

**ANALISIS REGRESI *ELASTIC-NET* UNTUK MENGATASI MASALAH
MULTIKOLINEARITAS PADA DATA DEMAM BERDARAH *DENGUE*
(DBD) DI INDONESIA**

(Skripsi)

OLEH

ADINDA SARIANTI



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

ABSTRACT

ELASTIC-NET REGRESSION ANALYSIS TO OVERCOME MULTICOLLINEARITY PROBLEMS IN DENGUE HEMORRHAGIC FEVER (DHF) DATA IN INDONESIA

By

ADINDA SARIANTI

Multicollinearity is a common problem that arises when there is a strong relationship between independent variables in a regression model. The Elastic-Net regression method can be used to solve this problem, as it is able to shrink the regression coefficient to exactly zero and also perform simultaneous variable selection. This study was conducted with the aim of determining the performance of the Elastic-Net method in overcoming the problem of multicollinearity and comparing its estimated value with the Ordinary Least Squared (OLS) on dengue case data in Indonesia as well as to find out what factors affect dengue cases in Indonesia. From the results of the study, it can be concluded that the Elastic-Net method successfully overcomes the multicollinearity problem and produces a better regression coefficient compared to OLS based on AIC measurements. The analysis showed that the factors that influence dengue cases in Indonesia are number of environmental health workers (X_2), the percentage of households with proper sanitation (X_3), the percentage of households with decent housing (X_4), the percentage of households with decent drinking water (X_5), the amount of rainfall (X_6), number of rainy days (X_7), and population density (X_8).

Keywords: Regression Analysis, Multicollinearity, Elastic-Net, AIC, Dengue Hemorrhagic Fever (DHF).

ABSTRAK

ANALISIS REGRESI *ELASTIC-NET* UNTUK MENGATASI MASALAH MULTIKOLINEARITAS PADA DATA DEMAM BERDARAH *DENGUE* (DBD) DI INDONESIA

Oleh

ADINDA SARIANTI

Multikolinearitas adalah masalah umum yang muncul ketika ada hubungan yang kuat antar variabel bebas dalam model regresi. Metode Regresi *Elastic-Net* dapat digunakan untuk mengatasi masalah ini, karena metode ini mampu menyusutkan koefisien regresi tepat nol dan juga melakukan seleksi variabel secara simultan. Penelitian ini dilakukan dengan tujuan untuk mengetahui kinerja metode *Elastic-Net* dalam mengatasi masalah multikolinearitas dan membandingkan nilai dugaannya dengan Metode Kuadrat Terkecil (MKT) pada data kasus DBD di Indonesia serta mengetahui faktor-faktor apa saja yang mempengaruhi kasus DBD di Indonesia. Dari hasil penelitian, dapat disimpulkan bahwa metode *Elastic-Net* berhasil mengatasi masalah multikolinearitas dan menghasilkan dugaan koefisien regresi yang lebih baik dibandingkan dengan MKT berdasarkan pada pengukuran AIC. Adapun dari analisis yang dilakukan diperoleh faktor-faktor yang mempengaruhi kasus DBD di Indonesia yaitu, jumlah tenaga kesehatan lingkungan (X_2), persentase rumah tangga dengan sanitasi layak (X_3), persentase rumah tangga dengan rumah layak huni (X_4), persentase rumah tangga dengan air minum layak (X_5), jumlah curah hujan (X_6), jumlah hari hujan (X_7), dan kepadatan penduduk (X_8).

Kata Kunci: Analisis Regresi, Multikolinearitas, *Elastic-Net*, AIC, Demam Berdarah *Dengue* (DBD).

**ANALISIS REGRESI *ELASTIC-NET* UNTUK MENGATASI MASALAH
MULTIKOLINEARITAS PADA DATA DEMAM BERDARAH *DENGUE*
(DBD) DI INDONESIA**

Oleh

ADINDA SARIANTI

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

Judul Skripsi

: **ANALISIS REGRESI *ELASTIC-NET*
UNTUK MENGATASI MASALAH
MULTIKOLINEARITAS PADA DATA
DEMAM BERDARAH *DENGUE* (DBD) DI
INDONESIA**

Nama Mahasiswa

: **Adinda Sarianti**

Nomor Pokok Mahasiswa

: **1957031013**

Program Studi

: **Matematika**

Fakultas

: **Matematika dan Ilmu Pengetahuan Alam**



Prof. Ir. Netti Herawati, M.Sc., Ph.D.
NIP. 196501251990032001

Dr. Agus Sutrisno, S.Si., M.Si.
NIP. 197008311999031002

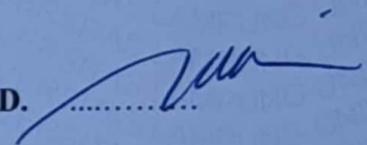
2. **Ketua Jurusan Matematika**

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

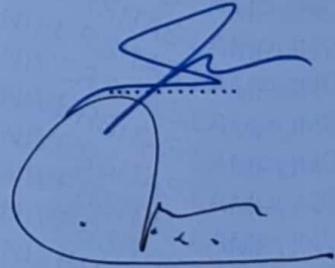
MENGESAHKAN

1. Tim Penguji

Ketua : **Prof. Ir. Netti Herawati, M.Sc., Ph.D.**



Sekretaris : **Dr. Agus Sutrisno, S.Si., M.Si.**



Penguji
Bukan Pembimbing : **Dr. Subian Saidi, S.Si., M.Si.**

.....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : **8 Agustus 2024**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan dibawah ini:

Nama Mahasiswa : **ADINDA SARIANTI**

Nomor Pokok Mahasiswa : **1957031013**

Jurusan : **Matematika**

Judul Skripsi : **ANALISIS REGRESI *ELASTIC-NET*
UNTUK MENGATASI MASALAH
MULTIKOLINEARITAS PADA DATA
DEMAM BERDARAH *DENGUE* (DBD) DI
INDONESIA**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 22 Agustus 2024
Penulis



Adinda Sarianti
NPM. 1957031013

RIWAYAT HIDUP

Penulis bernama lengkap Adinda Sarianti yang lahir di Kotabumi, Lampung Utara pada tanggal 17 September 2001 dan merupakan anak kedua dari lima bersaudara, putri dari pasangan Bapak Hendry dan Ibu Fitri. Penulis menyelesaikan pendidikan Sekolah Dasar (SD) di SDN 1 Labuhan Ratu pada tahun 2007 sampai 2013, Sekolah Menengah Pertama (SMP) di SMPN 22 Bandar Lampung pada tahun 2013 sampai 2016, dan Sekolah Menengah Atas (SMA) di SMA Al Kautsar Bandar Lampung pada tahun 2016 sampai 2019.

Pada tahun 2019 penulis diterima sebagai mahasiswa S1 Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur mandiri. Selama aktif menjadi mahasiswa, penulis ikut serta dalam UKMF Kerohanian Islam (ROIS) sebagai anggota Biro Kemuslimahan Tahun 2020.

Sebagai bentuk penerapan bidang ilmu di dunia kerja, penulis melaksanakan Kerja Praktek (KP) pada tahun 2022 di Dinas Pendidikan dan Kebudayaan Provinsi Lampung dan sebagai bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Kuripan, Kecamatan Pesisir Utara, Kabupaten Pesisir Barat.

KATA INSPIRASI

“Maka sesungguhnya bersama kesulitan ada kemudahan”
(QS. Al-Insyirah : 5)

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya”
(QS. Al-Baqarah : 286)

“Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sebelum mereka
mengubah keadaan diri mereka sendiri”
(QS. Ar-Ra'd : 11)

PERSEMBAHAN

Alhamdulillah rabbil'alamin, puji dan syukur kepada Allah SWT atas nikmat serta hidayahnya sehingga skripsi ini dapat terselesaikan dengan baik, serta salawat dan salam kepada junjungan kita Nabi Muhammad SAW.

Kupersembahkan karya ini kepada:

Diri Sendiri

Terima kasih untuk setiap hal yang sudah diusahakan dan selalu berusaha untuk menjadi pribadi yang lebih baik.

Ayah, Ibu, Kakak, dan Adik

Orang tuaku tercinta, Bapak Hendry dan Ibu Fitri, kakakku tersayang Firinda Soniya, serta adik-adikku tersayang, Adam Kurnianda, Fahira Almadina, dan Fahari Islamy yang selalu memberikan doa, dukungan, dan kasih sayang.

Dosen

Dosen-dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, dan memberikan arahan serta ilmu yang berharga.

Sahabat-sahabatku

Terima kasih kepada semua orang-orang baik yang telah memberikan pengalaman, semangat, motivasi, serta doa-doanya dan senantiasa memberikan dukungan dalam hal apapun.

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji dan syukur penulis ucapkan kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Regresi *Elastic-Net* untuk Mengatasi Masalah Multikolinearitas pada Data Demam Berdarah *Dengue* (DBD) di Indonesia”.

Penulis menyadari bahwa dalam penulisan skripsi ini tidak terlepas dari bimbingan, dukungan, bantuan, dan doa dari berbagai pihak. Oleh karena itu, dengan ketulusan hati penulis ingin menyampaikan terima kasih kepada:

1. Ibu Prof. Ir. Netti Herawati, M.Sc., Ph.D. selaku Dosen Pembimbing I dan Dosen Pembimbing Akademik yang senantiasa membimbing, memberi masukan, saran serta mendukung penulis dalam menyelesaikan skripsi ini.
2. Bapak Dr. Agus Sutrisno, S.Si., M.Si. selaku Dosen Pembimbing II yang telah memberikan bimbingan serta saran yang membantu kepada penulis dalam proses penyelesaian skripsi ini.
3. Bapak Dr. Subian Saidi, S.Si., M.Si. selaku Dosen Pembahas atas ketersediaannya untuk membahas serta memberikan kritik dan saran serta evaluasi kepada penulis dalam penyelesaian skripsi ini.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.
5. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.
6. Seluruh dosen, staff, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.

7. Ayah, ibu, kakak, adik, dan seluruh keluarga besar yang selalu memberikan kasih sayang, dukungan, nasihat, motivasi serta doa kepada penulis.
8. Sahabat-sahabat terbaik yang tidak lelah memberikan semangat, doa, dan waktunya untuk membantu penulis dalam menyelesaikan skripsi ini.
9. Tim penutupan, Fitdes, Yuni, Siti, Clara, Mia, Lisna, Vista, Amanda, Herlina, Amel, Nufus, dan Ikhsan. Terima kasih untuk semangat dan dukungan yang sangat berarti.
10. Teman-teman Matematika 2019, terima kasih atas kebersamaannya.
11. Almamater tercinta, Universitas Lampung.
12. Seluruh pihak yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis mengharapkan masukan serta saran untuk dijadikan pelajaran kedepannya.

Bandar Lampung, 22 Agustus 2024
Penulis

Adinda Sarianti

DAFTAR ISI

	Halaman
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvi
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah.....	1
1.2 Tujuan Penelitian	3
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA	4
2.1 Analisis Regresi	4
2.2 Regresi Linear Berganda.....	5
2.3 Metode Kuadrat Terekcil (MKT).....	6
2.4 Multikolinearitas	9
2.4.1 Dampak Multikolinearitas.....	9
2.5 Ukuran Pemusatan dan Penskalaan (<i>Centering and Scaling</i>).....	10
2.6 <i>Elastic-Net</i>	11
2.7 <i>Akaike Information Criterion</i> (AIC)	12
2.8 Validasi Silang (<i>Cross Validation</i>)	13
2.9 Demam Berdarah <i>Dengue</i> (DBD).....	13
III. METODOLOGI PENELITIAN	15
3.1 Waktu dan Tempat Penelitian	15
3.2 Data Penelitian	15
3.3 Metode Penelitian	16
IV. HASIL DAN PEMBAHASAN	18
4.1 Deteksi Multikolinearitas.....	18
4.2 Analisis Model dengan Metode Kuadrat Terkecil (MKT)	19
4.3 Analisis Model dengan Regresi <i>Elastic-Net</i>	20
4.3.1 Pemilihan Nilai Parameter Penyusutan (λ) Optimum.....	20
4.3.2 Nilai Koefisien Regresi <i>Elastic-Net</i>	21

4.4 Perbandingan Nilai Koefisien Regresi Model MKT dan <i>Elastic-Net</i>	22
4.5 Pemilihan Model Terbaik	23
V. KESIMPULAN	25
DAFTAR PUSTAKA	26
LAMPIRAN	28

DAFTAR TABEL

Tabel	Halaman
1. Variabel Data	15
2. Nilai Korelasi antar Variabel Bebas.....	18
3. Nilai VIF Masing-Masing Variabel Bebas	19
4. Estimasi Parameter dengan MKT	20
5. Nilai Koefisien Regresi pada Model MKT dan <i>Elastic-Net</i>	22
6. Perbandingan Nilai AIC pada MKT dan <i>Elastic-Net</i>	23

DAFTAR GAMBAR

Gambar	Halaman
1. Plot <i>Cross Validation</i> Parameter Penyusutan (λ) Regresi <i>Elastic-Net</i>	21
2. Plot Koefisien Regresi <i>Elastic-Net</i>	22

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan analisis data yang bertujuan untuk menjelaskan bagaimana variabel terikat berhubungan dengan satu atau lebih variabel bebas (Hosmer & Lemeshow, 2000). Namun dalam penerapannya tidak sedikit muncul masalah spesifik ketika proses analisis dilakukan, salah satunya yaitu masalah multikolinearitas.

Multikolinearitas merupakan salah satu masalah yang sering muncul saat melakukan analisis regresi berganda, yang terjadi ketika dua atau lebih variabel bebas memiliki korelasi yang kuat (Mait dkk., 2021). Kehadiran multikolinearitas dapat mengakibatkan estimasi koefisien dalam model regresi menjadi tidak stabil sehingga dapat menghasilkan *standard error* yang besar (Altelbany, 2021). Beberapa metode bisa digunakan untuk menangani masalah multikolinearitas diantaranya dengan menambah sampel baru, mengganti atau mengeluarkan variabel yang mempunyai korelasi yang tinggi, regresi komponen utama, dan beberapa metode lainnya (Montgomery *et al.*, 2012). Menurut Fikri (2023), masalah multikolinearitas juga dapat diatasi dengan menggunakan metode regularisasi, yang dapat menyusutkan koefisien estimasi menjadi nol. Adapun metode regularisasi yang umum digunakan adalah Regresi Ridge, *Least Absolute Shrinkage and Selection Operator* (LASSO), dan *Elastic-Net*.

Metode Regresi *Elastic-Net* merupakan metode yang menggabungkan nilai penalti dari kedua metode, yaitu Regresi Ridge dan LASSO, dimana regresi ini mampu mengatasi kekurangan dari Regresi Ridge dan LASSO yang dapat menyusutkan koefisien regresi tepat nol juga menyeleksi peubah secara simultan (Zou & Hastie, 2005).

Untuk mencapai output optimal dalam penelitian yang dilakukan, tidak sedikit peneliti yang menggunakan metode regularisasi. Seperti pada penelitian yang dilakukan oleh Herawati *et al.* (2024), mengenai kinerja Regresi Ridge, LASSO, dan *Elastic-Net* dalam menangani masalah multikolinearitas menggunakan data simulasi dan aplikasinya pada data *real*. Penelitian lain juga dilakukan oleh Altalbany (2021), peneliti membahas tentang penanganan masalah multikolinearitas pada data simulasi yang kemudian dibandingkan kinerja masing-masing model Regresi Ridge, Regresi LASSO, dan Regresi *Elastic-Net*. Penelitian-penelitian tersebut menunjukkan bahwa Regresi *Elastic-Net* memberikan hasil terbaik.

Dalam penelitian ini peneliti akan menggunakan data kesehatan, khususnya data kasus Demam Berdarah *Dengue* (DBD) di Indonesia. Di Indonesia, DBD menjadi salah satu tantangan utama dalam bidang kesehatan masyarakat dan dapat menyerang semua golongan umur. Penyakit DBD merupakan penyakit menular tidak langsung yang ditularkan melalui gigitan nyamuk. Gejala penyakit DBD ditandai dengan demam tinggi selama 2-7 hari disertai sakit kepala, nyeri sendi (*myalgia*) dan otot (*arthralgia*) serta ruam pada kulit (Setiawan dkk., 2009). Adapun faktor-faktor yang berpengaruh terhadap kasus ini diantaranya jumlah rumah sakit, rumah tangga dengan sanitasi layak, perilaku masyarakat, iklim, dan faktor-faktor lainnya. Setelah dilakukan penelitian terdahulu, peneliti menemukan adanya masalah multikolinearitas pada data, mengingat banyaknya faktor yang mempengaruhi kasus ini menyebabkan kemungkinan antar variabel bebas memiliki korelasi yang tinggi menjadi sangat besar, sehingga rawan mengalami masalah multikolinearitas.

Bedasarkan uraian diatas, peneliti pun tertarik untuk melakukan penelitian mengenai penanganan masalah multikolinearitas menggunakan Regresi *Elastic-Net* pada data DBD di Indonesia.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui kinerja metode Regresi *Elastic-Net* dalam mengatasi masalah multikolinearitas.
2. Membandingkan nilai duga koefisien regresi menggunakan *Elastic-Net* dengan Metode Kuadrat Terkecil (MKT).
3. Mengetahui faktor-faktor yang mempengaruhi angka kasus *Incidance Rate* (IR) DBD di Indonesia.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Menambah wawasan bagi peneliti mengenai metode Regresi *Elastic-Net* dalam mengatasi masalah multikolinearitas.
2. Menambah referensi bagi pembaca mengenai penerapan metode Regresi *Elastic-Net* dalam mengatasi masalah multikolinearitas.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan analisis data yang bertujuan untuk menjelaskan bagaimana variabel terikat berhubungan dengan satu atau lebih variabel bebas (Hosmer & Lemeshow, 2000). Terdapat dua jenis analisis regresi, yaitu regresi linear sederhana yang digunakan untuk data dengan satu variabel terikat dan satu variabel bebas, serta regresi linear berganda yang digunakan untuk data dengan satu variabel terikat dan lebih dari satu variabel bebas.

Secara matematis, hubungan antara satu variabel bebas dan satu variabel terikat dapat dinyatakan dalam persamaan regresi berikut:

$$Y_i = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

untuk $i = 1, 2, \dots, n$

dengan:

Y_i = Variabel terikat untuk pengamatan ke- i

X = Variabel bebas

β_0 = Konstanta (parameter)

β_1 = Koefisien regresi

ε = Sisaan (galat)

2.2 Regresi Linear Berganda

Regresi linear berganda merupakan jenis analisis regresi yang menganalisis hubungan linear antara satu variabel terikat dengan dua atau lebih variabel bebas. Menurut Draper & Smith (1998), persamaan regresi linear berganda dapat dirumuskan dalam bentuk matematis sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (2.2)$$

untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$

dengan:

- Y_i = Variabel terikat untuk pengamatan ke- i
- β_0 = Konstanta (parameter)
- β_k = Koefisien regresi ke- k
- X_{ki} = Variabel bebas pengamatan ke- i
- ε = Sisaan (galat) pengamatan ke- i

Adapun persamaan linear untuk pendugaan garis regresi linear berganda dapat dituliskan sebagai berikut:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} \quad (2.3)$$

untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$

dengan:

- \hat{y}_i = Nilai dugaan variabel terikat pengamatan ke- i
- b_0 = Konstanta (parameter)
- b_k = Koefisien regresi ke- k
- x_{ki} = Nilai variabel bebas pengamatan ke- i

Dalam bentuk matriks, persamaan regresi linear berganda dapat ditulis sebagai berikut:

$$Y = X\beta + \varepsilon \quad (2.4)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{1i} & X_{12} & \dots & X_{1k} \\ 1 & X_{2i} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}$$

dengan:

- Y = Vektor variabel terikat berukuran $n \times 1$
- X = Matriks variabel bebas berukuran $n \times k$
- β = Vektor parameter berukuran $k \times 1$
- ε = Vektor galat berukuran $n \times 1$

2.3 Metode Kuadrat Terkecil (MKT)

Metode Kuadrat Terkecil (MKT) atau yang biasa juga dikenal dengan *Ordinary Least Square* (OLS) merupakan metode yang dipakai untuk mengestimasi parameter $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ dengan cara meminimalkan Jumlah Kuadrat Galat (JKG) (Montgomery *et al.*, 2012).

Nilai β dapat diperoleh dengan meminimalkan bentuk kuadrat:

$$Q(\beta_j) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j \right)^2 \quad (2.5)$$

Jumlah kuadrat galat ε_i^2 dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\varepsilon_i^T \varepsilon_i = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_n] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 = \sum \varepsilon_i^2 \quad (2.6)$$

Berdasarkan persamaan umum regresi linear berganda dengan matriks diperoleh:

$$\varepsilon = Y - X\beta \quad (2.7)$$

Oleh karena itu, perkalian matriks galat dapat ditulis sebagai berikut:

$$\varepsilon_i^T \varepsilon_i = (Y - X\beta^T)(Y - X\beta)$$

$$\begin{aligned}
&= Y^t Y - Y^T X \beta - X^T \beta^T Y + X^T \beta^T X \beta \\
&\quad (\text{karena } X^T \beta^T Y = Y^T X \beta) \\
&= Y^T Y - 2Y^T X \beta + X^T \beta^T X \beta
\end{aligned}$$

Selanjutnya adalah menentukan turunan $\varepsilon_i^T \varepsilon_i$ terhadap β secara parsial dan menyamakannya dengan nol.

$$\begin{aligned}
\frac{\partial \varepsilon_i^T \varepsilon_i}{\partial \beta} &= 0 - 2Y^T X \beta + \beta^T X^T X \beta = 0 \\
-2Y^T X \beta + \beta^T X^T X \beta &= 0 \\
-2X^T Y + 2X^T X \beta &= 0 \\
2X^T X \beta &= 2X^T Y \\
X^T X \beta &= X^T Y \\
(X^T X)^{-1} X^T X \beta &= (X^T X)^{-1} X^T Y \\
I \beta &= (X^T X)^{-1} X^T Y \\
\beta &= (X^T X)^{-1} X^T Y
\end{aligned}$$

Sehingga diperoleh penduga untuk MKT sebagai berikut:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.8)$$

Adapun sifat-sifat dari penduga MKT dapat diuraikan sebagai berikut:

1. $\hat{\beta}$ linear

$\hat{\beta}$ linear jika β merupakan fungsi linear dari β

$$\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T Y \\
&= (X^T X)^{-1} X^T (X \beta + \varepsilon) \\
&= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \varepsilon \\
&= I \beta + (X^T X)^{-1} X^T \varepsilon
\end{aligned}$$

2. $\hat{\beta}$ adalah penduga tak bias jika $E(\hat{\beta}) = \beta$

$$\begin{aligned}
E(\hat{\beta}) &= E((X^T X)^{-1} X^T Y) \\
&= E((X^T X)^{-1} X^T X \beta + \varepsilon) \\
&= E((X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \varepsilon)
\end{aligned}$$

$$\begin{aligned}
&= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T E(\varepsilon) \\
&= (X^T X)^{-1} X^T X \beta \\
&= I \beta \\
&= \beta
\end{aligned}$$

Karena $E(\hat{\beta}) = \beta$, maka $\hat{\beta}$ merupakan penduga tak bias dari β

3. $\hat{\beta}$ memiliki varians minimum

$$\begin{aligned}
\text{Var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\
&= E[((X^T X)^{-1} X^T Y - \beta)((X^T X)^{-1} X^T Y - \beta)^T] \\
&= E[((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)^T] \\
&= E[((X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \varepsilon - \beta)] \\
&\quad ((X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \varepsilon - \beta)^T] \\
&= E[(I\beta + (X^T X)^{-1} X^T \varepsilon - \beta)(I\beta + (X^T X)^{-1} X^T \varepsilon - \beta)^T] \\
&= E[(\beta + (X^T X)^{-1} X^T \varepsilon - \beta)(\beta + (X^T X)^{-1} X^T \varepsilon - \beta)^T] \\
&= E[((X^T X)^{-1} X^T \varepsilon)((X^T X)^{-1} X^T \varepsilon)^T] \\
&= E[((X^T X)^{-1} X^T \varepsilon^T \varepsilon X (X^T X)^{-1}] \\
&= (X^T X)^{-1} X^T X (X^T X)^{-1} E[\varepsilon \varepsilon^T] \\
&= (X^T X)^{-1} \sigma^2
\end{aligned}$$

$\text{Var}(\hat{\beta}) = (X^T X)^{-1} \sigma^2$ merupakan varians terkecil dari semua estimasi linear tak bias.

Jika terpenuhi sifat linear, tak bias, dan mempunyai varians minimum pada estimator MKT, maka dapat dikatakan bahwa estimator tersebut bersifat *Best Linear Unbiased Estimator* (BLUE).

2.4 Multikolinearitas

Menurut Gujarati & Porter (2009), multikolinearitas adalah adanya hubungan linear yang kuat di antara beberapa atau semua variabel bebas dalam model regresi.

Terdapat beberapa cara untuk mengetahui apakah multikolinearitas muncul dalam model regresi, antara lain:

1. Melihat korelasi yang kuat antar variabel bebas
Korelasi yang kuat antar variabel bebas, misalnya sebesar 0.8, dapat mengindikasikan adanya masalah multikolinearitas dalam model.
2. Melihat nilai *Variance Inflation Factor* (VIF) pada model regresi
Nilai VIF dapat diperoleh menggunakan rumus berikut:

$$VIF_{(j)} = \frac{1}{1 - R_j^2} \quad (2.9)$$

Dengan R_j^2 merupakan koefisien determinasi yang didapat dari variabel bebas X_j yang diregresikan dengan variabel regresi lainnya. Jika nilai VIF yang diperoleh lebih besar dari 10, maka hal ini menunjukkan adanya masalah multikolinearitas.

2.4.1 Dampak Multikolinearitas

Menurut Gujarati & Porter (2009), masalah multikolinearitas dapat memberikan dampak pada model regresi, antara lain:

1. Adanya multikolinearitas menyebabkan varian dan kovarian pada penduga MKT menjadi besar, sehingga akan sulit untuk mendapatkan pendugaan yang tepat.
2. Selang kepercayaan cenderung lebih besar dikarenakan adanya nilai *standard error* yang tinggi.

3. Pengaruh variabel bebas terhadap variabel terikat dalam statistik uji t cenderung tidak signifikan.
4. Nilai statistik uji t menjadi tidak signifikan meskipun nilai R^2 secara keseluruhan sangat tinggi.
5. Perubahan kecil pada data dapat membuat penduga MKT dan *standard error* menjadi sangat sensitif.

2.5 Ukuran Pemusatan dan Penskalaan (*Centering and Scaling*)

Menurut Kutner *et al.* (2005), bagian dari pembakuan (*standardized*) variabel mencakup pemusatan dan penskalaan data. Ketika data tidak distandarisasi maka akan sulit untuk membandingkan nilai koefisien regresinya yang dapat mengakibatkan kesimpulan atau keputusan yang salah. Dalam konteks pembakuan variabel, transformasi korelasi adalah salah satu modifikasi sederhana yang dapat dilakukan. Adapun pemusatan adalah selisih antara setiap nilai pengamatan dan rata-rata semua pengamatan dalam variabel. Sedangkan penskalaan mencakup penyesuaian pengamatan ke dalam satuan deviasi estándar dari variabel. Namun pada Regresi *Elastic-Net* diasumsikan bahwa variabel terikat (Y) berpusat di sekitar nol untuk menghindari masalah intersep yang seharusnya tidak dikenai penalti, sedangkan variabel bebas (X) distandarisasi (Zou & Hastie, 2005). Variabel terikat (Y) yang berpusat di sekitar nol dan pembakuan variabel bebas (X_1, X_2, \dots, X_k) ditunjukkan pada rumus berikut:

$$Y - \text{mean}(Y)$$

$$X_j^* = \frac{X_{ij} - \bar{X}_j}{S_{xj}} \text{ dengan } S_{xj} = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n - 1}} \quad (2.10)$$

untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$

dengan:

$$\bar{X}_j = \text{Rata-rata dari pengamatan } X_j$$

S_{x_j} = Standar deviasi dari X_j

Model regresi berganda yang telah distandarisasi merupakan hasil dari transformasi model regresi berganda, yang dikenal sebagai transformasi korelasi. Berikut adalah penulisan model regresi yang telah distandarisasi:

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \dots + \beta_k^* X_{ki}^* + \varepsilon_i^* \quad (2.11)$$

2.6 *Elastic-Net*

Elastic-Net merupakan metode regresi penalti (*penalized regression*) yang mengkombinasikan nilai penalti Regresi Ridge dan Regresi LASSO, dimana metode ini dapat menangani masalah multikolinearitas. Menurut Zou & Hastie (2005), metode *Elastic-Net* mampu menyusutkan koefisien regresi hingga nol, dan juga melakukan seleksi variabel secara simultan serta dapat memilih kelompok variabel yang berkorelasi.

Elastic-Net dikembangkan untuk mengatasi kekurangan yang ada pada Regresi LASSO, berikut adalah kekurangan dari Regresi LASSO:

1. Saat jumlah variabel (p) melebihi jumlah pengamatan (n), LASSO hanya memilih n variabel untuk dimasukkan ke dalam model.
2. Apabila terdapat beberapa variabel yang berkorelasi tinggi, LASSO akan memilih salah satu variabel secara acak.
3. Saat jumlah variabel (p) lebih sedikit dari jumlah pengamatan (n), kinerja LASSO akan didominasi oleh Regresi Ridge.

Penalti *Elastic-Net* dituliskan sebagai berikut:

$$\sum_{j=1}^p [\alpha |\beta_j| + (1 - \alpha) \beta_j^2]$$

Penduga koefisien pada *Elastic-Net* dapat dituliskan sebagai berikut:

$$\hat{\beta}^{net} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (2.12)$$

dimana $\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}$, $0 \leq \alpha \leq 1$.

Nilai α merupakan kombinasi koefisien penyusutan antara Regresi Ridge dan Regresi LASSO. Jika nilai $\alpha = 0$ maka penalti yang digunakan adalah penalti Regresi Ridge, namun jika nilai $\alpha = 1$ maka penalti yang digunakan adalah penalti Regresi LASSO.

2.7 Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) merupakan metode untuk memilih model terbaik yang diperkenalkan oleh Akaike dan Schwarz. Menurut Herawati *et al.* (2024), penentuan model terbaik didasarkan pada nilai AIC terkecil.

Penghitungan nilai AIC dilakukan menggunakan rumus berikut:

$$AIC = n \ln \left(\frac{JKG}{n} \right) + 2p \quad (2.13)$$

dengan:

JKG = Jumlah kesalahan kuadrat galat

n = Jumlah data

p = Jumlah parameter

2.8 Validasi Silang (*Cross Validation*)

Validasi silang atau *Cross Validation* (CV) adalah metode yang digunakan untuk memperkirakan galat prediksi guna meningkatkan ketepatan dalam pemilihan model (James *et al.*, 2013). Metode validasi silang bekerja dengan cara membagi data menjadi dua bagian, yaitu data *training* dan data *testing*. Selain itu, validasi silang juga dapat membatasi terjadinya *overfitting* melalui pengujian model pada tahap *training*. Menurut Efron & Tibshirani (1993), metode *k-fold* adalah salah satu pendekatan dalam validasi silang, dimana seluruh data dibagi secara acak menjadi *k-fold*. Lipatan (*fold*) pertama digunakan sebagai data uji, sementara sisa dari *k-1 fold* lainnya digunakan untuk melatih model. Langkah ini diulang sampai *k* kali hingga model telah dilatih dan dievaluasi pada seluruh kumpulan data. Persamaan berikut digunakan oleh CV untuk mengestimasi nilai galat prediksi (\widehat{PE}):

$$\widehat{PE} = CV = \frac{1}{k} \sum_{i=1}^k (y_i - \hat{y}_{-i(k)})^2 \quad (2.14)$$

dimana $\hat{y}_{-i(k)}$ merupakan nilai estimasi y saat *fold* ke- k tidak digunakan untuk menduga model. Dalam prakteknya, sering kali penggunaan *k-fold* CV menggunakan $k = 5$ atau $k = 10$. Penggunaan $k = 5$ atau $k = 10$ pada *k-fold* karena nilai-nilai tersebut dapat menghasilkan perkiraan tingkat kesalahan pengujian yang tidak mengalami bias yang terlalu tinggi maupun varian yang sangat tinggi.

2.9 Demam Berdarah *Dengue* (DBD)

Demam Berdarah *Dengue* (DBD) merupakan penyakit yang disebabkan oleh virus *dengue* dan ditularkan melalui gigitan nyamuk *Aedes aegypti*. Menurut Taryono dkk. (2018), virus *dengue* membutuhkan waktu sekitar 3-14 hari untuk berkembang di dalam tubuh manusia sebelum gejala mulai muncul, dengan gejala umumnya muncul antara hari keempat hingga ketujuh. Sementara itu, di dalam

tubuh nyamuk, virus *dengue* berkembang selama sekitar 8-10 hari. Adapun gejala penyakit DBD ditandai dengan demam tinggi selama 2-7 hari disertai sakit kepala, nyeri sendi (*myalgia*) dan otot (*arthralgia*) serta ruam pada kulit (Setiawan dkk., 2009). Jika penyakit ini tidak ditangani, maka dapat memicu terjadinya Kejadian Luar Biasa (KLB) hingga berakhir kematian. Banyak faktor yang mempengaruhi penyebaran penyakit ini antara lain faktor iklim, sosial ekonomi, dan faktor lingkungan. Sejak pertama kali ditemukan di Indonesia pada tahun 1968, jumlah kasus ini terus meningkat dan wilayah penyebarannya pun meluas. Meskipun pemerintah telah melakukan berbagai upaya pencegahan dan penanggulangan sejak dimulainya program pada tahun 1970, namun upaya tersebut belum berhasil menekan penyebaran penyakit ini seperti yang diharapkan. Dalam upaya mencapai nol kematian akibat *dengue* pada tahun 2030, Rencana Strategis (Renstra) Kementerian Kesehatan 2020-2024 menetapkan target indikator yaitu 95% kabupaten/kota dengan *Incidence Rate* (IR) DBD $\leq 10/100.000$ penduduk pada tahun 2024 (Kemenkes, 2022).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung pada semester genap tahun ajaran 2023/2024.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari situs resmi Kementerian Kesehatan Republik Indonesia yaitu data kasus Demam Berdarah *Dengue* (DBD) di Indonesia Tahun 2021 dengan data faktor iklim yang diperoleh dari situs resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), dengan jumlah data yang digunakan sebanyak 34 data observasi yang terdiri dari 9 variabel bebas dan 1 variabel terikat yang akan diteliti.

Tabel 1. Variabel Data

Variabel	Keterangan
Y	<i>Incidance Rate</i> (IR) Demam Berdarah <i>Dengue</i> (DBD) (persen)
X_1	Jumlah Rumah Sakit
X_2	Jumlah Tenaga Kesehatan Lingkungan (jiwa)
X_3	Persentase Rumah Tangga dengan Sanitasi Layak (persen)

Tabel 1. (Lanjutan)

X_4	Persentase Rumah Tangga dengan Rumah Layak Huni (persen)
X_5	Persentase Rumah Tangga dengan Air Minum Layak (persen)
X_6	Jumlah Curah Hujan (mm)
X_7	Jumlah Hari Hujan
X_8	Kepadatan Penduduk (per km ²)
X_9	Jumlah Penduduk Miskin (ribu orang)

3.3 Metode Penelitian

Penelitian ini dilakukan melalui studi pustaka, yaitu dengan menelaah referensi dari buku, skripsi, dan jurnal. Untuk mempermudah perhitungan dan memperoleh hasil yang tepat, penulis menggunakan *software* RStudio versi 4.2.1 dalam penelitian ini. Berikut adalah langkah-langkah yang dilakukan dalam penelitian ini:

1. Mengidentifikasi multikolinearitas dengan mempertimbangkan korelasi antar variabel bebas dan nilai *Variance Inflation Factor* (VIF)
2. Menganalisis model regresi menggunakan MKT
 - a. Melakukan proses *center* dan *scale* data, karena data memiliki satuan yang berbeda-beda
 - b. Analisis model regresi menggunakan MKT
3. Melakukan analisis model regresi dengan metode *Elastic-Net*
 - a. Melakukan proses *center* dan *scale* data, karena data memiliki satuan yang berbeda-beda
 - b. Menentukan nilai λ yang optimal menggunakan metode validasi silang
 - c. Mendapat nilai duga Regresi *Elastic-Net* berdasarkan nilai λ yang optimal
 - d. Menganalisis model regresi yang dihasilkan menggunakan metode Regresi *Elastic-Net*
4. Membandingkan nilai koefisien yang dihasilkan oleh MKT dan Regresi *Elastic-Net*

5. Menghitung nilai AIC untuk setiap model yang dihasilkan oleh MKT dan *Elastic-Net*
6. Menentukan model terbaik berdasarkan nilai AIC terkecil dari masing-masing model yang diperoleh dari MKT dan *Elastic-Net*

V. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan untuk mengatasi masalah multikolinearitas pada data DBD di Indonesia tahun 2021 menggunakan Regresi *Elastic-Net*, maka dapat disimpulkan bahwa model Regresi *Elastic-Net* memberikan hasil yang lebih baik dalam menduga variabel bebas dibandingkan dengan model MKT berdasarkan nilai AIC yang diperoleh. Sehingga diperoleh model terbaik dengan Regresi *Elastic-Net* pada data kasus DBD di Indonesia tahun 2021 sebagai berikut:

$$\hat{Y} = -7,026044 \times 10^{-15} - 2,064105 X_2 + 1,085229 X_3 - 8,090547 X_4 + 5,838272 X_5 - 4,457860 X_6 - 1,258460 X_7 - 5,628242 X_8$$

Dari model tersebut maka dapat diketahui bahwa kasus DBD dipengaruhi oleh jumlah tenaga kesehatan lingkungan (X_2), persentase rumah tangga dengan sanitasi layak (X_3), persentase rumah tangga dengan rumah layak huni (X_4), persentase rumah tangga dengan air minum layak (X_5), jumlah curah hujan (X_6), jumlah hari hujan (X_7), dan kepadatan penduduk (X_8).

DAFTAR PUSTAKA

- Altelbany, S. 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*. **5**(1): 131–142.
- Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). 2023. Jumlah Curah Hujan dan Hari Hujan. <https://www.bmkg.go.id/>. Diakses pada 24 Agustus 2023.
- Draper, N.R. & Smith, H. 1998. *Applied Regression Analysis*. 3rd Edition. John Wiley & Sons, New York.
- Efron, B. & Tibshirani, R. 1993. *An Introduction to the Bootstrap*. 1st Edition. Chapman & Hall, New York.
- Fikri, A.F., Agwil., & Agustina, D. 2023. Performa Teknik Regularisasi dalam Penanganan Masalah Multikolinearitas. *DJMA: Diophantine Journal of Mathematics and Its Applications*. **2**(1): 46-51.
- Gujarati, D.N. & Porter, D.C. 2009. *Basic Econometrics*. McGraw-Hill Irwin, Boston.
- Herawati, N., Wijayanti, A., Sutrisno, A., Nusyirwan, & Misgiyati. 2024. The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application. *Journal of Modern Applied Statistical Methods*. **23**(2): 3-13.
- Hosmer, D.W. & Lemeshow, S. 2000. *Applied Logistic Regression*. 2nd Edition. John Wiley and Sons, New York.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. *An Introduction to Statistical Learning with Applications in R*. Springer, New York.
- Kementerian Kesehatan Republik Indonesia. 2022. *Laporan Tahunan 2022 Demam Berdarah Dengue*. Kemenkes RI, Jakarta.
- Kementerian Kesehatan Republik Indonesia. 2023. Profil Kesehatan Indonesia Tahun 2021. <https://repository.kemkes.go.id/book/828>. Diakses pada 6 Agustus 2023.
- Kurniasih, D., Mariani, S., & Sugiman. 2013. Efisiensi Relatif Estimator Fungsi Kernel Gaussian Terhadap Estimator Polinomial dalam Peramalan USD terhadap JPY. *UNNES Journal of Mathematics*. **2**(2): 80-84.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. 2005. *Applied Linear Statistical Models*. 5th Edition. McGraw-Hill Irwin, Boston.
- Mait, Y.A., Salaki, D.T., & Komalig, H.A.H. 2021. Kajian Model Prediksi Metode Least Absolute Shrinkage and Selection Operator (LASSO) pada Data Mengandung Multikolinearitas. *Jurnal Matematika dan Aplikasi*. **10**(2): 69-75.
- Montgomery, D.C., Peck, E.A., & Vining, G.G. 2012. *Introduction to Linear Regression Analysis*. 4th Edition. John Wiley and Sons, New York.
- Setiawan, B., Chen, K., & Pohan, H.T. 2009. Diagnosis dan Terapi Cairan pada Demam Berdarah Dengue. *Scientific Journal of Pharmaceutical and Medical Application*. **22**(1): 1-11.
- Taryono, A.P.N., Ispriyanti, D., & Prahutama, A. 2018. Analisis Faktor-Faktor yang Mempengaruhi Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Provinsi Jawa Tengah dengan Metode Spatial Autoregressive Model dan Spatial Durbin Model. *Indonesian Journal of Applied Statistics*. **1**(1): 1-13.
- Zou, H. & Hastie, T. 2005. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. **67**(2): 301-320.