Regresi Ridge (L_2)	20
2.4.2.	Regresi Lasso (L_1)	21
2.5.	<i>Least Absolute Shrinkage and selection Operator</i> (LASSO)	21
2.6.	Algoritma <i>Least Angle Regression</i> (LARS)	23
2.7.	<i>Akaike Information Criterion</i> (AIC)	25
III	METODE PENELITIAN	26
3.1.	Metode Penelitian	26
3.2.	Jenis dan Sumber Literatur	26
3.3.	Studi Kasus	27
3.3.1.	Variabel Data Penelitian	27
3.4.	Software yang Digunakan	29
3.5.	Langkah-langkah Analisis	29
3.6.	Source code	30
3.7.	Diagram Analisis Data (<i>Flowchart</i>)	31
IV	HASIL DAN PEMBAHASAN	32
4.1.	Eksplorasi Data	32
4.2.	Statistik Deskriptif	33
4.3.	Standarisasi Data (<i>Z-Score</i>)	35
4.4.	Analisis Regresi Linear Berganda	35
4.5.	Uji Hipotesis	37
4.5.1.	Uji F (Uji Serentak)	37

4.5.2. Uji t (Uji Parsial)	38
4.6. Uji Asumsi Klasik	41
4.6.1. Uji Normalitas	42
4.6.2. Uji Autokorelasi	43
4.6.3. Uji Heteroskedastisitas	45
4.6.4. Uji Multikolinearitas	46
4.7. <i>Least Absolute Shrinkage and Selection Operator Operator (LASSO)</i> dengan Algoritma LARS	47
4.8. Menentukan Nilai Lambda Optimal dengan AIC	51
4.9. Pemilihan Model Terbaik regresi LASSO	53
V PENUTUP	56
5.1. Kesimpulan	56
5.2. Saran	57
DAFTAR PUSTAKA	57
LAMPIRAN	61
A DATA PENELITIAN	61
B SOURCE CODE PROGRAM R	65
C TABEL TINJAUAN PUSTAKA	72
D BIOGRAFI PENULIS	74

DAFTAR TABEL

3.1	Variabel Data Penelitian	28
4.1	Statistik Deskriptif	34
4.2	Nilai <i>p-value</i> Setiap Variabel	38
4.3	Hasil uji VIF Data Kriminalitas di Indonesia Tahun 2021	46
4.4	Tahapan-Tahapan Masuknya Variabel Bebas pada Regresi LASSO	48
4.5	Nilai Lambda	50
4.6	Nilai Lambda	52
4.7	Model Regresi LASSO dengan AIC (λ)	53
4.8	Koefisien Regresi LASSO Berdasarkan AIC	54

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

1.1	Jumlah Kejahatan dan Jumlah Risiko kejahatan, 2019-2021	5
2.1	Perbandingan Kondisi Distribusi β	19
3.1	Flowchart Penelitian	31
4.1	Ilustrasi Plot Variabel Independen dengan Variabel Dependen	32
4.2	Plot Autokorelasi Data Kriminalitas di Indonesia Tahun 2021	45
4.3	Ilustrasi Plot Tahapan Variabel Bebas kedalam Model	47
4.4	Plot Nilai AIC dan Lambda	53



DAFTAR LAMBANG

Y	= Variabel dependen
X_n	= variabel independen
β_1, \dots, β_n	= parameter regresi
ε	= error
X	= matrix variabel independen
β	= vektor parameter regresi ukuran $(k+1) \times 1$
L	= Kuadrat residual
$\varepsilon_i^T \varepsilon_i$	= perkalian matriks <i>error</i>
$\hat{\beta}$	= vektor koefisien
$X'X$	= matriks korelasi peubah X
n	= Jumlah observasi
p	= Banyaknya variabel independen di tambah intersep
σ^2	= Varians <i>error</i>
$Var(b_{k-1})$	= Varians b_{k-1}
r_{XX}	= Korelasi antar variabel independen
R_j^2	= koefisien determinasi antar X_j dengan variabel yang lain.
j	= banyaknya model regresi.
$f(t_i)$	= fungsi <i>spline</i> dengan pendekatan titik knot

- λ = parameter penghalus
 D = matriks semi definit positif=diag $(0_{p+1}, 1_k)$
 $\sum_{i=1}^n [y_i - f(t)]^2$ = jumlah kuadrat *error*
 $\lambda\beta^T D\beta$ = fungsi terpenalti
 $w_j (j = 1, 2, \dots, p)$ = bobot yang dikenalan regularisasi pada koefisien regresi j
 ε^{**} = vektor *error* ukuran $(n \times 1)$
 t = nilai parameter tuning dengan $t \leq 0$
 k = jumlah parameter yang diestimasi dalam model regresi
 θ = Parameter yang diestimasi

INTISARI

OPTIMISASI ANALISIS REGRESI LASSO PADA KASUS KRIMINALITAS DI INDONESIA TAHUN 2021 DENGAN ALGORITMA LARS MENGGUNAKAN AIC

LINGGAR UTAMI

20106010003

Multikolinearitas merupakan permasalahan umum dalam analisis regresi, dimana beberapa variabel independen mempunyai korelasi yang tinggi antar variabelnya. Identifikasi multikolinearitas menggunakan nilai *Variance Inflation Factor* (VIF) atau pemeriksaan koefisien korelasi antar variabel dalam suatu model regresi. Permasalahan multikolinearitas memiliki dampak signifikan, termasuk peningkatan variansi dan timbulnya bias, yang pada akhirnya dapat menyebabkan overfitting dan menurunkan signifikansi koefisien. Oleh karena itu, penanganan multikolinearitas menjadi suatu aspek kritis untuk memastikan stabilitas dan akurasi model regresi yang dikembangkan. Untuk mengatasi permasalahan tersebut menggunakan metode LASSO (*Least Absolute Shrinkage and Selection Operator*). Regresi LASSO menggunakan metode penyusutan koefisien variabel independen efektif mengarahkan koefisien variabel yang memiliki korelasi tinggi mendekati atau mencapai nilai nol. Pencarian koefisien regresi LASSO dilakukan dengan menggunakan algoritma *Least Absolute Regression* (LARS). Dalam penelitian ini, pemilihan lambda optimal pada regresi LASSO menggunakan kriteri *Akaike's Information Criterion* (AIC). Penelitian ini diaplikasikan pada data Kriminalitas di Indonesia tahun 2021. Kesimpulan yang diperoleh berdasarkan penelitian yang telah dilakukan terdapat satu dari tujuh variabel independen yang masuk ke dalam model regresi LASSO yaitu jumlah penduduk di Indonesia (X_2) dimana variabel tersebut berpengaruh terhadap angka kriminalitas di Indonesia tahun 2021.

Kata kunci: Multikolinearitas, LASSO, LARS, AIC, Kriminalitas

ABSTRACT

OPTIMIZATION OF LASSO REGRESSION ANALYSIS ON CRIME CASES IN INDONESIA IN 2021 WITH THE LARS ALGORITHM USING AIC

LINGGAR UTAMI

20106010003

Multicollinearity is a common problem in regression analysis, where several independent variables have a high correlation between their variables. Identification of multicollinearity uses the Value of Variance Inflation Factor (VIF) or an examination of the correlation coefficient between variables in a regression model. Multicollinearity problems have significant impacts, including an increase in variance and the incidence of bias, which can ultimately lead to overfitting and lower coefficient significance. Therefore, addressing multicollinearity is a critical aspect to ensure the stability and accuracy of the regression model developed. To overcome these problems, the LASSO method (Least Absolute Shrinkage and Selection Operator) is used. LASSO regression uses a method of shrinking the coefficients of independent variables effectively directing the coefficients of variables that have a high correlation to approach or reach zero values. The search for LASSO regression coefficients is carried out using the *Least Absolute Regression* (LARS) algorithm. In this study, the selection of optimal lambda in LASSO regression uses the Akaike's Information Criterion (AIC). This research is applied to crime data in Indonesia in 2021. The conclusion obtained based on the research that has been done is that there is one of the seven independent variables included in the LASSO regression model, namely the total population in Indonesia (X_2) where this variable affects the crime rate in Indonesia in 2021.

Key words: Multicollinearity, LASSO, LARS, AIC, Crime

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Analisis regresi merupakan metode statistika yang digunakan untuk memahami dan menjelaskan hubungan antara satu atau lebih variabel independen (bebas) dengan variabel dependen (tak bebas). Pada analisis regresi dibedakan menjadi dua jenis yaitu, analisis regresi linear sederhana dan analisis regresi linear berganda. Regresi linear sederhana merupakan analisis regresi dengan tepat satu variabel independen dan satu variabel dependen, sedangkan analisis regresi berganda merupakan analisis regresi dengan satu variabel dependen dan lebih dari satu variabel independen. Hasil analisis regresi linear dapat digunakan untuk membuat prediksi, mengidentifikasi faktor-faktor yang mempengaruhi variabel dependen, serta mengukur kekuatan dan arah hubungan antara variabel-variabel yang diteliti. Pada analisis regresi metode yang dapat digunakan untuk menentukan estimasi terhadap parameter-parameter yaitu Metode Maksimum Likelihood (*Maximum Likelihood Estimation*). Pada metode tersebut perlu adanya empat asumsi BLUE (Best, Linear, Unbiased, and Estimator)(Andana et al., 2017).

Maksimum Likelihood (*Maximum Likelihood Estimation* atau MLE) merupakan salah satu metode dalam statistika yang bertujuan untuk memaksimalkan fungsi likelihood untuk mengestimasi parameter dari suatu model statistika dengan probabilitas maksimum. Metode ini menggunakan nilai-nilai di dalam ruang parameter untuk mengestimasi nilai parameter yang tidak diketahui. Dalam konteks

ini, MLE digunakan untuk mendapatkan estimasi parameter populasi yang tidak diketahui berdasarkan data sampel. Proses ini esensial karena memungkinkan kita untuk menyusun perkiraan yang optimal terhadap nilai-nilai yang mungkin di dalam model statistika yang digunakan. Asumsi Klasik adalah serangkaian prasyarat yang diperlukan untuk mendapatkan estimasi parameter yang optimal dan menghasilkan statistik inferensial yang baik. Untuk memenuhi asumsi klasik tersebut perlu adanya uji asumsi klasik dimana terdapat empat uji asumsi, yaitu uji normalitas, uji heterokedasitas, uji autokorelasi, dan uji multikolinearitas. Pada analisis regresi juga sering kali terdapat model yang tidak memenuhi uji asumsi salah satunya adalah uji multikolinearitas, dimana hal ini disebabkan karena nilai koefisien yang tidak stabil. Multikolinearitas merupakan kondisi dimana terdapat korelasi antar variabel bebas sehingga variabel tersebut tidak bersifat *orthogonal*. suatu variabel bebas dikatakan *orthogonal* apabila variabel bebas memiliki nilai korelasi antar variabel sama dengan nol. Cara menangani penyimpangan asumsi ini adalah dengan menggunakan metode analisis regresi *Least Absolute Shrinkage and Selection Operator (LASSO)* (Purba, 2021).

Regresi terpenalti, atau sering disebut *penalized regression*, adalah pendekatan dalam statistika yang mengintegrasikan elemen regularisasi ke dalam model regresi. Tujuan utama regresi terpenalti adalah untuk mengendalikan overfitting dan meningkatkan kinerja model dengan menambahkan penalti pada fungsi objektif. Menurut Pekhimenko (2006) regresi terpenalti yang umum digunakan terdapat dua jenis yaitu regresi L_1 (Lasso) dan regresi L_2 (Ridge). Pendekatan ini membantu dalam memilih dan mempertahankan variabel yang penting, sambil mengurangi dampak variabel yang kurang relevan atau berkontribusi kecil terhadap prediksi.

Metode *Least Absolute Shrinkage and Selection Operator (LASSO)* meru-

pakan pengembangan dari analisis regresi ridge yang pertama kali dikenalkan oleh Tibshirani pada tahun 1996. Regresi LASSO digunakan untuk mengatasi permasalahan multikolinearitas, dimana regresi LASSO dapat menyusutkan koefisien regresi dari variabel bebas berkorelasi tinggi dengan *error* mendekati nol atau tepat pada nol (Tibshirani, 1996). Penalti penyusutan pada metode LASSO menyebabkan terjadinya penyusutan nilai estimasi koefisien parameter, sehingga variabel bebas yang berpengaruh terhadap model akan tetap dimasukkan ke dalam model regresi, sedangkan variabel bebas yang kurang berpengaruh estimasi koefisien parameter akan disusutkan sampai nol dan diseleksi dari model hingga model menjadi lebih efisien. Regresi LASSO juga berperan sebagai seleksi variabel-variabel sehingga menghasilkan model dengan variabel terbaik. Menurut Tibshirani (1996) regresi LASSO juga dapat digunakan pada data yang kontinu dan memerlukan variabel independen yang berdistribusi normal baku. Perhitungan yang tepat pada metode regresi LASSO dapat menggunakan algoritma *Least Angle Regression (LAR)*.

Algoritma *Least Angle Regression (LAR)* pertama kali ditemukan oleh Bradley Efron dkk pada tahun 2002 (Robbani et al., 2019). Algoritma LAR merupakan sebuah algoritma untuk menghasilkan model linear yang baik pada analisis regresi. Algoritma LARS (*Least Angle Regression*) pada regresi Lasso adalah metode komputasional yang digunakan untuk menentukan koefisien model regresi Lasso secara iteratif dengan memilih dan menambahkan variabel yang memiliki korelasi paling tinggi dengan residu. Modifikasi algoritma LAR disebut dengan algoritma LARS. Memodifikasi algoritma LAR dapat memberikan taksiran koefisien pada regresi LASSO dan memiliki langkah yang lebih efisien dibandingkan regresi LASSO itu sendiri (Hastie et al., 2009). Pada algoritma ini dimulai dengan menyetarakan koefisien sama dengan nol.

Salah satu metode yang bisa digunakan untuk menentukan model regresi terbaik adalah metode *Akaike Information Criterion (AIC)*. AIC dikembangkan oleh Hirotugu Akaike pada tahun 1974 dan menjadi salah satu kriteria yang paling umum digunakan dalam pemilihan model dan didasarkan pada *Maximum likelihood Estimation (MLE)*. Tujuan utama AIC adalah meminimalkan kesalahan prediksi model dengan mempertimbangkan kualitas model dan kompleksitasnya. AIC mencoba mencapai keseimbangan antara kesesuaian model terhadap data (kesalahan penjelasan) dan jumlah parameter yang digunakan dalam model (kesalahan penyesuaian)(Fathurahman, 2010).

Kriminalitas adalah salah satu isu sosial yang signifikan di seluruh dunia, termasuk di Indonesia. Kriminalitas merujuk pada serangkaian tindakan yang dianggap melanggar hukum dan norma-norma sosial dalam suatu masyarakat. Dapat mencakup berbagai kejahatan, mulai dari pelanggaran kecil hingga kejahatan yang serius. Definisi kriminalitas dapat bervariasi tergantung pada sistem hukum dan nilai-nilai yang dianut oleh masyarakat tersebut. Tingkat kriminalitas yang tinggi dapat mengancam keamanan masyarakat, mengganggu ketertiban umum, dan memiliki dampak ekonomi yang merugikan. Oleh karena itu, pemahaman yang lebih tentang faktor-faktor yang mempengaruhi tindak kriminal sangat penting untuk merancang kebijakan yang efektif dalam memitigasi dan mengendalikan tindakan kriminal. Tindak kriminal juga bisa melibatkan tindakan yang merugikan individu atau kelompok, seperti penipuan, penganiayaan, atau tindak kejahatan terhadap properti.

Tingkat kriminalitas dipengaruhi oleh sejumlah faktor kompleks. Termasuk di antaranya adalah kondisi sosial ekonomi, tingkat pendidikan, ketidaksetaraan, dan kondisi psikologis. Selain itu, lingkungan tempat tinggal dan faktor so-

sial seperti pekerjaan dan pendapatan juga dapat memainkan peran penting dalam membentuk perilaku kriminal. Dari kondisi itulah akan muncul tindakan-tindakan kriminalitas yang dapat merugikan berbagai pihak karena ketidakmampuan dalam meningkatkan kesejahteraan (Ningsih et al., 2023).



Gambar 1.1 Jumlah Kejahatan dan Jumlah Risiko kejahatan, 2019-2021

Berdasarkan gambar 1.1 pada tahun 2019, tercatat sebanyak 269.324 kejadian kejahatan. Angka ini menurun menjadi 247.218 kejadian pada tahun 2020, dan pada tahun 2021, jumlah kejadian kriminalitas mencapai 239.481 kejadian. Indikator tingkat kejahatan (crime rate) juga menurun selama periode 2019-2020, yaitu dari 103 pada tahun 2019 menjadi 94 pada tahun 2020, kemudian menurun lagi menjadi 90 pada tahun 2021. Pada tahun 2019 data survei yang mencerminkan persentase penduduk yang menjadi korban kejahatan dalam periode 2019-2021 juga menunjukkan pola penurunan yang serupa dengan data registrasi dari 1,01 persen menjadi 0,78 persen pada tahun 2020, kemudian kembali menurun menjadi 0,47 persen pada tahun 2021. Selama periode 2019-2021, persentase kasus yang dilaporkan ke polisi tidak pernah melebihi 25 persen. Pada tahun 2021, persentasenya mencapai 23,31 persen, sedikit mengalami kenaikan jika dibandingkan dengan tahun 2020 (23,46 persen) dan 2019 (22,19 persen).

Pada tahun 2021, Indonesia mengalami berbagai tantangan terkait dengan

kriminalitas, mulai dari tindak pencurian hingga kekerasan. Faktor-faktor yang berkontribusi terhadap tingkat kriminalitas di Indonesia pada tahun tersebut bisa sangat beragam, termasuk aspek sosial, ekonomi, demografi, dan geografis. Oleh karena itu, diperlukan sebuah penelitian yang cermat untuk mengidentifikasi variabel-variabel yang paling berpengaruh terhadap tingkat kriminalitas di Indonesia tahun 2021.

Dampak kriminalitas melibatkan lebih dari kerugian materi dan keamanan pribadi. Secara sosial, kriminalitas dapat menciptakan ketidakstabilan dan kekawatiran di dalam masyarakat. Terjadinya kejahatan dapat mengakibatkan hilangnya rasa keamanan, membatasi kebebasan individu, dan menciptakan lingkungan yang kurang kondusif untuk pertumbuhan ekonomi dan perkembangan masyarakat. Dalam ajaran Islam, Allah SWT menegaskan kepentingan untuk tidak mengarahkan diri sendiri menuju kebinasaan dan mendorong perilaku baik. Prinsip ini mencerminkan nilai-nilai moral Islam yang menekankan untuk menjauhi segala bentuk tindakan yang merugikan masyarakat dimana seperti yang tertuang dalam Al-Quran Surah Al-Baqarah ayat 195, sebagai berikut:

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

وَأَنْفِقُوا فِي سَبِيلِ اللَّهِ وَلَا تُلْقُوا بِأَيْدِيكُمْ إِلَى
التَّهْلُكَةِ وَأَحْسِنُوا إِنَّ اللَّهَ يُحِبُّ الْمُحْسِنِينَ ﴿١٩٥﴾

Artinya: "Dan belanjakanlah (harta bendamu) di jalan Allah, dan janganlah kamu menjatuhkan dirimu sendiri ke dalam kebinasaan, dan berbuat baiklah, karena sesungguhnya Allah menyukai orang-orang yang berbuat baik."

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengisi celah pe-

ngetahuan dengan mengoptimalkan analisis regresi Lasso dengan algoritma LARS menggunakan kriteria AIC untuk memahami faktor-faktor yang memengaruhi kriminalitas di Indonesia tahun 2021. Hasil penelitian diharapkan dapat memberikan pandangan yang lebih baik dan solusi yang lebih efektif dalam penanganan masalah kriminalitas di Indonesia.

1.2. Rumusan Masalah

Berdasarkan uraian pada latar belakang diperoleh rumusan masalah dari penelitian ini yaitu sebagai berikut:

1. Bagaimana menerapkan metode analisis regresi LASSO dengan algoritma LARS menggunakan kriteria AIC dalam mengatasi kemultikolinearitasan data regresi?
2. Bagaimana model terbaik yang dihasilkan berdasarkan regresi LASSO dengan algoritma LARS menggunakan kriteria AIC pada data kriminalitas di Indonesia tahun 2021?

1.3. Batasan Masalah

Agar fokus pada sasaran yang diharapkan maka perlu dilakukan pembatasan masalah dalam penelitian ini sebagai berikut:

1. Metodologi penelitian analisis regresi LASSO dengan algoritma Least Angle Regression (LARS).
2. Pemilihan model terbaik menggunakan kriteria AIC.
3. Data yang digunakan adalah data kriminalitas di Indonesia tahun 2021 menurut provinsi.

4. *Software* yang digunakan adalah R versi .4.3.1.

1.4. Tujuan Penelitian

1. Menerapkan metode regresi LASSO dengan Algoritma LARS menggunakan kriteria AIC untuk mengatasi kemultikolinearitasan pada data regresi.
2. Mengidentifikasi model terbaik dengan regresi LASSO dengan algoritma LARS menggunakan kriteria AIC pada data kriminalitas di Indonesia tahun 2021.

1.5. Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan manfaat dibidang akademik maupun non-akademi, serta dapat meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi kasus kriminalitas khususnya di Indonesia pada tahun 2021. Penelitian ini juga diharapkan dapat membantu mengoptimalkan model prediksi dengan regresi LASSO yang menggunakan algoritma LARS dan kriteria AIC untuk kasus kriminalitas di Indonesia kedepannya, serta penelitian ini dapat berguna dalam penelitian-penelitian berikutnya yang melibatkan analisis regresi LASSO menggunakan LARS dan pemilihan variabel bebas dalam suatu permasalahan.

1.6. Sistematika Penulisan

Sistematika pada penelitian ini yaitu:

BAB 1 PENDAHULUAN

Pada bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II DASAR TEORI

Pada bab ini menjelaskan terkait landasan teori yang menjadi acuan, berisi tentang

konsep, hipotesis dan rancangan penelitian.

BAB III METODE PENELITIAN

Pada bab ini membahas terkait metode penelitian yang akan digunakan dalam pemecahan masalah yang berisi tentang pendekatan penelitian, sumber data, metode pengolahan data dan diagram analisis data.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan dilakukan proses analisis data yang kemudian hasilnya diinterpretasikan secara terperinci.

BAB V PENUTUP

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

1.7. Tinjauan Pustaka

Pada penelitian yang dilakukan oleh (Mahalani & Rifai, 2022) dengan judul *Least Absolute Shrinkage And Selection Operator (LASSO) untuk Mengatasi Multikolinearitas pada Model Regresi Linear Berganda* membahas mengenai penerapan metode regresi *Least Absolute Shrinkage And Selection Operator (LASSO)* untuk mengatasi masalah multikolinearitas dalam model regresi linear berganda. Penelitian ini menerapkan metode regresi LASSO pada data jumlah penduduk miskin di Indonesia pada tahun 2020. Pada penelitian ini menghasilkan kesimpulan terdapat tiga variabel bebas yang menunjukkan masalah multikolinearitas pada studi kasus yaitu pengeluaran per kapita, indeks pembangunan manusia, dan rata-rata lama sekolah. Dengan menggunakan regresi LASSO, variabel produk domestik bruto menjadi variabel pertama yang masuk kedalam model regresi dikarenakan memiliki korelasi yang paling tinggi dengan *error*. Hal ini didasarkan pada analisis koefisien korelasi, hasil estimasi koefisien, dan nilai VIF menggunakan metode regresi LAS-

SO. Metode perhitungan yang digunakan dalam penelitian ini adalah metode regresi *Least Absolute Shrinkage And Selection Operator* (LASSO) dengan algoritma *Least Angle Regression* (LARS).

Penelitian yang dilakukan oleh (Rahmawati & Suratman, 2022) dengan judul *Performa Regresi Ridge dan Regresi LASSO pada Data dengan Multikolinearitas*. Metode yang digunakan pada penelitian ini adalah regresi ridge dan regresi LASSO dengan *mean squared error* (MSE). Hasil penelitian menunjukkan bahwa dalam mengatasi multikolinearitas pada model regresi linear berganda, regresi ridge memiliki *mean squared error* (MSE) yang lebih kecil dibandingkan regresi lasso. Akan tetapi, regresi lasso lebih unggul dalam seleksi variabel dan interpretasi model karena mampu mengurangi beberapa variabel bebas yang tidak signifikan. Dengan demikian, regresi lasso dapat memberikan interpretasi model yang lebih efisien.

Pada penelitian yang dilakukan oleh (Robbani et al., 2019) yang berjudul *Regresi Least Absolute Shrinkage and Selection Operator (LASSO) pada Kasus Inflasi di Indonesia Tahun 2014-2017*. Penelitian ini memaparkan penggunaan metode regresi LASSO dalam menganalisis inflasi di Indonesia dari tahun 2014 hingga 2017. Metode regresi LASSO digunakan untuk mengatasi masalah multikolinearitas dalam model regresi linear berganda dengan menyusutkan koefisien taksiran mendekati angka nol dan melakukan seleksi variabel independen. Algoritma LARS digunakan untuk memecahkan regresi LASSO secara komputasional. Hasil penelitian menunjukkan bahwa variabel-variabel bebas seperti Produk Domestik Bruto, Ekspor Bersih, Nilai Tukar Rupiah, Tingkat Bunga, Harga Beras, dan Upah Tenaga Kerja Petani memiliki dampak terhadap inflasi sebagai variabel terikat.

Penelitian yang dilakukan (Pardede et al., 2022) dengan judul *Penerapan Regresi Least Absolute Shrinkage and Selection Operator (LASSO) Untuk Meng-*

identifikasi Variabel yang Berpengaruh terhadap Kejadian Stunting di Indonesia. Pada penelitian ini menjelaskan tentang penggunaan metode regresi Least Absolute Shrinkage and Selection Operator (LASSO) untuk mengidentifikasi variabel yang mempengaruhi stunting di Indonesia. Penelitian ini bertujuan untuk mengatasi masalah multikolinearitas antara variabel independen dan memilih model terbaik untuk memprediksi stunting pada anak-anak. Hasil penelitian menunjukkan bahwa lima variabel, yaitu pemberian ASI eksklusif, konsumsi protein, imunisasi DPT-HB, tinggi badan ibu, dan diare, memiliki pengaruh signifikan terhadap stunting di Indonesia..

Penelitian yang dilakukan oleh (Altalbany, 2021) dengan judul *Evaluation of Ridge, Elastic Net and Lasso Regression Methods in Precedence of Multicollinearity Problem: A Simulation Study*. Penelitian ini membahas tentang perbandingan kinerja metode regresi Ridge, Lasso, dan Elastic Net dalam menangani multikolinearitas dalam analisis regresi berganda menggunakan data simulasi. Studi tersebut menemukan bahwa metode Elastic Net lebih unggul dalam memperkirakan koefisien regresi saat multikolinearitas rendah, sedang, dan tinggi untuk ukuran sampel apa pun. Metode Lasso adalah pengestimasi koefisien regresi yang paling akurat saat data mengandung multikolinearitas yang tinggi dengan ukuran sampel kurang dari 10.000 observasi. Artikel ini memberikan gambaran tentang dasar-dasar regresi, metode regresi Ridge, Lasso, dan Elastic Net, serta aplikasinya dalam menangani multikolinearitas.

Penelitian yang dilakukan oleh (Sartika et al., 2020) dengan judul *Analisis Regresi dengan Metode Least Absolute Shrinkage and Selection Operator (Lasso) dalam Mengatasi Multikolinearitas*. Penelitian ini membahas tentang penggunaan metode LASSO menggunakan algoritma LARS dalam analisis regresi untuk meng-

atasi masalah multikolinearitas. Penelitian ini menggunakan data tentang kasus pneumonia pada balita di Kota Pontianak dan Kabupaten Mempawah. Hasil analisis menunjukkan bahwa metode LASSO dapat mengatasi masalah multikolinearitas dan menghasilkan model regresi terbaik.

Berdasarkan beberapa penelitian yang telah dipaparkan, dapat ditarik kesimpulan bahwasanya banyaknya penggunaan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) menggunakan *Least Angle Regression* (LARS) sebagai penyelesaian masalah kemultikolinearitasan pada data regresi. Akan tetapi, banyak penelitian yang hanya sebatas penggunaan regresi LASSO dengan algoritma LARS saja untuk mengatasi kemultikolinearitasan pada data regresi dan belum banyak yang menggunakan *Akaike Information Criterion* (AIC) sebagai metode praduga parameter.

Dari yang telah diuraikan sebelumnya, maka akan dilakukan penelitian mengenai penyelesaian masalah kemultikolinearitasan pada data menggunakan regresi LASSO dengan algoritma LARS dan pemilihan parameternya menggunakan *Akaike Information Criterion* (AIC) pada data kriminalitas di Indonesia tahun 2021.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian terkait regresi LASSO dengan algoritma LARS pada data angka kriminalitas di Indonesia tahun 2021, maka dapat diambil kesimpulan sebagai berikut:

1. Multikolinearitas pada model regresi dapat dideteksi dengan cara melihat nilai $VIF > 10$ maka model terdapat multikolinearitas. Pada penelitian ini dari tujuh variabel independen, terdapat dua variabel yang terdeteksi multikolinearitas yaitu variabel X_2 dan variabel X_7 . Melalui implementasi metode analisis regresi LASSO dengan algoritma LARS diperoleh delapan tahapan LASSO, dimana pada tahap pertama seluruh variabel independen menjadi nol. selanjutnya variabel X_2 pertama kali masuk ke dalam model diikuti variabel X_4, X_1, X_5, X_6, X_7 , dan variabel yang terakhir masuk kedalam model variabel X_3 .
2. Berdasarkan analisis regresi LASSO data kriminalitas Indonesia tahun 2021 menggunakan algoritma LARS dengan kriteria AIC, diperoleh model terbaik sebagai berikut:

$$\hat{Y} = -1.628007 \times 10^{-17} + (1.731027 \times 10^{-1})X_2$$

Berdasarkan persamaan model 4.8, menjelaskan bahwasanya terdapat satu dari tujuh variabel independen yang masuk ke dalam model terbaik yaitu jumlah

penduduk di Indonesia (X_2). Model terbaik dipilih berdasarkan kombinasi variabel yang menghasilkan nilai lambda dan AIC optimal. Dengan mempertimbangkan trade-off antara ketepatan dan kompleksitas model, sehingga model mampu memberikan estimasi yang baik mengenai keanekaragaman data kriminalitas. Dengan demikian, model yang menghasilkan nilai lambda dan AIC optimal dapat dianggap sebagai model optimal untuk analisis data kriminalitas di Indonesia tahun 2021.

5.2. Saran

Setelah dilakukannya penelitian dan diperoleh kesimpulan terkait regresi LASSO menggunakan algoritma LARS pada data kriminalitas di Indonesia tahun 2021, penulis menyampaikan beberapa saran sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah data kriminalitas di Indonesia tahun 2021 dengan tujuh variabel independen dimana data tersebut sudah dua tahun lalu. Untuk meningkatkan relevansi penelitian, disarankan untuk menggunakan dataset terbaru dalam analisis regresi LASSO dengan algoritma LARS, mengatasi kemultikolinearitas pada data kriminalitas di Indonesia. Dengan data yang lebih baru, penelitian dapat memberikan gambaran yang lebih akurat tentang faktor-faktor yang memengaruhi kriminalitas.
2. Pemilihan model terbaik pada penelitian ini menggunakan kriteria AIC sedangkan masih banyak metode yang dapat digunakan untuk pemilihan model terbaik regresi LASSO seperti metode BIC, GCV, CV MSE, CP Mallows, dan masih banyak lainnya. Diharapkan juga mempertimbangkan metode analisis dan model prediktif terbaru untuk penelitian yang lebih inovatif.

DAFTAR PUSTAKA

- Akaike, H. (1987). Factor analysis and aic. *Psychometrika*, 52:317–332, <https://api.semanticscholar.org/CorpusID:47952861>.
- Altalbany, S. I. (2021). Evaluation of ridge, elastic net and lasso regression methods in precedence of multicollinearity problem: A simulation study. *Journal of Applied Economics and Business Studies*, <https://api.semanticscholar.org/CorpusID:236668371>.
- 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.
- Azizah, I. N., Arum, P. R., & Wasono, R. (2021). Model terbaik uji multikolinearitas untuk analisis faktor-faktor yang mempengaruhi produksi padi di kabupaten blora tahun 2020. In *Prosiding Seminar Nasional UNIMUS*, volume 4.
- Bjerkholt, O. (2018). Ragnar frisch (1895-1973). *Memorandum (institute of Pacific Relations, American Council)*, <https://api.semanticscholar.org/CorpusID:159243350>.
- Draper, N. R. & Smith, H. (1998). *Applied regression analysis*, volume 326. John Wiley & Sons.
- Efron, B., Hastie, T. J., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32:407–499, <https://api.semanticscholar.org/CorpusID:121570279>.

- Fathurahman, M. (2010). Pemilihan model regresi terbaik menggunakan akaike's information criterion. *J. EKSPONENSIAL*, 1(2):26–33.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- Islamiyati, A. (2019). Penggunaan regresi spline terpenalti dalam mengidentifikasi pola perubahan nilai kurs rupiah di indonesia.
- Kurniawan, P. S. et al. (2018). Maximum likelihood estimator untuk mengestimasi model regresi isotoniik dengan pendekatan polinomial bernstein pada kasus satu variabel independen.
- Kusuma, G. W. & Wulansari, I. Y. (2019). Analisis kemiskinan dan kerentanan kemiskinan dengan regresi ridge, lasso, dan elastic-net di provinsi jawa tengah tahun 2017. In *Seminar Nasional Official Statistics*, volume 2019, pages 503–513.
- Mahalani, A. J. & Rifai, N. A. K. (2022). Least absolute shrinkage and selection operator (lasso) untuk mengatasi multikolinearitas pada model regresi linear berganda. In *Bandung Conference Series: Statistics*, volume 2, pages 119–125.
- Ningsih, D. R., Intan, P. K., & Yuliati, D. (2023). Pemodelan tindak pidana kriminalitas di kota tangerang menggunakan metode regresi lasso. *ESTIMASI: Journal of Statistics and Its Application*, pages 64–77.
- Pardede, T. T., Sumargo, B., & Rahayu, W. (2022). Penerapan regresi least absolute shrinkage and selection operator (lasso) untuk mengidentifikasi variabel

- yang berpengaruh terhadap kejadian stunting di indonesia. *Jurnal Statistika dan Aplikasinya*, 6(1):37–48.
- Pekhimenko, G. (2006). Penalized logistic regression for classification. *Dept. Comput. Sci., Univ. Toronto, Toronto, ON M5S3L1*.
- Purba, S. A. (2021). Estimasi parameter data berdistribusi normal menggunakan maksimum likelihood berdasarkan newton raphson. *Jurnal Sains Dasar*, <https://api.semanticscholar.org/CorpusID:247426512>.
- Rahmawati, F. & Suratman, R. Y. (2022). Performa regresi ridge dan regresi lasso pada data dengan multikolinearitas. *Leibniz: Jurnal Matematika*, 2(2):1–10.
- Robbani, M., Agustiani, F., & Herrhyanto, N. (2019). Regresi least absolute shrinkage and selection operator (lasso) pada kasus inflasi di indonesia tahun 2014-2017. *Jurnal EurekaMatika*, 7(2):1–16.
- Sartika, I., Debatara, N. N., & Imro'ah, N. (2020). Analisis regresi dengan metode least absolute shrinkage and selection operator (lasso) dalam mengatasi multikolinearitas. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 9(1).
- Sarwoko, M. (2005). Dasar-dasar ekonometrika. *Yogyakarta: Andi Offset*.
- Sugiyono, P. (2017). Metode penelitian bisnis: pendekatan kuantitatif, kualitatif, kombinasi, dan r&d. *Penerbit CV. Alfabeta: Bandung*, 225:87.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.
- Tibshirani, R. & Wasserman, L. (2017). Sparsity, the lasso, and friends. *Lecture notes from "Statistical Machine Learning," Carnegie Mellon University, Spring*.