

SKRIPSI

**OPTIMALISASI REGRESI BINOMIAL NEGATIF DENGAN
REGULARISASI LASSO MENGGUNAKAN K-FOLD CROSS
VALIDATION**

**(STUDI KASUS : JUMLAH KASUS DBD DI JAWA TENGAH
TAHUN 2024)**



ROBIT MA'RUFİ SABILA

22106010081

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

2026

**OPTIMALISASI REGRESI BINOMIAL NEGATIF DENGAN
REGULARISASI LASSO MENGGUNAKAN K-FOLD CROSS
VALIDATION**

**(STUDI KASUS : JUMLAH KASUS DBD DI JAWA TENGAH
TAHUN 2024)**

Skripsi

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



diajukan oleh

ROBIT MA'RUFİ SABILA

22106010081

Kepada

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

2026



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 : Robit Ma'rufi Sabila

NIM : 22106010081

Judul Skripsi : Optimalisasi Regresi Binomial Negatif dengan Regularisasi Lasso
Menggunakan *K-Fold Cross Validation* (Studi Kasus: Jumlah Kasus DBD di Jawa Tengah tahun 2024)

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

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

Wassalamu'alaikum wr. wb.

Pembimbing II

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

NIP. 19741003 200003 2 002

Yogyakarta, 24 Juli 2026

Pembimbing I

Aulia Khifah Futhona, M.Sc.

NIP. 19920605 201903 2 021



PENGESAHAN TUGAS AKHIR

Nomor : B-1869/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : Optimalisasi Regresi Binomial Negatif dengan Regularisasi Lasso Menggunakan K-Fold Cross Validation (Studi Kasus: Jumlah Kasus DBD di Jawa Tengah tahun 2024)

yang dipersiapkan dan disusun oleh:

Nama : ROBIT MA'RUFU SABILA
Nomor Induk Mahasiswa : 22106010081
Telah diujikan pada : Senin, 03 Agustus 2026
Nilai ujian Tugas Akhir : A-

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

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Aulia Khifah Futhona, M.Sc.
SIGNED

Valid ID: 6a7ee4f27b1a3



Penguji I

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

Valid ID: 6a83984987329



Penguji II

Lilih Deva Martias, M.Sc.
SIGNED

Valid ID: 6a7eff6ed14e0



Yogyakarta, 03 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a83d0f80bf78

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Robit Ma'rufi Sabila

NIM : 22106010081

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, 27 Juli 2026



Robit Ma'rufi Sabila

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

TIADA HALAMAN YANG LEBIH MENENANGKAN SELAIN LEMBAR
PERSEMBAHAN INI, SEBAB TERDAPAT MEREKA YANG SELALU
MENGHADIRKAN DOA, CINTA, DAN KEKUATAN DALAM SETIAP LANGKAH
PERJUANGAN PENULIS.

SKRIPSI INI PENULIS PERSEMBAHKAN KEPADA KEDUA ORANG TUA
TERCINTA, KAKAK SERTA ADIK² PENULIS YANG SELALU MEMBERIKAN
DO'A, KASIH DAN CINTA NYA.

SEMUA KELUARGA, SAHABAT DAN ORANG-ORANG YANG PERNAH PENULIS
KENAL. ALMAMATER KAMPUS TERCINTA YANG BANYAK MEMBERIKAN
WARNA DALAM HIDUP SELAMA 4 TAHUN, UNIVERSITAS ISLAM NEGERI
SUNAN KALIJAGA YOGYAKARTA

"MATEMATIKA ANGKATAN 2022 TERCINTA"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

“Bagi setiap umat ada kiblat yang dia menghadap ke arahnya. Maka, berlomba-lombalah kamu dalam berbagai kebajikan. Di mana saja kamu berada, pasti Allah akan mengumpulkan kamu semuanya. Sesungguhnya Allah Mahakuasa atas segala sesuatu.”

(QS. Al Baqarah ayat 148)

“Dialah yang menjadikan bumi untuk kamu dalam keadaan mudah dimanfaatkan. Maka, jelajahilah segala penjurunya dan makanlah sebagian dari rezeki-Nya. Hanya kepada-Nya kamu (kembali setelah) dibangkitkan.”

(QS. Al Mulk ayat 15)

“Apabila kamu sudah memutuskan menekuni suatu bidang, jadilah orang yang konsisten. Itu adalah kunci keberhasilan yang sebenarnya.”

(B.J. Habibie.)

“Jangan pernah memilih dalam membantu seseorang. Bantulah mereka yang membutuhkan bantuanmu, karena kita tidak tahu, siapa yang akan membantu kita di masa yang akan datang.”

(Keluarga Penulis)

PRAKATA

Alhamdulillah rabbil'alamin, puji syukur kehadiran Allah Swt. yang telah melimpahkan rahmat, nikmat serta hidayah-Nya. Shalawat serta salam semoga selalu tercurahkan kepada junjungan kita, Nabi Muhammad Saw. Berkat rahmat Allah Yang Maha Kuasa, penulis dapat menyelesaikan skripsi ini yang berjudul "Optimalisasi Regresi Binomial Negatif dengan Regularisasi Lasso menggunakan K-Fold Cross Validation".

Penulis mengucapkan terima kasih kepada Ibu Aulia Khifah Futhona, M.Sc., selaku pembimbing I dan Ibu Sri Utami Zuliana, S.Si., M.Sc., Ph.D., selaku pembimbing II yang telah meluangkan waktu dan tenaga untuk memberikan saran dan bimbingan serta masukannya kepada penulis. Penulis menyadari bahwa proses penulisan skripsi ini terdapat banyak hambatan dan rintangan. Namun berkat adanya bimbingan, motivasi, dan dorongan dari berbagai pihak, *alhamdulillah* skripsi ini dapat terselesaikan. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah Swt. yang senantiasa memberikan kesehatan, kekuatan, dan kemudahan kepada penulis.
2. Bapak Prof. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
3. Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
4. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi

Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

5. Bapak Deddy Rahmadi, M.Sc., selaku dosen pembimbing akademik yang telah memberikan pengarahan kepada penulis dari awal perkuliahan hingga akhir masa studi.
6. Seluruh dosen dan staf Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang senantiasa memberikan ilmu dan layanan terbaik kepada penulis dari awal hingga akhir perkuliahan.
7. Kedua orang tua tercinta, yang telah memberikan kasih, doa dan dukungannya sehingga penulis sampai ke titik ini.
8. Kakak, adik-adik serta keluarga besar penulis yang selalu memberikan semangat, doa, dan dukungannya yang sangat berarti bagi penulis.
9. Sahabat "Pentagon Rajin Tahajud" yang telah memberikan dukungan dan kebersamaan kepada penulis sejak awal masa perkuliahan hingga tahap akhir penyusunan skripsi ini.
10. Teman-teman satu bimbingan kelompok belajar "Skripsi Poisson" yang telah bersama-sama menemani dan membantu selama bimbingan.
11. Teman-teman "KKN 117 Gunawijaya Magelang" yang telah memberikan pengalaman, kebersamaan, dan kenangan berharga selama pelaksanaan KKN.
12. Teman-teman seperjuangan di perpustakaan yang telah menemani, membantu, dan memberikan semangat kepada penulis selama proses perkuliahan dan penyusunan skripsi.

13. Sobot “PPT” yang telah menjadi tempat berbagi cerita, tawa, dukungan, dan kebersamaan bagi penulis selama masa perkuliahan hingga penyusunan skripsi ini.
 14. Sobot grup “Namanya Ini Aja” yang telah kebersamai penulis sejak masa SMP hingga sekarang serta memberikan dukungan, kebersamaan, dan banyak kenangan berharga.
 15. Keluarga Matematika angkatan 2022 yang telah kebersamai selama di bangku perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.
 16. Semua pihak yang telah memberikan dukungan, bantuan, dan doa, yang tidak dapat disebutkan satu per satu. Kebaikan, perhatian, dan bantuan kalian sangat berarti dalam perjalanan penulis menyelesaikan studi ini. Semoga segala kebaikan yang diberikan dibalas dengan keberkahan dan kebahagiaan.
- Penulis menyadari bahwa skripsi ini masih belum sempurna, sehingga penulis sangat mengharapkan kritik dan saran yang membangun. Akhir kata, penulis berharap semoga Allah SWT membalas semua kebaikan dari pihak-pihak yang telah membantu menyelesaikan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi semua orang, khususnya dalam pengembangan ilmu pengetahuan. *Aamiin.*

Yogyakarta, 23 Mei 2026

Robit Ma'rufi Sabila

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN TUGAS AKHIR	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMBANG	xv
INTISARI	.xviii
ABSTRACT	xix
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah	6
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
1.6. Tinjauan Pustaka	7
1.7. Sistematika Penulisan	16
II DASAR TEORI	17
2.1. Model linear	17

2.2. <i>Generalized Linear Models</i>	19
2.2.1. Komponen GLM	19
2.2.2. Asumsi GLM	23
2.3. Analisis Regresi Poisson	25
2.4. <i>Maximum Likelihood Estimation</i> Regresi Poisson	28
2.5. Overdispersi	30
2.6. Regresi Binomial Negatif	32
2.7. <i>Maximum Likelihood Estimation</i> Regresi Binomial Negatif	34
2.8. Algoritma <i>Newton Raphson</i>	41
2.9. Data Dimensi Tinggi	42
2.10. Multikolinearitas	43
2.11. Overfitting	45
2.12. <i>Least Absolute Shrinkage and selection Operator</i> (Lasso)	47
2.13. K-Fold Cross Validation	51
III METODE PENELITIAN	55
3.1. Metode Penelitian	55
3.2. Jenis dan Sumber Literatur	55
3.3. Studi Kasus	57
3.4. Variabel Data Penelitian	58
3.5. Software yang digunakan	58
3.6. Langkah-langkah Analisis	59
3.7. Source code	60
3.8. Diagram Analisis Data (<i>Flowchart</i>)	61
IV HASIL DAN PEMBAHASAN	62
4.1. Deskripsi Data	62
4.2. Statistik Deskriptif	64

4.3. Pemodelan Regresi Poisson	66
4.4. Uji Equidispersi	68
4.5. Pemodelan Regresi Binomial Negatif	70
4.6. Uji Multikolinearitas	72
4.7. Standarisasi Data	74
4.8. Overfitting	76
4.9. Regularisasi <i>Least Absolute Shrinkage and Selection Operator</i> (Lasso)	77
4.10. Penentuan Nilai Parameter Lambda Optimal Menggunakan K-Fold Cross Validation (KCV)	80
4.11. Seleksi Variabel dengan Lambda Optimal	83
4.12. Pemodelan regresi Binomial Negatif Akhir	85
V PENUTUP	87
5.1. Kesimpulan	87
5.2. SARAN	89
DAFTAR PUSTAKA	89
LAMPIRAN	93
A DATA PENELITIAN	93
B SOURCE CODE PROGRAM R	96
C CURRICULUM VITAE	103

DAFTAR TABEL

1.1	Tinjauan Pustaka	13
2.1	Fungsi <i>Link</i> Umum untuk Berbagai Variabel Respons	21
2.2	Ilustrasi K-Fold Cross-Validation dengan $K = 5$	52
3.1	Deskripsi Variabel Penelitian	58
4.1	Analisis Statistik Deskriptif	65
4.2	Hasil Estimasi Parameter Model Regresi Poisson	67
4.3	Hasil uji equidispersi	68
4.4	Hasil Estimasi Parameter Regresi Binomial Negatif	70
4.5	Hasil Uji Multikolinearitas (VIF)	73
4.6	Sebagian Nilai Parameter λ lintasan penalti Lasso	79
4.7	Sebagian Nilai Parameter λ dan Nilai <i>Cross Validation Error</i>	81
4.8	Hasil Parameter Optimal Model Binomial Negatif Lasso	83
4.9	Koefisien Regresi LASSO dengan λ Optimal	84
4.10	Koefisien Model Regresi Binomial Negatif Akhir	85
1.1	Data Penelitian	93

DAFTAR GAMBAR

2.1	Multikolinearitas	45
2.2	Ilustrasi <i>Underfitting</i> , <i>Balanced</i> , dan <i>overfitting</i>	46
3.1	Flowchart Penelitian	61
4.1	Plot Jumlah Kasus DBD di Jawa Tengah Tahun 2024	63
4.2	Boxplot Sebelum Standarisasi Data	75
4.3	Boxplot Setelah Standarisasi Data	76
4.4	Penentuan Parameter λ dengan K-Fold Cross Validation	78
4.5	Plot Nilai KCV dan Lambda	82

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR LAMBANG

y_i	: variabel dependen / respon
$x_1, x_2, \dots, x_p$	: variabel independen / prediktor
β_0	: konstanta
$\beta_1, \beta_2, \dots, \beta_p$	: koefisien regresi
ε	: <i>error</i>
Σ	: penjumlahan total
$\mathbf{y}$	: vektor variabel dependen
$\mathbf{X}$	: matriks variabel independen
$\boldsymbol{\beta}$	: vektor parameter regresi
$\boldsymbol{\varepsilon}$	: vektor komponen <i>error</i> acak
σ^2	: variansi <i>error</i>
η	: prediktor linear
$\boldsymbol{\beta}^T$	: transpose vektor parameter regresi
x_i^T	: transpose vektor prediktor pada pengamatan ke- i
$g(\mu)$	: fungsi penghubung (<i>link function</i>)
μ	: rata-rata (mean)
$g^{-1}(\eta)$	: invers fungsi penghubung (<i>inverse link function</i>)
π	: probabilitas kejadian
$\text{logit}(\pi)$	: fungsi logit
β	: parameter regresi
$f(y; \mu)$	: fungsi peluang distribusi Poisson

e	: bilangan eksponensial natural
Var	: variansi
E	: ekspektasi / rata-rata
$M_Y(t)$	: fungsi pembangkit momen variabel acak Y
$\ln$	: logaritma natural
$P(y_i)$	: probabilitas variabel dependen pada pengamatan ke- i
D^2	: deviansi
db	: derajat bebas (<i>degree of freedom</i>)
χ^2	: statistik Pearson Chi-Square
α	: parameter dispersi regresi Binomial Negatif
$\exp$	: fungsi eksponensial
$f(y; \mu, \alpha)$	: fungsi massa peluang distribusi Binomial Negatif
$\Gamma(\cdot)$	: fungsi gamma
$L(\cdot)$	: fungsi likelihood
$\prod$	: notasi perkalian
$\frac{\partial}{\partial \beta}$	: turunan parsial terhadap parameter β
$\ell(\cdot)$	: fungsi log-likelihood
$\psi(\cdot)$	: fungsi digamma
p	: banyaknya variabel independen
n	: banyaknya observasi
l_p	: fungsi likelihood penalti
VIF	: Variance Inflation Factor untuk variabel
R^2	: koefisien determinasi
$P(\beta)$	: fungsi penalti Lasso
λ	: parameter penalti Lasso

$\ \beta\ _1$	: norma- L_1 vektor parameter regresi
$\hat{\beta}_{\text{Lasso}}$	: estimator parameter regresi Lasso
$\arg \min$	: argumen yang meminimumkan suatu fungsi
t	: nilai parameter tuning dengan $t \geq 0$
$Q(\beta, \alpha)$	: fungsi objektif regresi Binomial Negatif Lasso
$\text{sign}(\beta_j)$	: fungsi tanda (<i>sign function</i>)
$(\hat{\beta}, \hat{\alpha})$	: estimator parameter hasil optimasi
Err	: <i>error</i> prediksi
$E[L(Y, \hat{f}(X))]$	: nilai harapan fungsi kerugian model
$\hat{f}(X)$	: fungsi prediksi hasil estimasi model
$L(\cdot)$	: fungsi kerugian (<i>loss function</i>)
K	: jumlah fold
CV	: Cross Validation
$\hat{\lambda}$	: parameter optimal hasil validasi silang

INTISARI

Optimalisasi Regresi Binomial Negatif dengan Regularisasi Lasso

menggunakan K-Fold Cross Validation

(Studi kasus : Jumlah Kasus DBD di Jawa Tengah Tahun 2024)

Robit Ma'rufi Sabila

22106010081

Regresi merupakan salah satu metode statistika untuk menganalisis hubungan antara variabel dependen dan independen. Pada *count data*, model regresi yang umum digunakan adalah regresi Poisson. Regresi Poisson memiliki asumsi equidispersi, yaitu nilai rata-rata dan variansi dari variabel dependen bernilai sama. Namun dalam praktiknya, regresi Poisson sering mengalami overdispersi yaitu nilai variansi lebih besar dari rata-ratanya, sehingga regresi Poisson kurang sesuai digunakan. Salah satu metode untuk mengatasi overdispersi adalah regresi Binomial Negatif, karena memiliki parameter dispersi yang dapat mengakomodasi overdispersi pada data. Selain overdispersi, multikolinearitas yang ditandai dengan tingginya korelasi antarvariabel independen dapat menyebabkan koefisien regresi menjadi kurang stabil. Penggunaan data berdimensi tinggi juga dapat meningkatkan risiko overfitting. Oleh karena itu, diperlukan metode seleksi variabel untuk menghasilkan model yang lebih sederhana dan efisien. Untuk mengatasi masalah tersebut, digunakan regularisasi Lasso dengan menerapkan penalti L_1 pada koefisien regresi sehingga menghasilkan estimasi parameter yang lebih stabil sekaligus melakukan seleksi variabel. Penelitian ini menggunakan jumlah kasus DBD di Jawa Tengah tahun 2024 sebagai variabel dependen dan 12 variabel independen yang berkaitan dengan faktor kesehatan, sanitasi, dan sosial ekonomi. Penelitian ini bertujuan untuk menerapkan regresi Binomial Negatif dengan regularisasi Lasso menggunakan K-Fold Cross Validation. Hasil penelitian menunjukkan bahwa nilai λ optimal sebesar 0,4247 menghasilkan satu variabel terpilih, yaitu Jumlah Sarana Kesehatan Puskesmas (JSKP), sehingga diperoleh model yang lebih parsimonious dan efisien. **Kata kunci** : Binomial Negatif, Lasso, K-Fold Cross Validation, Demam Berdarah (DBD)

ABSTRACT

Optimization of Negative Binomial Regression with Lasso Regularization Using *K-Fold Cross-Validation*

(Case Study: The Number of Dengue Hemorrhagic Fever (DHF) Cases in Central
Java in 2024)

Robit Ma'rufi Sabila

22106010081

Regression is one of the statistical methods used to analyze the relationship between dependent and independent variables. For *count data*, the Poisson regression model is commonly used. Poisson regression assumes equidispersion, where the mean and variance of the dependent variable are equal. However, in practice, Poisson regression often encounters overdispersion, where the variance is greater than the mean, making Poisson regression less appropriate. One method for addressing overdispersion is Negative Binomial regression, which has a dispersion parameter that can accommodate overdispersion in the data. In addition to overdispersion, multicollinearity, characterized by high correlations among independent variables, can cause regression coefficients to become unstable. The use of high-dimensional data can also increase the risk of overfitting. Therefore, a variable selection method is needed to produce a simpler and more efficient model. To address these issues, Lasso regularization is applied by imposing an L_1 penalty on the regression coefficients, thereby producing more stable parameter estimates while simultaneously performing variable selection. This study uses the number of Dengue Hemorrhagic Fever (DHF) cases in Central Java in 2024 as the dependent variable and 12 independent variables related to health, sanitation, and socioeconomic factors. This study aims to apply Negative Binomial regression with Lasso regularization using K-Fold Cross Validation. The results show that the optimal λ value of 0.4247 selects one variable, namely the Number of Community Health Centers (JSKP), resulting in a more parsimonious and efficient model.

Keywords: Negative Binomial, Lasso, K-Fold Cross Validation, Dengue Hemorrhagic Fever (DHF)

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Analisis regresi merupakan teknik statistik yang digunakan untuk menyelidiki dan memodelkan hubungan antarvariabel (Montgomery et al., 2021). Teknik ini digunakan dalam berbagai bidang, seperti fisika, kimia, ekonomi, manajemen, teknik, biologi, ilmu hayati dan ilmu sosial. Regresi juga sering dipakai dalam proses penambangan data dan menjadi dasar dalam ilmu serta analisis data. Analisis regresi juga merupakan teknik statistik yang paling banyak digunakan karena penerapannya yang luas pada berbagai permasalahan. Meskipun analisis regresi linear banyak digunakan, model ini memiliki keterbatasan karena mengasumsikan bahwa *error* berdistribusi normal dengan variansi konstan. Selain itu, regresi linear kurang sesuai digunakan ketika variabel dependen bersifat diskrit, tidak negatif, dan memiliki hubungan tertentu antara nilai rata-rata dan variansi, karena melanggar asumsi regresi linear. Oleh karena itu, dikembangkan *Generalized Linear Model* (GLM).

Generalized Linear Model (GLM) merupakan model regresi yang tidak mengikuti normal tetapi masih mengikuti distribusi keluarga eksponensial. Model GLM berfungsi apabila variabel dependennya berupa *count data* bilangan bulat non negatif atau data cacah. Salah satu bentuk khusus dari GLM adalah regresi Poisson. Regresi Poisson adalah model regresi yang digunakan pada data yang variabel dependennya berdistribusi tidak normal dan merupakan *count data*. Regresi

Poisson juga dapat digunakan untuk data *crossection*, *time series*, maupun data panel. Regresi Poisson mempunyai asumsi *equidispersi*, yaitu kondisi di mana pada variabel dependen, nilai rata-rata sama dengan nilai variansinya (Wibawa et al., 2023). Namun, dalam banyak kasus, asumsi dasar regresi Poisson sering tidak terpenuhi karena adanya overdispersi.

Overdispersi merupakan kondisi ketika variansi variabel dependen lebih besar daripada nilai rata-ratanya (Sundari & Sihombing, 2021). Overdispersi dapat dilihat berdasarkan nilai *pearson chi-squares* dan deviasi yang dibagi dengan derajat bebas yang bernilai lebih dari 1. Apabila nilai deviasi dan derajat bebas menghasilkan nilai melebihi angka 1, maka pada model regresi Poisson menunjukkan adanya kondisi overdispersi. Kondisi ini menyebabkan model regresi Poisson menjadi kurang baik. Hal ini terjadi karena adanya tingkat kesalahan yang tinggi, serta menyebabkan interpretasi menjadi tidak valid. Solusi yang dapat dilakukan untuk menangani overdispersi pada regresi Poisson adalah dengan mengganti asumsi distribusi Poisson dengan distribusi Binomial Negatif (Widodo & Ariani, 2018).

Regresi Binomial Negatif merupakan pengembangan dari regresi Poisson yang lebih fleksibel dalam memodelkan *count data*, khususnya apabila terjadi overdispersi (Hilbe, 2011). Regresi ini merupakan model regresi non linear yang terbentuk dari kombinasi distribusi Poisson dan Gama (*Poisson Gama Mixture*) serta termasuk dalam keluarga *Generalized Linear Model (GLM)*, sehingga mampu menangani variasi data yang berbeda-beda antar pengamatan. Berbeda dengan model Poisson yang mengharuskan variansi sama dengan rata-ratanya, model Binomial Negatif memiliki model variansi yang lebih umum, yaitu sebagai fungsi dari rata-rata, dan dilakukan dengan menambahkan parameter dispersi. Model ini

lebih fleksibel karena dapat menangkap variasi data yang lebih besar, sehingga memberikan hasil estimasi yang lebih akurat, dibandingkan dengan model regresi Poisson pada kondisi overdispersi (Hilbe, 2011).

Selain permasalahan asumsi equidispersi, permasalahan multikolinearitas dan overfitting juga merupakan isu fundamental pada metode regresi klasik, terutama apabila melibatkan banyak variabel independen. Pada model berbasis *Generalized Linear Model* (GLM), seperti regresi Poisson dan Binomial Negatif, parameter model diestimasi menggunakan *Maximum Likelihood Estimation* (MLE), yaitu metode yang menentukan parameter dengan memaksimalkan fungsi likelihood dari data yang diamati (Nugraha, 2014). Dalam praktiknya, fungsi likelihood sering diubah menjadi fungsi log-likelihood karena lebih mudah untuk dihitung dan dioptimalkan tanpa mengubah nilai maksimum yang dihasilkan. Proses optimasi pada MLE umumnya dilakukan menggunakan metode numerik, seperti algoritma Newton-Raphson atau *Iteratively Reweighted Least Squares* (IRLS), karena solusi analitik biasanya sulit diperoleh (Hilbe, 2011). Selain itu, estimator yang dihasilkan melalui MLE memiliki sifat konsisten, tidak bias, dan efisien secara asimtotik. Oleh karena itu, diperlukan metode yang tidak hanya mampu mengestimasi parameter secara efisien melalui MLE, tetapi juga dapat mengatasi permasalahan multikolinearitas dan overfitting melalui proses seleksi variabel.

Metode seleksi variabel seperti subset selection dan regresi ridge telah dikembangkan untuk mengatasi keterbatasan model regresi klasik. Kedua teknik ini memiliki standar untuk memperbaiki keterbatasan. *Subset selection* menyediakan model yang dapat diinterpretasikan, namun bisa sangat bervariasi karena merupakan proses diskrit di mana variabel bisa dihapus atau dipertahankan

dari model, namun perubahan kecil pada data dapat menghasilkan model yang sangat berbeda. Sementara *regresi ridge* merupakan proses yang berkelanjutan dengan mengecilkan koefisien agar model lebih stabil, namun *ridge* tidak menetapkan koefisien apapun ke 0 sehingga menghasilkan model yang sulit diinterpretasikan. Kedua teknik ini dapat memperbaiki keterbatasan, namun keduanya masih memiliki kelemahan.

Selain permasalahan asumsi distribusi, analisis regresi dengan banyak variabel independen berpotensi mengalami ketidakstabilan model dan kesulitan dalam interpretasi parameter (Hastie et al., 2009). Oleh karena itu, diperlukan suatu metode optimisasi yang mampu melakukan seleksi variabel secara simultan. Salah satu metode regularisasi yang banyak digunakan adalah *Least Absolute Shrinkage and Selection Operator* (Lasso). Metode Lasso bekerja dengan menambahkan penalti norma L_1 pada fungsi likelihood sehingga mampu mengecilkan koefisien regresi dan menetapkan yang lain ke 0.

Pada fungsi objektif, yaitu penjumlahan nilai absolut koefisien regresi, penalti tersebut memberikan hukuman terhadap koefisien yang bernilai besar sehingga nilai koefisien dapat mengecil bahkan menjadi nol. Variabel dengan koefisien sama dengan nol dianggap tidak memberikan kontribusi yang signifikan terhadap model sehingga dikeluarkan dari model. Dengan demikian, metode Lasso mencoba mempertahankan keunggulan dari metode sebelumnya yaitu *subset selection* dan *ridge regression* (Tibshirani, 1996).

Untuk mendapatkan model yang terbaik, digunakan K-Fold Cross Validation sebagai validasi model. Cross Validation adalah metode untuk melakukan validasi model terbaik. Model ini dibentuk dengan melakukan penyusunan ulang (resampling) pada data untuk membaginya menjadi 2 bagian, yaitu data training

dan data testing yang berfungsi untuk menguji keefektifan dari model. Metode yang sering digunakan dari Cross Validation adalah metode K-Fold Cross Validation, karena metode ini akan menghasilkan model yang tidak bias(Widyaningsih et al., 2021).

Penelitian ini menggunakan data jumlah kasus Demam Berdarah Dengue (DBD) di Jawa Tengah tahun 2024 yang merupakan infeksi yang disebabkan oleh virus dengue yang dapat menyebabkan demam akut. Penyakit DBD adalah penyakit menular yang disebabkan oleh virus dengue dan ditularkan melalui nyamuk *Aedes sp.* Infeksi dari virus dengue yaitu hasil interaksi multifaktorial antara faktor hospes, agen penyakit, dan lingkungan faktor determinan seperti, perubahan demografi, peningkatan urbanisasi, kepadatan penduduk, mobilitas penduduk, dan lainnya(Ri, 2010)

Berdasarkan uraian di atas, penelitian ini bertujuan untuk menerapkan metode regularisasi Lasso pada regresi Binomial Negatif yang diaplikasikan pada data DBD dalam menangani masalah multikolinearitas dan overfitting menggunakan teknik K-Fold Cross Validation untuk mendapatkan model terbaik yang berfungsi sebagai validasi model.

1.2. Rumusan Masalah

1. Bagaimana langkah-langkah dalam melakukan analisis regresi Binomial Negatif dengan regularisasi Lasso?
2. Bagaimana pengoptimalan model regresi Binomial Negatif dengan regularisasi Lasso menggunakan K-Fold Cross Validation ?
3. Bagaimana model terbaik regresi Binomial Negatif dengan regularisasasi Lasso pada data jumlah kasus DBD di Jawa Tengah tahun 2024?

1.3. Batasan Masalah

1. Penelitian dibatasi dengan penggunaan model regresi Binomial Negatif dengan regularisasi Lasso
2. Penelitian hanya berfokus pada teknik *K-Fold Cross Validation*
3. Penelitian ini hanya menggunakan data Demam Berdarah Dengue (DBD) di Jawa Tengah 2024
4. Analisis dilakukan menggunakan software R versi 4.5.0

1.4. Tujuan Penelitian

1. Mengetahui langkah-langkah dalam melakukan analisis regresi Binomial Negatif dengan regularisasi Lasso
2. Mengoptimalkan regresi Binomial Negatif dengan regularisasi Lasso dalam memodelkan kasus DBD menggunakan K-Fold Cross Validation
3. Mengetahui model terbaik regresi Binomial Negatif dengan regularisasi Lasso pada data yang di teliti

1.5. Manfaat Penelitian

1. Penelitian ini diharapkan dapat memberikan manfaat baik di bidang akademik maupun praktis
2. Meningkatkan pemahaman tentang penerapan regresi Binomial Negatif dengan regularisasi Lasso menggunakan K-Fold Cross Validation dalam menangani masalah overdispersi pada data DBD di Jawa Tengah 2024

3. Penelitian ini diharapkan dapat membantu mengoptimalkan model prediksi jumlah kasus Demam Berdarah Dengue (DBD) di Jawa Tengah, dengan menggunakan teknik regresi Binomial Negatif dengan regularisasi Lasso untuk memilih variabel yang relevan serta K-Fold Cross Validation untuk memperoleh model dengan performa prediksi yang lebih baik.
4. Hasil penelitian ini dapat berguna bagi penelitian penelitian selanjutnya yang melibatkan analisis regresi Binomial Negatif dengan regularisasi Lasso menggunakan K-Fold Cross Validation.

1.6. Tinjauan Pustaka

Sebagai dasar teori dalam penulisan skripsi ini, penulis merujuk pada berbagai sumber seperti buku, majalah, dan jurnal. Berikut ini adalah tinjauan pustaka dalam penelitian ini.

1. Nugraha (2014) membahas regresi Poisson sebagai salah satu metode dalam kerangka *Generalized Linear Model* (GLM) yang digunakan untuk memodelkan variabel dependen berupa *count data* serta merupakan pengembangan dari regresi Binomial Negatif. Penulis menguraikan karakteristik distribusi Poisson serta model regresi Poisson untuk menjelaskan hubungan antara variabel dependen dan variabel independen. Selain itu, dijelaskan pula proses estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE), yaitu dengan memaksimumkan fungsi likelihood. Dalam buku tersebut juga dijelaskan bahwa salah satu keterbatasan regresi Poisson adalah asumsi equidispersi, sehingga apabila terjadi overdispersi, diperlukan model alternatif seperti regresi Binomial Negatif yang lebih fleksibel dalam menangkap variasi data. Buku ini juga menyajikan penerapan model regresi

menggunakan perangkat lunak R sebagai ilustrasi dalam analisis *count data*.

2. Wibawa et al. (2023) bertujuan untuk mengetahui variabel yang berpengaruh terhadap faktor-faktor hipertensi di kota kendari tahun 2019 dengan menggunakan pemodelan regresi Poisson. Data terdiri dari satu variabel dependen dan lima variabel independen. Metode yang dilakukan dalam penelitian ini adalah regresi Poisson. Metode ini dilakukan karena data merupakan *count data* dan model regresi Poisson memiliki asumsi equidispersi (kondisi di mana nilai rata-rata sama dengan nilai variansi dari variabel dependen bernilai sama). Hasil penelitian menunjukkan bahwa faktor pelayanan kesehatan paling signifikan yang mempengaruhi jumlah hipertensi di Kota Kendari tahun 2019.
3. Hilbe (2011) membahas secara mendalam mengenai regresi Binomial Negatif sebagai pengembangan dari regresi Poisson dalam kerangka *Generalized Linear Model* (GLM). Penulis menjelaskan bahwa regresi Binomial Negatif digunakan untuk mengatasi permasalahan overdispersi pada *count data*, yaitu kondisi ketika variansi lebih besar daripada nilai rata-rata. Dalam buku ini juga diuraikan bahwa model Binomial Negatif memberikan fleksibilitas yang lebih baik dibandingkan regresi Poisson karena mampu menangkap variasi data yang lebih besar. Selain itu, dijelaskan proses estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) serta interpretasi koefisien dalam model regresi Binomial Negatif. Buku ini menjadi rujukan penting dalam penerapan model Binomial Negatif pada *count data*.
4. Widodo & Ariani (2018) bertujuan untuk memilih model terbaik yang digunakan untuk memodelkan kasus terjadinya pasien yang menderita DBD

di Jawa Tengah tahun 2016. Data pada penelitian ini terdiri dari satu variabel dependen dan 5 variabel independen. Metode yang digunakan yaitu Analisis deskriptif untuk mengetahui gambaran umum data, regresi Poisson digunakan pada *count data* sedangkan Binomial Negatif sebagai alternatif dari data overdispersi. Hasil penelitian menunjukkan model terbaik yaitu Regresi Binomial Negatif dibandingkan dengan regresi Poisson.

5. Sundari & Sihombing (2021) bertujuan untuk memperoleh model terbaik dari tiga model regresi yang diujikan yaitu model Regresi Negative Binomial, Zero-Truncated Poisson dan Zero Negative Binomial, serta memodelkan hubungan pengaruh faktor iklim terhadap jumlah penderita DBD di kota Bogor. Data yang digunakan adalah data jumlah penderita demam berdarah per tahun di Kotamadya Bogor yang bersumber dari Jawa Barat Dalam Angka 2011-2020 (Badan Pusat Statistik Jawa Barat) dengan satu variabel dependen dan lima variabel independen. Metode regresi Negative Binomial, Zero-Truncated Poisson dan Zero Negative Binomial dipilih dalam penelitian ini karena dapat mengatasi masalah overdispersi (nilai ekspektasi tidak sama dengan nilai variannya) sehingga dilakukan penanganan dengan ketiga metode tersebut. Model terpilih dengan kriteria terbaik adalah Negative Binomial. Pada taraf nyata 5% terdapat satu variabel independen yang memiliki pengaruh nyata terhadap jumlah DBD di Kota Bogor dan pada taraf nyata 15% terdapat tiga variabel independen yang memiliki pengaruh nyata terhadap jumlah DBD di Kota Bogor.
6. Suryani et al. (2021) bertujuan untuk memodelkan faktor-faktor yang mempengaruhi jumlah kasus DBD di Jawa Tengah menggunakan *Geographically Weighted Negative Binomial Regression* (GWNBR) dengan

menggunakan pembobot fixed exponential kernel. Data yang digunakan dalam penelitian ini merupakan jumlah kasus DBD di Jawa Tengah pada tahun 2017 (Y). Untuk variabel independen terdiri dari 5 variabel yaitu, Persentase Rumah Sehat (X1), Persentase Penduduk Miskin (X2), Persentase Kualitas Air Bersih (X3), Persentase Penduduk dengan Air Minum Layak (X4), dan Rasio Tenaga Medis (X5). Penelitian ini dilakukan menggunakan regresi Poisson karena data merupakan *count data* dan jika dalam regresi Poisson terdapat overdispersi, hal ini dapat diatasi dengan regresi binomial negatif. Sementara untuk melihat efek spasial, menggunakan metode *Geographically Weighted Negative Binomial Regression* (GWNBR). Hasil penelitian dengan regresi Poisson memperoleh tiga variabel yang memiliki pengaruh signifikan terhadap kasus demam berdarah. Untuk regresi binomial negatif, dua variabel memiliki pengaruh signifikan terhadap kasus DHF. Sementara metode GWNBR memperoleh dua kelompok kabupaten/kota berdasarkan variabel yang signifikan. Variabel yang mempengaruhi jumlah kasus DHF di semua kabupaten/kota di Jawa Tengah adalah persentase rumah sehat, persentase kualitas air bersih, dan rasio tenaga medis.

7. Al-Farisi (2024) bertujuan untuk menganalisis dan membandingkan model Regresi Poisson, Regresi Binomial Negatif, dan Regresi Hurdle Binomial Negatif pada data yang mengalami overdispersi. Data yang digunakan adalah data jumlah kasus kematian penderita Demam Berdarah Dengue (DBD) di Provinsi Jawa Timur tahun 2022 beserta beberapa faktor yang diduga mempengaruhinya. Penelitian ini menggunakan metode *Maximum Likelihood Estimation* (MLE) untuk mengestimasi parameter model, sedangkan pemilihan model terbaik dilakukan menggunakan kriteria *Akaike*

Information Criterion (AIC). Hasil penelitian menunjukkan bahwa model Regresi Binomial Negatif merupakan model terbaik dengan nilai AIC sebesar 191,94. Model yang diperoleh melibatkan tiga variabel prediktor yang berpengaruh terhadap jumlah kasus kematian penderita DBD, yaitu X_1 , X_5 , dan X_6 .

8. Tibshirani (1996) membahas metode baru untuk estimasi parameter pada model linear menggunakan *Least Absolute Shrinkage and Selection Operator* (Lasso) yang berfungsi untuk meminimumkan jumlah kuadrat residual dengan kendala bahwa jumlah nilai absolut dari koefisien regresi tidak melebihi suatu konstanta. Karena kendala tersebut, metode ini sering menghasilkan beberapa koefisien yang bernilai tepat nol, sehingga menghasilkan model yang mudah diinterpretasikan seperti seleksi subset, dan menunjukkan regresi ridge yang lebih stabil yang dapat memberikan solusi yang lebih baik dalam situasi banyak variabel independen.
9. Iswari (2026) bertujuan untuk memperoleh model Regresi Binomial Negatif dengan regularisasi Lasso yang optimal pada fitur tekstual promosi di platform X menggunakan kriteria *Akaike Information Criterion* (AIC). Data yang digunakan merupakan data sekunder berupa 1.063 unggahan promosi akun @ShopeeID di platform X periode 1 Januari hingga 27 November 2025, dengan jumlah *likes* sebagai variabel dependen dan 75 fitur tekstual hasil pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai variabel independen. Metode Regresi Binomial Negatif dipilih karena variabel respons berupa *count data* yang terindikasi mengalami overdispersi, sedangkan regularisasi Lasso diterapkan untuk melakukan penyusutan koefisien sekaligus seleksi fitur tekstual yang relevan. Pemilihan

nilai parameter penalti λ dilakukan menggunakan kriteria AIC sehingga diperoleh model yang lebih sederhana (*parsimonious*) dengan keseimbangan yang baik antara kecocokan model dan kompleksitasnya.

10. Widyarningsih et al. (2021) bertujuan untuk mengetahui faktor-faktor yang memengaruhi banyaknya publikasi ilmiah yang dihasilkan oleh mahasiswa PhD Biokimia tahun 1997. Data yang dipakai peneliti ialah jumlah penulisan publikasi yang dihasilkan oleh seorang mahasiswa PhD Biokimia yang terdiri dari 915 mahasiswa sebagai sampel dengan 6 variabel yaitu, banyaknya artikel yang dihasilkan oleh mahasiswa Ph.D. dalam 3 tahun terakhir, jenis kelamin, status pernikahan, banyaknya anak dibawah 5 tahun, prestise dari program Ph.D., dan banyaknya artikel oleh mentor dalam 3 tahun terakhir. Karena variabel dependen merupakan *count data*, metode analisis yang digunakan adalah Regresi Poisson. Namun karena data mengalami overdispersi, akan digunakan Regresi Binomial Negatif. Perbandingan beberapa model Regresi Poisson dan Binomial Negatif dilakukan untuk menentukan model terbaik dengan K-Fold Cross Validation sebagai validasi model. Hasil penelitian menunjukkan bahwa model terbaik yang didapatkan adalah model Regresi Binomial Negatif dengan variabel independen jenis kelamin, status pernikahan, banyaknya anak dibawah 5 tahun, prestise, dan banyaknya artikel oleh mentor dalam 3 tahun terakhir.

Tabel 1.1 Tinjauan Pustaka

No	Peneliti	Persamaan	Perbedaan	
			Penelitian Terdahulu	Penelitian Ini
1	Nugraha (2014)	Menggunakan pendekatan Maximum Likelihood Estimation (MLE) dalam estimasi parameter model	Pemilihan model terbaik menggunakan uji kecocokan model (<i>goodness of fit</i>), uji rasio likelihood, dan koefisien determinasi (R^2)	Pemilihan model terbaik menggunakan K-Fold Cross Validation
2	(Wibawa et al., 2023)	Menggunakan uji multikolinearitas dan menggunakan asumsi equidispersi	Menggunakan data hipertensi, Fokus pada pemodelan regresi Poisson dan uji parameter (uji serentak dan parsial)	Pemodelan menggunakan regresi Binomial Negatif dengan regularisasi Lasso dan pemilihan model terbaik menggunakan K-Fold Cross Validation.
3	Hilbe (2011)	Menggunakan model regresi Binomial Negatif, menangani overdispersi	Model digunakan untuk mengatasi overdispersi dengan pendekatan <i>Maximum Likelihood Estimation</i> (MLE) dalam kerangka GLM	Menggunakan regresi Binomial Negatif dengan regularisasi Lasso dan pemilihan model terbaik menggunakan K-Fold Cross Validation
4	Widodo & Ariani (2018)	Menggunakan regresi Binomial Negatif, menangani overdispersi	Menggunakan regresi Poisson dan Regresi Binomial Negatif sebagai pembanding dan sebagai alternatif untuk data yang mengalami overdispersi	Menggunakan regresi Binomial Negatif dengan regularisasi Lasso menggunakan <i>K-Fold Cross Validation</i>

Dilanjutkan ke halaman berikutnya

No	Peneliti	Persamaan	Perbedaan	
			Penelitian Terdahulu	Penelitian Ini
4	Widodo & Ariani (2018)	Menggunakan regresi Binomial Negatif, menangani overdispersi	Menggunakan regresi Poisson dan Regresi Binomial Negatif sebagai pembanding dan sebagai alternatif untuk data yang mengalami overdispersi	Menggunakan regresi Binomial Negatif dengan regularisasi Lasso menggunakan <i>K-Fold Cross Validation</i>
5	Sundari & Sihombing (2021)	Menggunakan data DBD, Menangani masalah overdispersi	Pemilihan model terbaik dilakukan untuk mengatasi overdispersi dengan membandingkan regresi <i>Negative Binomial</i> , <i>Zero-Truncated Poisson</i> , dan <i>Zero-Inflated Negative Binomial</i>	Menggunakan regresi Binomial Negatif dengan regularisasi Lasso serta validasi menggunakan K-Fold Cross Validation.
6	Suryani et al. (2021)	Menggunakan data DBD, penaksiran parameter dengan <i>Maximum Likelihood Estimation</i> (MLE)	Menggunakan model <i>Geographically Weighted Negative Binomial Regression</i> (GWNBR), pemilihan model berdasarkan AIC serta uji spasial	Model Binomial Negatif, Fokus pada pengendalian overfitting dan seleksi variabel menggunakan penalti Lasso serta evaluasi prediksi dengan K-Fold Cross Validation.

Dilanjutkan ke halaman berikutnya

No	Peneliti	Persamaan	Perbedaan	
			Penelitian Terdahulu	Penelitian Ini
7	Al-Farisi (2024)	Menggunakan model Regresi Binomial Negatif untuk menangani data overdispersi.	Menganalisis dan membandingkan Regresi Poisson, Regresi Binomial Negatif, dan Regresi Hurdle Binomial Negatif. Pemilihan model terbaik dilakukan menggunakan kriteria AIC.	Mengoptimalkan Regresi Binomial Negatif dengan regularisasi Lasso serta menentukan parameter penalti menggunakan K-Fold Cross Validation sehingga diperoleh model yang lebih sederhana melalui seleksi variabel.
8	Tibshirani (1996)	Menggunakan regularisasi Lasso	Lasso diperkenalkan dalam konteks regresi linear dengan asumsi distribusi normal	Menggunakan regularisasi Lasso yang diterapkan pada regresi Binomial Negatif untuk <i>count data</i>
9	Iswari (2026)	Menggunakan Regresi Binomial Negatif dengan regularisasi Lasso.	Pemilihan parameter penalti dilakukan menggunakan Akaike Information Criterion (AIC) pada data fitur tekstual promosi di platform X.	Pemilihan parameter penalti dilakukan menggunakan K-Fold Cross Validation pada data jumlah kasus DBD di Jawa Tengah.
10	Widyaningsih et al. (2021)	Menggunakan <i>count data</i> , serta validasi model dengan K-Fold Cross Validation	Menggunakan regresi Poisson dan Binomial Negatif sebagai perbandingan model terbaik	Menggunakan model binomial negatif dengan regularisasi Lasso untuk seleksi variabel, serta K-Fold Cross Validation untuk validasi model

Berdasarkan penelitian terdahulu, regresi Binomial Negatif banyak

digunakan untuk memodelkan *count data* yang mengalami overdispersi, sedangkan regularisasi Lasso banyak diterapkan untuk seleksi variabel independen melalui pemberian penalti terhadap koefisien regresi. Selain itu, K-Fold Cross Validation digunakan untuk menentukan parameter penalti optimal pada model regularisasi. Penerapan ketiga metode tersebut umumnya masih dilakukan secara terpisah. Penelitian yang mengintegrasikan regresi Binomial Negatif dengan regularisasi Lasso dan K-Fold Cross Validation pada data kasus DBD masih terbatas. Oleh karena itu, penelitian ini menggabungkan ketiga pendekatan tersebut untuk memperoleh model yang optimal.

1.7. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan tugas akhir ini terbagi menjadi lima bab, yaitu sebagai berikut:

- BAB 1** : Bab ini membahas tentang latar belakang masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan dan tinjauan pustaka.
- BAB 2** : Bab ini membahas landasan teori yang menjadi dasar acuan, meliputi konsep-konsep dan rancangan penelitian.
- BAB 3** : Bab ini membahas tentang metode penelitian, jenis penelitian dan sumber literatur, variabel data penelitian, langkah-langkah analisis dan diagram analisis data.
- BAB 4** : Bab ini dilakukan proses analisis data yang diikuti dengan interpretasi hasil secara mendalam dan terperinci.
- BAB 5** : Bab ini membahas tentang kesimpulan penelitian dan saran dari penulis.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dijelaskan, diperoleh beberapa kesimpulan sebagai berikut:

1. Langkah-langkah penelitian diawali dengan pemodelan regresi Poisson sebagai model awal, kemudian dilakukan uji asumsi equidispersi menggunakan nilai dispersi untuk menentukan ada atau tidaknya overdispersi. Apabila terjadi overdispersi, maka digunakan regresi Binomial Negatif sebagai model yang lebih sesuai. Selanjutnya, dilakukan uji multikolinearitas menggunakan *Variance Inflation Factor* (VIF). Apabila terdapat variabel yang memiliki nilai VIF lebih dari 10, maka variabel tersebut mengindikasikan adanya multikolinearitas. Tahap berikutnya adalah melakukan standarisasi variabel independen dan menerapkan regularisasi Lasso untuk melakukan seleksi variabel dengan menentukan parameter penalti λ menggunakan metode K-Fold Cross Validation.
2. Tahapan pengoptimalan model regresi Binomial Negatif dengan regularisasi Lasso menggunakan metode K-Fold Cross Validation dilakukan melalui proses standarisasi data untuk menyamakan skala antar variabel, kemudian melakukan regularisasi Lasso dengan menambahkan penalti L1, dengan penentuan nilai parameter penalti menggunakan metode K-Fold Cross Validation, berdasarkan jumlah observasi, dipilih 5 Fold Cross Validation,

dimana data dibagi menjadi 5 kelompok (fold), setiap fold berisi 7 observasi, Pada setiap iterasi, 4 fold digunakan sebagai data pelatihan dan 1 fold sebagai data validasi. Proses diulang hingga setiap fold menjadi data validasi satu kali. Kemudian dilakukan evaluasi model menggunakan nilai lambda berdasarkan nilai *Cross Validation Error* minimum dan digunakan untuk membentuk model akhir serta menentukan variabel independen yang terpilih.

3. Model terbaik regresi Binomial Negatif dengan regularisasi Lasso pada data jumlah kasus DBD di Jawa Tengah tahun 2024 diperoleh berdasarkan nilai parameter penalti λ optimal yang dipilih menggunakan metode K-Fold Cross Validation. Berdasarkan hasil analisis, diperoleh nilai λ optimal sebesar 0,4247183649 dengan nilai Cross Validation Error minimum sebesar 1,487026. Hasil seleksi variabel menggunakan regularisasi Lasso menunjukkan bahwa dari 12 variabel independen yang digunakan, hanya variabel Jumlah Sarana Kesehatan Puskesmas (JSKP) yang memiliki koefisien tidak nol. Model regresi Binomial Negatif terbaik yang diperoleh adalah:

$$\hat{\mu} = \exp(4,62448 + 0,05719 JSKP)$$

Hasil tersebut menunjukkan bahwa variabel JSKP merupakan variabel yang dipertahankan dalam model setelah proses regularisasi Lasso. Regularisasi Lasso mampu menyederhanakan model dengan mereduksi jumlah variabel prediktor dari 12 variabel menjadi satu variabel, sehingga diperoleh model yang lebih parsimonious.

5.2. SARAN

Setelah dilakukan penelitian mengenai penerapan regresi Binomial Negatif dengan regularisasi Lasso menggunakan metode K-Fold Cross Validation, penulis memberikan beberapa saran sebagai berikut:

1. Penelitian ini menggunakan data jumlah kasus DBD di Jawa Tengah tahun 2024 dengan sejumlah variabel independen yang terbatas. Oleh karena itu, disarankan pada penelitian selanjutnya untuk menggunakan data yang lebih terbaru serta menambahkan variabel lain yang relevan
2. Penelitian selanjutnya disarankan untuk membandingkan metode regularisasi Lasso dengan metode regularisasi lain seperti *Ridge Regression*, *Elastic Net*, atau metode seleksi variabel lainnya untuk memperoleh model yang lebih optimal dan stabil.
3. Penelitian selanjutnya juga dapat mempertimbangkan penggunaan metode validasi lain selain K-Fold Cross Validation, seperti *Leave-One-Out Cross Validation* (LOOCV) atau *Repeated Cross Validation*, guna membandingkan kestabilan parameter penalti optimal yang diperoleh.

DAFTAR PUSTAKA

- Agresti, A. (2019). *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 3 edition, ISBN: 9781119405269.
- Al-Farisi, S. (2024). *Analisis Regresi Poisson, Regresi Binomial Negatif, dan Regresi Hurdle Binomial Negatif pada Data Overdispersi (Studi Kasus: Jumlah Kasus Kematian Penderita Demam Berdarah Dengue di Provinsi Jawa Timur Tahun 2022)*. PhD thesis, UIN Sunan Kalijaga Yogyakarta.
- Algamal, Z. Y. & Lee, M. H. (2015). Adjusted adaptive lasso in high-dimensional poisson regression model. *Modern Applied Science*, 9(4):170.
- Fitriyanti, W. & Kurniawan, U. (2019). Regresi negatif binomial bivariat untuk mengatasi overdispersi regresi poisson bivariat. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 7(1).
- Habsy, B. A., Mufidha, N., Shelomita, C., Rahayu, I., & Muckorobin, M. I. (2023). Filsafat dasar dalam konseling psikoanalisis: Studi literatur. *Indonesian Journal of Educational Counseling*, 7(2):189–199.
- Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Overview of r and rstudio. In *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*, pages 31–47. Springer.
- Hastie, T., Tibshirani, R., Friedman, J., et al. (2009). The elements of statistical learning.

- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
- Iswari, D. (2026). *Optimisasi Regresi Binomial Negatif Lasso Berbasis AIC pada Fitur Tekstual Promosi di Platform X (Studi Kasus: @ShopeeID)*. PhD thesis, UIN Sunan Kalijaga Yogyakarta.
- Martono, N. (2010). *Metode penelitian kuantitatif: Analisis Isi dan Analisis Data Sekunder (sampel halaman gratis)*. RajaGrafindo Persada.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Nugraha, J. (2014). *Pengantar analisis data kategorik: Metode dan aplikasi menggunakan program R*. Deepublish.
- Ramdhan, M. et al. (2021). *Metode penelitian*. Cipta Media Nusantara.
- Ri, K. (2010). Demam berdarah dengue. *Bul Jendela Epidemiol*, 2:48.
- Sadik, K. et al. (2021). *Kajian Regresi Logistik Terpenalti pada Data Dimensi Tinggi*. PhD thesis, IPB University.
- Salinas Ruíz, J., Montesinos López, O. A., Hernández Ramírez, G., & Crossa Hiriart, J. (2023). *Generalized linear mixed models with applications in agriculture and biology*. Springer Nature.
- Sari, R. A. (2017). Analisis faktor yang berpengaruh terhadap penyakit demam berdarah dengue (dbd) di kabupaten bantul menggunakan regresi poisson.
- Sundari, M. & Sihombing, P. R. (2021). Penanganan overdispersi pada regresi poisson:(studi kasus: Pengaruh faktor iklim terhadap jumlah penderita penyakit

- demam berdarah di kota bogor). *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 2(1):1–9.
- Suryani, I., Yasin, H., & Kartikasari, P. (2021). Pemodelan jumlah kasus demam berdarah dengue (dbd) di jawa tengah dengan geographically weighted negative binomial regression (gwnbr). *Jurnal Gaussian*.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.
- Wibawa, G. N. A., Abapihi, B., et al. (2023). Pemodelan regresi poisson terhadap faktor-faktor yang mempengaruhi terjadinya hipertensi di kota kendari: Faktor-faktor yang mempengaruhi terjadinya hipertensi di kota kendari. *Jurnal Matematika Komputasi dan Statistika*, 3(1):255–262.
- Widodo, E. & Ariani, P. M. (2018). Analisis faktor penyebab penyakit dbd di jawa tengah menggunakan regresi binomial negatif. *Jurnal Kesehatan Vokasional*, 3(1).
- Widyaningsih, Y., Arum, G. P., & Prawira, K. (2021). Aplikasi k-fold cross validation dalam penentuan model regresi binomial negatif terbaik. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 15(2):315–322.