PENANGANAN MASALAH MULTIKOLINEARITAS DENGAN REGRESI RIDGE, LASSO, DAN *ELASTIC-NET* PADA KASUS BALITA *STUNTING* DI INDONESIA

(Skripsi)

Oleh

AMELIANA WIJAYANTI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

ABSTRACT

HANDLING OF MULTICOLINEARITY PROBLEMS WITH RIDGE, LASSO, AND ELASTIC-NET REGRESSION IN CASES OF STUNTING TODDLERS IN INDONESIA

By

AMELIANA WIJAYANTI

Ridge, LASSO, and Elastic-Net regression are methods that are generally used in the context of linear regression when there are many predictor variables and multicollinearity problems exist. Ridge regression is used to reduce overfitting and control the variance of the model by adding a penalty term in the form of the sum of the squared coefficients in the estimated objective function. Similar to Ridge, LASSO uses a penalty in the form of the absolute number of coefficients in the estimated objective function. Meanwhile, Elastic-Net performs shrinkage and variable selection simultaneously by combining the penalties between Ridge Regression and LASSO. The purpose of this study is to determine the performance of Regression Ridge, LASSO, and Elastic-Net in dealing with multicollinearity problems in stunting toddler case data in Indonesia. The results of this study indicate that Elastic-Net is more effective in overcoming multicollinearity problems when compared to Ridge Regression and LASSO based on MSE and AIC measurements.

Keywords: Ridge Regression, LASSO, Elastic-Net, Multicollinearity, MSE, AIC.

ABSTRAK

PENANGANAN MASALAH MULTIKOLINEARITAS DENGAN REGRESI RIDGE, LASSO, DAN *ELASTIC-NET* PADA KASUS BALITA *STUNTING* DI INDONESIA

Oleh

AMELIANA WIJAYANTI

Regresi Ridge, LASSO, dan *Elastic-Net* merupakan metode yang umumnya digunakan dalam konteks regresi linear ketika terdapat banyak variabel prediktor dan adanya masalah multikolinearitas. Regresi Ridge digunakan untuk mengurangi overfitting dan mengendalikan varians model dengan menambahkan istilah penalti berupa jumlah kuadrat koefisien pada fungsi objektif yang diestimasi. Serupa dengan Ridge, LASSO menggunakan penalti berupa jumlah mutlak koefisien pada fungsi objektif yang diestimasi. Sedangkan *Elastic-Net* melakukan penyusutan dan seleksi peubah secara simultan dengan menggabungkan penalti antara Regresi Ridge dan LASSO. Tujuan dari penelitian ini yaitu mengetahui performa Regresi Ridge, LASSO, dan *Elastic-Net* dalam menangani masalah multikolinearitas pada data kasus balita *stunting* di Indonesia. Hasil dari penelitian ini mengindikasikan bahwa *Elastic-Net* secara lebih efektif mengatasi masalah multikolinearitas jika dibandingkan dengan Regresi Ridge dan LASSO berdasarkan pada pengukuran MSE dan AIC.

Kata Kunci: Regresi Ridge, LASSO, *Elastic-Net*, Multikolinearitas, MSE, AIC.

PENANGANAN MASALAH MULTIKOLINEARITAS DENGAN REGRESI RIDGE, LASSO, DAN *ELASTIC-NET* PADA KASUS BALITA *STUNTING* DI INDONESIA

Oleh

AMELIANA WIJAYANTI

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
2023

Judul Skripsi	: **PENANGANAN MASALAH MULTIKOLINEARITAS DENGAN REGRESI RIDGE, LASSO, DAN *ELASTIC-NET* PADA KASUS BALITA *STUNTING* DI INDONESIA**
Nama Mahasiswa	: **Ameliana Wijayanti**
Nomor Pokok Mahasiswa	: **1917031010**
Jurusan	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Ir. Netti Herawati, M.Sc.	**Agus Sutrisno, S.Si., M.Si.**
NIP. 196501251990032001	NIP. 197008311999031002

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman. S.Si.,M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Dr. Ir. Netti Herawati, M.Sc.**	………………..
Sekretaris	: **Agus Sutrisno, S.Si., M.Si.**	………………..
Penguji Bukan Pembimbing	: **Drs. Nusyirwan, M.Si**	………………..

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi : **06 Juni 2023**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa	: **Ameliana Wijayanti**
Nomor Pokok Mahasiswa	: **1917031010**
Jurusan	: **Matematika**
Judul Skripsi	: **PENANGANAN MASALAH MULTIKOLINEARITAS DENGAN REGRESI RIDGE, LASSO, DAN *ELASTIC-NET* PADA KASUS BALITA *STUNTING* 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, 06 Juni 2023
Penulis,

Ameliana Wijayanti
NPM. 1917031010

RIWAYAT HIDUP

Penulis bernama Ameliana Wijayanti lahir di Titiwangi pada 10 April 2001. Penulis merupakan anak pertama dari tiga bersaudara dari pasangan Bapak Sugianto dan Ibu Siti Masruroh.

Penulis menempuh pendidikan Sekolah dasar (SD) di SD Negeri 1 Waygelam pada tahun 2007 sampai dengan 2013. Kemudian melanjutkan ke Sekolah Menengah Pertama di SMP Negeri 1 Sidomulyo pada tahun 2013 sampai dengan 2016. Selanjutnya ke Sekolah Menengah Atas di SMA Negeri 1 Sidomulyo pada tahun 2016 sampai dengan 2019. Pada tahun 2019 penulis terdaftar sebagai mahasiswa Program Studi S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) serta menjadi penerima Beasiswa Bidikmisi.

Pada tahun 2020 dan 2021 penulis aktif di UKMF Natural, penulis diamanahkan menjadi Anggota Redaksi dan Multimedia. Pada tahun 2022 penulis melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik (BPS) Kabupaten Lampung Selatan sebagai aplikasi bidang ilmu di dunia kerja. Pada tahun yang sama, penulis melaksanakan Kuliah Kerja Nyata (KKN) di Pekon Margoyoso Kecamatan Sumber Rejo, Kabupaten Tanggamus, sebagai bentuk pengabdian kepada masyarakat.

KATA INSPIRASI

"... Dan Allah beserta orang-orang yang sabar."
(Q.S Al-Baqarah: 249)

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."
(Q.S Al-Baqarah: 286)

"Sesungguhnya bersama kesulitan ada kemudahan."
(Q.S Al-Insyirah: 6)

"Selagi hidup, berjuang. Sebab nanti setelah mati, kita juga pelan-pelan hilang."
(Boy Candra)

"Belajar itu tidak bisa dipaksakan, tapi perlu juga untuk dikondisikan.
(Merry Riana)

"Ikuti saja alurmu, dan fokus disana sampai jadi sesuatu."
(Ameliana Wijayanti)

PERSEMBAHAN

Dengan mengucap Alhamdulilah dan syukur kepada Allah SWT atas nikmat serta hidayahnya sehingga skripsi ini dapat terselesaikan dengan baik. Rasa syukur dan bahagia saya persembahkan karya ini kepada:

Kedua Orang Tua dan Adikku

Terima kasih kepada kedua orang tuaku tercinta, Bapak Sugianto dan Ibu Siti Masruroh atas segala pengorbanan, doa dan ridho kalian serta dukungannya selama ini. Terima kasih telah memberikan pelajaran berharga tentang makna perjalanan hidup yang sebenarnya sehingga kelak bisa manjadi orang yang bermanfaat bagi semua orang. Terimakasih juga kepada kedua adikku, Novita Lestari dan Ahmad Rifai yang selalu menghibur penulis.

Dosen Pembimbing dan Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Keluarga dan Sahabat-sahabatku

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

Almamater Tercinta

Universitas Lampung

SANWACANA

Segala puji dan syukur penulis ucapkan kepada Allah SWT atas segala nikmat dan karunia-Nya yang tak terhingga sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Penanganan Masalah Multikolinearitas dengan Regresi Ridge, LASSO, dan *Elastic-Net* pada Kasus Balita *Stunting* di Indonesia"**. Dalam penulisan skripsi ini tidak dapat terselesaikan tanpa adanya bimbingan, bantuan dan dukungan dari berbagai pihak.

Sehingga, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Dr. Ir. Netti Herawati, M.Sc. selaku dosen pembimbing I dan dosen pembimbing akademik yang senantiasa membimbing, memberi masukan, saran serta mendukung penulis dalam menyelesaikan skripsi ini.
2. Agus Sutrisno, S.Si., M.Si. selaku dosen pembimbing II memberikan bimbingan, pengarahan, serta saran sehingga penulis dapat menyelesaikan skripsi ini.
3. Drs. Nusyirwan, M.Si selaku dosen penguji yang telah memberikan kritik dan saran yang membangun sehingga skripsi ini dapat terselesaikan.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staff, karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Kedua orang tuaku, Bapak dan Ibu yang selalu mendukung, menemani, memberikan motivasi, mendoakan, serta memberikan semangat sehingga menguatkan penulis dalam menjalani setiap proses meraih gelar sarjana.

8. Terima kasih untuk diri sendiri yang telah berjuang, berproses, dan bertahan hingga dapat menyelesaikan skripsi.
9. Untuk kedua adikku Novita Lestari dan Ahmad Rifai, beserta keluarga besar yang selalu memberikan dukungan kepada penulis serta doa-doanya.
10. Untuk Muhammad Juan Pradana, yang selalu mendengarkan keluh kesah penulis, serta memberi dukungan hingga dapat menyelesaikan skripsi.
11. Untuk sahabatku Amalia Pratiwi, yang selalu memberikan semangat dan dukungan kepada penulis.
12. Untuk Novi, Wiranto, Elsa, Anis, Lina, Risma, Rizke, Bila, Puput yang selalu mendoakan, memberikan semangat, motivasi, pengertian, pencerahan, serta mendengarkan keluh kesah penulis.
13. Teman-teman Matematika 2019 serta Abang Yunda yang telah membantu serta memberikan semangat kepada penulis yang tidak dapat disebutkan satu persatu.
14. Orang-orang baik yang namanya tidak dapat saya sebutkan satu persatu yang telah menjadi bagian teman terbaik penulis yang selalu memberikan semangat dan menemani penulis dalam keadaan apapun serta telah memberikan pengalaman dan banyak cerita selama masa perkuliahan.

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

Bandar Lampung, 06 Juni 2023
Penulis,

Ameliana Wijayanti

DAFTAR ISI

halaman

DAFTAR TABEL vi

DAFTAR GAMBAR vii

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 4
2.3 Metode Kuadrat Terkecil (MKT) 5
2.4 Multikolinearitas 8
2.5 Ukuran Pemusatan dan Penskalaan (*Centering and Scaling*) 9
2.6 Regresi Ridge 10
2.7 *Least Absolute Shrinkage and Selection Operator* (LASSO) 13
2.8 *Elastic-Net* 15
2.9 Ketepatan Model 16
2.10 Validasi Silang 17
2.11 *Stunting* 18

III. METODOLOGI PENELITIAN 19

3.1 Waktu dan Tempat Penelitian 19
3.2 Data Penelitian 19
3.3 Metode Penelitian 20

IV. HASIL DAN PEMBAHASAN 22

4.1 Deteksi Multikolinearitas 22
4.2 Analisis Model Regresi dengan Metode Kuadrat Terkecil 23
4.3 Analisis dengan Regresi Ridge 24
4.4 Analisis dengan Regresi LASSO 27
4.5 Analisis dengan Regresi *Elastic-Net* 29

4.6 Perbandingan Koefisien Model MKT, Regresi Ridge, LASSO, dan *Elastic-Net* 31
4.7 Perbandingan Nilai Standard Error Parameter MKT, Regresi Ridge, LASSO, dan *Elastic-Net* 33
4.8 Pemilihan Model Terbaik 34

V. KESIMPULAN 37

DAFTAR PUSTAKA 38

LAMPIRAN 40

DAFTAR TABEL

Tabel Halaman

1. Variabel Data 19
2. Korelasi antarvariabel Independen... 22
3. Nilai VIF Variabel Independen.. 23
4. Estimasi Parameter dengan MKT 24
5. Perbandingan Koefisien Model MKT, Regresi Ridge, LASSO, dan *Elastic-Net*... 32
6. Nilai *Standard Error* Parameter MKT, Ridge, LASSO, dan *Elastic-Net*..... 33
7. Perbandingan Nilai MSE dan AIC Pada Regresi Ridge, LASSO, dan *Elastic-Net*... 34

DAFTAR GAMBAR

Gambar Halaman

1. Plot *Cross Validation* pada Regresi Ridge 25
2. Plot Koefisien Regresi Ridge. 25
3. Plot Nilai Koefisien Regresi Ridge. 26
4. Plot *Cross Validation* pada Regresi LASSO 27
5. Plot Koefisien Regresi LASSO 28
6. Plot Nilai Koefisien Regresi LASSO 28
7. Plot *Cross Validation* pada Regresi *Elastic-Net* 29
8. Plot Koefisien Regresi *Elastic-Net* 30
9. Plot Nilai Koefisien Regresi *Elastic-Net* 30
10. Nilai MSE MKT, Regresi Ridge, LASSO, dan *Elastic-Net* 35
11. Nilai AIC MKT, Regresi Ridge, LASSO, dan *Elastic-Net* 35

I.PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi adalah metode statistik yang diaplikasikan dalam mempelajari keterkaitan antara satu atau lebih variabel idependen dengan satu variabel dependen. Salah satu jenis dari analisis regresi yaitu regresi linear berganda. Adapun tujuan dari analisis regresi linear berganda yaitu untuk memahami keterkaitan antara satu variabel dependen dengan beberapa variabel independen secara simultan. Dalam penerapannya tidak jarang masalah spesifik muncul saat dilakukan analisis, salah satunya yaitu masalah multikolinearitas. Menurut Gujarati & Porter (2009), salah satu kondisi yang harus dipenuhi dalam analisis regresi linear berganda yaitu non-multikolinearitas.

Multikolinearitas terjadi apabila terdapat korelasi yang kuat antara dua atau lebih variabel independen dalam analisis regresi linear berganda. Variabel independen yang bersifat multikolinear memiliki hubungan linier yang kuat satu sama lain, yang dapat menyebabkan masalah dalam interpretasi dan estimasi parameter regresi sehingga penduga koefisien regresi yang diperoleh menjadi tidak efisien. Data yang terindikasi multikolinearitas dapat mengakibatkan ragam pada penduga kuadrat terkecil menjadi lebih besar dan dapat menurunkan ketepatan dari estimasi sehingga terjadi kesalahan dalam pengambilan keputusan. Menurut James, *et al*. (2013), untuk mengatasi masalah multikolinearitas, salah satu caranya yaitu dengan *shrinkage* (menyusutkan) parameter yang ditaksir. Teknik *shrinkage* sering disebut dengan istilah metode regularisasi. Metode regularisasi dapat menyusutkan koefisien mendekati angka nol yang relatif terhadap estimasi kuadrat terkecil. Metode regularisasi yang sering digunakan yaitu Regresi Ridge, *Least Absolute Shrinkage and Selection Operator* (LASSO), dan *Elastic-Net*.

Stunting merupakan kondisi dimana pertumbuhan terhambat pada anak-anak yang diakibatkan kekurangan gizi kronis, khususnya pada periode seribu hari pertama kehidupan (dari masa kehamilan sampai umur 2 tahun). Stunting dikarakterisasi oleh tinggi badan yang tidak wajar dengan usia anak, sehingga anak tampak lebih pendek dan lebih kecil dibandingkan dengan anak sebaya mereka. Berdasarkan hasil Survei Status Gizi Indonesia (SSGI), penurunan prevalensi *stunting* sebesar 24,4% pada tahun 2021 (Kemenkes, 2022). Namun nilai tersebut masih jauh dari target WHO yaitu kurang dari 20 %. Banyaknya faktor-faktor yang saling berkaitan yang mempengaruhi *stunting* pada balita memungkinkan antarvariabel bebasnya memiliki korelasi yang tinggi sangatlah besar, sehingga rawan terindikasi adanya masalah multikolinearitas.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Nabila (2018), mengenai penanganan masalah multikolinieritas dengan Regresi Ridge dibandingkan dengan Regresi Komponen Utama, diperoleh hasil terbaik menggunakan Regresi Ridge. Kemudian penelitian lainnya dalam menangani masalah multikolinearitas pernah dilakukan oleh Fadilah (2018), yaitu menangani masalah multikolinearitas dengan menggunakan Regresi LASSO dan penelitian oleh Nandari (2020), mengenai penanganan masalah multikolinieritas dengan Regresi *Elastic-Net*.

Berdasarkan uraian diatas, penulis tertarik untuk membandingkan Regresi Ridge, LASSO, dan *Elastic-Net* untuk mengatasi masalah multikolinearitas pada faktor-faktor yang berpengaruh terhadap kasus *stunting* pada balita di Indonesia dengan kriteria pembanding yang digunakan pada ketiga metode yaitu menggunakan nilai *Mean Squared Error* (MSE) dan *Akaike information criterion* (AIC). Oleh karena itu, penulis tertarik mengangkat penelitian ini dengan judul yaitu "*Penanganan Masalah Multikolinearitas dengan Regresi Ridge, LASSO, dan Elastic-Net pada Kasus Balita Stunting di Indonesia".*

1.2 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Mengetahui performa Regresi Ridge, LASSO, dan *Elastic-Net* dalam menangani masalah multikolinearitas.
2. Mendapatkan nilai duga koefisien regresi yang diperoleh menggunakan Regresi Ridge, LASSO*, dan Elastic-Net.*

1.3 Manfaat Penelitian

Manfaat penelitian ini diantaranya adalah menambah wawasan bagi penulis mengenai penerapan Regresi Ridge, LASSO, dan *Elastic-Net* dalam menangani masalah multikolinearitas. Serta sebagai bahan sumber acuan bagi pembaca yang ingin melakukan penelitian tentang penerapan Regresi Ridge, LASSO, dan *Elastic-Net* dalam menangani masalah multikolinearitas.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Menurut Draper & Smith (1998), Analisis regresi didefinisikan sebagai alat statistik dalam memahami hubungan antara variabel independen dan variabel dependen, serta memberikan penjelasan atau prediksi yang akurat mengenai variabel dependen berdasarkan variabel independen. Jenis analisis regresi yang melibatkan hubungan antara satu variabel independen dan satu variabel dependen disebut regresi linear sederhana. Sedangkan regresi linear berganda didefinisikan sebagai adanya hubungan linear antara variabel dependen dengan dua atau lebih variabel independen. Suatu model regresi dalam analisis regresi akan diestimasi dan digunakan untuk menggambarkan suatu pola atau bentuk hubungan antara variabel-variabel. Model regresi itu sendiri merupakan sebuah persamaan matematika yang memungkinkan kita untuk mengestimasi nilai dari variabel dependen berdasarkan nilai-nilai variabel independen.

2.2 Regresi Linear Berganda

Regresi linear berganda merupakan suatu teknik statistik yang diaplikasikan dalam memodelkan keterkaitan antara variabel dependen (Y) dengan beberapa variabel independen $(X_1, X_2, \dots, X_n)$. Tujuan dari analisis regresi linear berganda yaitu untuk menduga atau memperkirakan nilai Y atas X. Menurut Draper & Smith (1998), secara keseluruhan, model regresi linear berganda dapat dirumuskan dengan:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \tag{2.1}$$

dengan:

Y	= Variabel dependen
β_0	= Konstanta
$\beta_1, \beta_2, \dots, \beta_n$	= Koefisien regresi
$X_1, X_2, \dots, X_n$	= Variabel independen
ε	= Kesalahan penduga (*error*)

Dalam bentuk matriks, persamaan tersebut dapat dituliskan sebagai:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

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

dengan:

$\boldsymbol{Y}$	= Vektor variabel dependen berukuran n × 1
$\boldsymbol{X}$	= Matriks variabel independen berukuran n × k
$\boldsymbol{\beta}$	= Vektor parameter berukuran k × 1
$\boldsymbol{\varepsilon}$	= Vektor galat berukuran n ×1, dengan $\varepsilon \sim N(0, \sigma^2)$

2.3 Metode Kuadrat Terkecil (MKT)

Salah satu teknik yang bisa digunakan untuk memperkirakan parameter adalah *Ordinary Least Squares* (OLS) atau Metode Kuadrat Terkecil (MKT). Metode ini sangat populer dalam memperkirakan estimasi parameter regresi. Menurut Hastie, *et al*. (2008), pendekatan MKT digunakan untuk memperkirakan β dengan cara meminimalkan jumlah kuadrat galat (JKG).

Untuk memperoleh nilai β, langkahnya adalah dengan meminimalkan bentuk kuadrat:

$$Q(\hat{\beta}_j) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon_i^T \varepsilon_i = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\, \beta_j \right)^2 \tag{2.3}$$

Persamaan tersebut dalam bentuk matriks menjadi

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta} \tag{2.4}$$

Sehingga, perkalian matriks galat dapat ditulis sebagai:

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

Selanjutnya, $\boldsymbol{\varepsilon_i^T \varepsilon_i}$ akan diturunkan secara parsial terhadap β, kemudian ditetapkan sama dengan nol.

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

Sehingga mendapatkan estimator untuk MKT sebagai berikut:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-\boldsymbol{1}} \boldsymbol{X^T Y} \tag{2.5}$$

Adapun sifat-sifat MKT sebagai berikut:

1) $\hat{\beta}$ linear

   $\hat{\beta}$ linear apabila $\hat{\beta}$ merupakn 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) $\hat{\beta}$ tak bias

$\hat{\beta}$ adalah penduga tak bias apabila $E(\hat{\beta}) = \beta$

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

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

3) $\hat{\beta}$ memiliki variansi minimum

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

$Var(\hat{\beta}) = (X^T X)^{-1} \sigma^2$ merupakan varians paling kecil di antara semua estimasi linear tidak bias.

Apabila estimator MKT memenuhi sifat linear, tak bias dan mempunyai varians minimal maka estimator tersebut bersifat *Best Linear Unbiased Estimator* (BLUE).

2.4 Multikolinearitas

Istilah multikolinearitas awalnya diusulkan oleh Ragnar Frisch tahun 1934. Menurut Gujarati & Porter (2009), multikolinearitas atau kolinearitas ganda adalah adanya keterkaitan linier yang sangat erat antara beberapa atau semua variabel independen dalam model regresi. Multikolinearitas dapat menghasilkan koefisien regresi yang diperoleh dari analisis regresi berganda menjadi sangat lemah atau tidak memberikan hasil analisis yang mampu mewakili pengaruh variabel independen yang terlibat (Montgomery, *et al.*, 2012). Dampak lain dari adanya multikolinearitas antara lain yaitu :

1. Nilai standar *error* cenderung akan semakin besar bersamaan dengan tingginya tingkat korelasi antarvariabel independen.
2. Nilai selang kepercayaan cenderung untuk lebih besar akibat besarnya standar *error*, sehingga sangat sulit untuk menyangkal hipotesis nol dalam penelitian apabila terdapat multikolinearitas.
3. Dalam kasus multikolinearitas yang tinggi, mengakibatkan kemungkinan atau risiko gagal menolak hipotesis yang salah meningkat
4. Apabila multikolinearitas tinggi, memungkinkan diperoleh R^2 yang tinggi pula namun tidak dapat menjelaskan sifat maupun pengaruh dari variabel independen yang bersangkutan.

Menurut Montgomery, *et al*. (2012), terdapat beberapa metode dalam mengidentifikasi adanya masalah multikolinearitas, antara lain sebagai berikut:

a) Melihat korelasi antarvariabel independen

   Multikolinearitas dapat diduga dengan melihat tingginya korelasi antarvariabel independen. Apabila nilai dari koefisien korelasi antarvariabel independen diatas 0.5, diduga terjadi masalah multikolinearitas atau kolinearitas ganda antarvariabel independen.

b) Menggunakan *Variance Inflation Factor* (VIF)

VIF merupakan alternatif cara untuk mengidetifikasi adanya masalah multikolinearitas. Peningkatan variansi bergantung dari σ^2 dan VIF itu sendiri. Untuk mencari nilai VIF, rumus yang digunakan adalah sebagai berikut:

$$VIF_{(j)} = \frac{1}{1 - R_j^2} \tag{2.6}$$

R_j^2 merupakan koefisien determinasi ke-j ($j = 1,2, \dots, k$) yang diperoleh dari variabel independen X_j yang diestimasi dengan menggunakan variabel independen lainnya. Jika perolehan nilai VIF > 10, maka secara signifikan disimpulkan bahwa terjadi masalah multikolinearitas.

Menurut James, *et al*. (2013), untuk mengatasi masalah multikolinearitas, salah satu caranya yaitu dengan *shrinkage* (menyusutkan) parameter yang ditaksir. Teknik shrinkage sering disebut sebagai teknik regularisasi. Regularisasi memungkinkan pengurangan parameter mendekati nol relatif terhadap perkiraan kuadrat terkecil. Beberapa metode regularisasi yang umum digunakan adalah Regresi Ridge, LASSO, dan *Elastic-Net*.

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

Standarisasi merupakan bagian dari mengubah variabel menjadi bentuk yang seragam. Pembaruan sederhana dari normalisasi atau standarisasi variabel ini adalah transformasi korelasi. Pemusatan adalah selisih antara setiap pengamatan dan rata-rata dari semua pengamatan untuk variabel tersebut. Di sisi lain, penskalaan melibatkan menyajikan pengamatan dalam satuan standar deviasi dari pengamatan untuk variabel tersebut. (Kutner, *et al*., 2005). Dalam konteks ini, model regresi linear berganda akan distandarisasi.

Berikut merupakan *standardized* variabel dependen (Y) dan variabel independen ($X_1, X_2, \ldots, X_k$):

$$Y_i{}^* = \frac{Y_i - \bar{Y}}{S_Y} \quad dengan\ S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} \tag{2.7}$$

$$X_j{}^* = \frac{X_{ij} - \bar{X}_J}{S_{Xj}} \quad dengan\ S_{Xj} = \sqrt{\frac{\sum_{i=1}^{n}(X_{ij} - \bar{X}_J)^2}{n-1}} \tag{2.8}$$

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

dengan:

$\bar{Y}$ = *Mean* dari Y

$\bar{X}_J$ = *Mean* dari pengamatan X_j

S_Y = Standar deviasi dari Y

S_{Xj} = Standar deviasi dari X_j

Model regresi berganda yang telah diubah ke bentuk baku adalah hasil dari transformasi model regresi berganda (yang didefinisikan sebagai transformasi korelasi). Model regresi yang terstandarisasi (*standardized regression model*) dapat ditulis sebagai berikut:

$$Y_i{}^* = \beta_1{}^* X_{1i}{}^* + \beta_2{}^* X_{2i}{}^* + \cdots + \beta_k{}^* X_{ki}{}^* + \varepsilon_i{}^* \tag{2.9}$$

2.6 Regresi Ridge

Metode Regresi Ridge adalah suatu pendekatan atau teknik yang dikembangkan untuk mengatasi masalah multikolinearitas dan menjaga kestabilan koefisien regresi. Regresi Ridge pertama kali diajukan oleh A. E. Hoerl pada tahun 1962 untuk mengendalikan ketidakstabilan pada penduga Metode Kuadrat Terkecil (MKT). Menurut Herawati, *et al.* (2018), Regresi Ridge adalah sebuah variasi atau adaptasi dari Metode Kuadrat Terkecil (MKT) yang menghasilkan estimasi yang cenderung memiliki bias pada koefisien regresi. Regresi Ridge dapat mengurangi pengaruh akibat adanya masalah multikolinearitas dengan cara menentukan penduga bias yang memiliki varians yang lebih minimum dari varians penduga MKT. Meskipun metode ini cenderung memiliki bias pada koefisien regresi, penduga ini dapat menghampiri nilai koefisien regresi yang sebenarnya. Regresi Ridge serupa dengan MKT yaitu meminimumkan jumlah kuadrat galat atau *sum of squares error* (SSE) pada estimasi parameter regresi. Namun, Regresi Ridge memasukkan komponen penyusutan untuk mengurangi SSE.

Dalam proses penaksiran koefisien model regresi, estimator Regresi Ridge didapatkan melalui meminimalkan jumlah kesalahan kuadrat,

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{T}}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})$$

dengan syarat:

$$\boldsymbol{\beta_R}^{\boldsymbol{T}}\boldsymbol{\beta_R} = \boldsymbol{c}^2$$
$$\boldsymbol{\beta_R}^{\boldsymbol{T}}\boldsymbol{\beta_R} - \boldsymbol{c}^2 = \boldsymbol{0}$$

dengan menggunakan metode *lagrange multiplier*, diperoleh:

$$\boldsymbol{L} = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{T}}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) + \lambda\boldsymbol{I}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \boldsymbol{c}^{\boldsymbol{2}})$$
$$\boldsymbol{L} = \boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{Y} - \boldsymbol{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} + \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{2}} + \lambda\boldsymbol{I}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \boldsymbol{c}^{\boldsymbol{2}})$$
$$\boldsymbol{L} = \boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{Y} - \boldsymbol{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} + \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{2}} + \lambda\boldsymbol{I}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \lambda\boldsymbol{I}\boldsymbol{c}^{\boldsymbol{2}}$$

Dengan menggunakan persyaratan minimum dari persamaan diatas, penaksir koefisien Regresi Ridge adalah sebagai berikut:

$$\frac{\partial \boldsymbol{L}}{\partial \widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}} = -\boldsymbol{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y} + \boldsymbol{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} + 2\lambda\boldsymbol{I}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} = \boldsymbol{0}$$
$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} + \lambda\boldsymbol{I}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$(\boldsymbol{X^T X} + \lambda \boldsymbol{I})\boldsymbol{\widehat{\beta}_R} = \boldsymbol{X^T Y} \tag{2.10}$$
$$\boldsymbol{\widehat{\beta}_R} = (\boldsymbol{X^T X} + \lambda \boldsymbol{I})^{\boldsymbol{-1}}\boldsymbol{X^T Y}$$

Regresi Ridge menggunakan pendekatan dengan menambahkan konstanta bias λ pada diagonal matriks $\boldsymbol{X^T X}$, yang berdampak pada koefisien penduga Ridge yang dipengaruhi oleh besarnya bias λ. Dalam Regresi Ridge, variabel dependen (*Y*) dan variabel independen (*X*) disesuaikan menjadi bentuk standar menggunakan X^* dan Y^* berdasarkan rumus berikut:

$$Y_i{}^* = \frac{Y_i - \bar{Y}}{\sqrt{n-1}\, S_Y} \;\; dan\, X_j{}^* = \frac{X_{ij} - \bar{X_J}}{\sqrt{n-1}\, S_j} \tag{2.11}$$

Menurut Kutner, *et al.* (2005), persamaan umum model Regresi Ridge sebagai berikut:

$$Y_i{}^* = \beta_1{}^* X_1{}^* + \beta_2{}^* X_2{}^* + \cdots + \beta_p{}^* X_{ip}{}^* + \varepsilon^* \tag{2.12}$$

Persamaan dalam notasi matriks, ditulis dengan:

$$\boldsymbol{Y}^* = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^* \tag{2.13}$$

dengan:

$\boldsymbol{Y}^*$ = Vektor variabel dependen berukuran n × 1
$\boldsymbol{X}^*$ = Matriks variabel independen berukuran n × *k*
$\boldsymbol{\beta}^*$ = Vektor dari koefisien Regresi Ridge berukuran k × 1
$\boldsymbol{\varepsilon}^*$ = Vektor galat berukuran n × 1

Secara umum, formula yang digunakan untuk mengestimasi Regresi Ridge adalah sebagai berikut:

$$\hat{\beta}^{RIDGE} = \arg\min_{\beta} ||y - \boldsymbol{X}\beta||_2^2 + \lambda||\beta||_2^2 \tag{2.14}$$

atau dapat ditulis sebagai berikut:

$$\hat{\beta}^{RIDGE} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + (1-\alpha)\,\lambda\sum_{j=1}^{p}\beta_j^{\,2} \tag{2.15}$$

pada Regresi Ridge nilai α bernilai 0 sehingga,

$$\hat{\beta}^{RIDGE} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^{\,2} \tag{2.16}$$

dengan $\lambda \geq 0$ merupakan *tuning parameter* dan $\sum_{j=1}^{p}\beta_j^{\,2}$ merupakan penalti penyusutan. Penalti penyusutan tidak memberikan dampak apa pun ketika λ=0, sehingga Regresi Ridge akan menghasilkan estimasi koefisien regresi yang serupa dengan MKT. Ketika $\lambda \to \infty$, ini akan berdampak pada penalti pengurangan yang semakin besar dan estimasi koefisien mendekati nol. Dalam Regresi Ridge, λ optimal ditentukan dengan validasi silang.

2.7 ***Least Absolute Shrinkage and Selection Operator*** **(LASSO)**

Least Absolute Shrinkage and Selection Operator (LASSO) merupakan bagian dari metode penyusutan seperti Regresi Ridge yang dapat menanggulangi masalah multikolinearitas. LASSO dikenalkan oleh Tibshirani tahun 1996. Menurut Tibshirani (1996), LASSO mampu menyusutkan parameter regresi dari variabel independen yang memiliki korelasi tinggi menjadi nol atau mendekati nol.

Persamaan umum dalam metode LASSO dapat ditulis sebagai berikut:

$$\boldsymbol{Y}^{**} = \boldsymbol{X}^{**}\boldsymbol{\beta}^{**} + \boldsymbol{\varepsilon}^{**} \tag{2.17}$$

dengan:

$\boldsymbol{Y}^{**}$ = Vektor variabel dependen berukuran $n \times 1$

$\boldsymbol{X}^{**}$ = Matriks variabel independen berukuran $n \times k$

$\boldsymbol{\beta}^{**}$ = Vektor dari koefisien LASSO berukuran $k \times 1$

$\boldsymbol{\varepsilon}^{**}$ = Vektor galat berukuran $n \times 1$

Menurut Zhao & Yu (2006), Estimasi dalam metode LASSO diformulasikan sebagai:

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \|y - \boldsymbol{X}\beta\|_2^2 + \lambda\|\beta\|_1 \tag{2.18}$$

atau dapat dituliskan sebagai,

$$\hat{\beta}^{LASSO} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \alpha\,\lambda\sum_{j=1}^{p}\left|\beta_j\right| \tag{2.19}$$

pada Regresi LASSO nilai α bernilai 1 sehingga penduga Regresi LASSO adalah sebagai berikut:

$$\hat{\beta}^{LASSO} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \tag{2.20}$$

Formula LASSO dan Regresi Ridge hampir sama, hanya berbeda penaltinya saja. Pada Ridge, penalti penyusutannya $\sum_{j=1}^{p} {\beta_j}^2$, sedangkan LASSO diganti dengan $\sum_{j=1}^{p}\left|\beta_j\right|$. Sehingga, LASSO dapat mereduksi koefisien menjadi nol, menghasilkan model dengan variabel independen yang lebih sedikit. Penalti penyusutan pada LASSO mengakibatkan penyusutan nilai estimasi koefisien, mempertahankan variabel independen yang memiliki pengaruh yang signifikan pada model, dan menghilangkan variabel independen yang memiliki pengaruh yang lebih rendah untuk meningkatkan efisiensi model.

2.8 *Elastic-Net*

Elastic-net merupakan suatu metode atau teknik dalam analisis regresi yang menggabungkan elemen-elemen dari Regresi Ridge dan LASSO. Metode ini menggunakan kombinasi dari penalti L1 (LASSO) dan penalti L2 (Regresi Ridge) untuk melakukan seleksi variabel dan mengontrol multikolinearitas (Hastie, *et al*., 2008). Menurut Zou & Hastie (2005), metode *Elastic-Net* dapat menyusutkan koefisien regresi tepat menjadi nol, selain itu metode ini juga dapat melakukan seleksi variabel secara simultan dengan penalti *Elastic-Net* yang dinyatakan dengan:

$$\sum_{j=1}^{p}\left[\alpha\left|\beta_j\right| + (1-\alpha){\beta_j}^2\right] \tag{2.21}$$

Elastic-Net dapat digunakan untuk mengatasi masalah dari LASSO. Dimana Regresi LASSO memiliki kekurangan antara lain:

1. Ketika jumlah variabel (*p*) lebih besar daripada jumlah pengamatan (*n*) atau $p > n$, LASSO akan memilih hanya *n* variabel yang akan dimasukkan ke dalam model.
2. Jika terdapat sekelompok variabel dengan korelasi tinggi, maka LASSO cenderung memilih hanya satu variabel dari kelompok tersebut dan tidak peduli yang mana yang terpilih.
3. Ketika jumlah variabel (*p*) lebih kecil daripada jumlah pengamatan (*n*) atau $p < n$, kinerja LASSO lebih dipengaruhi oleh Regresi Ridge.

Elastic-net secara bersamaan melakukan pemilihan variabel otomatis dan kontinu, dan dapat memilih kelompok variabel berkorelasi.

Menurut Zou & Hastie (2005), untuk setiap λ_1 dan λ_2, ukuran standar dari *Elastic-Net* didefinisikan dengan:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| \qquad (2.22)$$

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

Sehingga penduga koefisien pada *Elastic-Net* dapat dituliskan sebagai:

$$\hat{\beta}^{net} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\left[(1-\alpha)\sum_{j=1}^{p} {\beta_j}^2 + \alpha \sum_{j=1}^{p} \left|\beta_j\right|\right] \qquad (2.23)$$

dengan α yang merupakan parameter gabungan Regresi Ridge $(\alpha = 1)$ dengan LASSO $(\alpha = 1)$. Ketika $\alpha = 0$, penalti penyusutan *Elastic-net* sama dengan penalti penyusutan Regresi Ridge. Apabila $\alpha = 1$, maka penalti penyusutan *Elastic-net* sama dengan penalti penyusutan LASSO.

2.9 Ketepatan Model

Pemilihan model terbaik dapat dilakukan dengan mengamati nilai *Mean Squared Error* (MSE). MSE adalah metode yang umum dan sederhana dalam mengukur tingkat kesalahan. MSE dihitung dengan mengkuadratkan selisih antara nilai prediksi dan nilai aktual. Semakin kecil nilai MSE, semakin akurat prediksi tersebut. Ghozali (2009), menyatakan bahwa MSE dapat dihitung dengan formula berikut:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2 \tag{2.24}$$

dengan:

Y_t = Nilai observasi data

$\hat{Y}_t$ = Nilai hasil prediksi

n = Jumlah data

Dalam penelitian ini, pemilihan model terbaik juga dilakukan dengan melihat nilai *Akaike information criterion* (AIC). Adapun rumus AIC sebagai berikut:

$$AIC_C = 2k + n \ln\left(\frac{JKG}{n}\right) \tag{2.25}$$

dimana $JKG = \sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2$, dan k merupakan jumlah parameter yang diestimasi oleh model. Dalam prakteknya, penentuan model terbaik dapat dilakukan dengan melihat nilai AIC yang paling rendah (Herawati, *et al.*, 2018).

2.10 Validasi Silang

Validasi silang (*cross validation*) adalah salah satu teknik yang digunakan dalam memperkirakan galat prediksi untuk meningkatkan keakuratan pemilihan model (James, *et al.*, 2013). Dalam CV, data dibagi menjadi dua yaitu data *training* dan data *testing*. Data *training* untuk membentuk model klasifikasi sedangkan data *testing* untuk mengukur ketepatan keberhasilan model. Menurut Efron & Tibshirani (1993), salah satu pendekatan dari *cross validation* yaitu dengan menerapkan *k-fold*. Dimana data dibagi kedalam *k-fold*. Dengan *fold* pertama sebagai data untuk validasi dan *k-1 fold* sisanya untuk untuk melatih model. Proses ini diulang hingga mencapai *fold* terakhir yang digunakan sebagai data validasi.. Sehingga formula *k-fold cross validation* dapat ditulis:

$$CV = \frac{1}{k}\sum_{i=1}^{k}\left(y_i - \hat{y}_{-i(k)}\right)^2 \qquad (2.26)$$

dimana $\hat{y}_{-i(k)}$ adalah dugaan y saat *fold* ke-*k* tidak digunakan dalam mengestimasi model. Dalam *k-fold cross validation* sering kali menggunakan $k =$ 5 atau $k = 10$. Menurut James, *et al*. (2013), keuntungan dalam menggunakan 5-*fold* atau 10-*fold cross validation* yaitu akan menghasilkan ragam rendah. Dengan *k-fold cross validation*, maka dapat diperoleh *tuning* parameter yang optimum dilihat dari nilai CV yang terkecil.

2.11 *Stunting*

Stunting merupakan gangguan pertumbuhan dan perkembangan pada anak balita akibat kekurangan gizi kronis dan infeksi berulang. Hal ini ditandai dengan tinggi badan yang berada di bawah standar. Berdasarkan data Survei SSGI tahun 2021, terlihat adanya penurunan prevalensi stunting dari 27,7% pada tahun 2019 menjadi 24,4% pada tahun 2021 (Kemenkes, 2022). Meskipun demikian, angka tersebut masih berada di bawah target yang ditetapkan oleh WHO, yaitu kurang dari 20%. *Stunting* merupakan masalah gizi yang menjadi masalah nasional karena berdampak negatif terhadap sumber daya manusia di masa mendatang. Faktor-faktor yang berhubungan langsung dengan pertumbuhan dan perkembangan anak usia dini, seperti misalnya. pemberian ASI eksklusif, program imunisasi dasar lengkap, berat badan lahir, faktor sosial ekonomi keluarga dan faktor kesehatan lingkungan. Upaya pemerintah sendiri untuk mencegah *stunting* dilaksanakan melalui program perbaikan gizi masyarakat yaitu Pemberian Makanan Tambahan (PMT) untuk meningkatkan status gizi anak, kebersihan lingkungan dengan peningkatan kualitas kebersihan lingkungan, jamban sehat perorangan, dan cuci tangan pakai sabun. Karena kebijakan menargetkan kepada warga miskin agar ada perubahan perilaku.

III.METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada Semester Ganjil tahun akademik 2022/2023 dengan melakukan penelitian secara studi pustaka di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang dipergunakan dalam penelitian ini adalah data sekunder yaitu data kasus *stunting* pada balita di Indonesia beserta faktor-faktor yang mempengaruhinya yang diperoleh dari *website* resmi Kementrian Kesehatan Republik Indonesia dengan data yang digunakan yaitu data tahun 2021. Dimana dalam penelitian ini jumlah data sebanyak 34 observasi dengan 13 variabel independen dan 1 variabel dependen yang akan diteliti.

Tabel 1. Variabel Data

Variabel	Keterangan
Y	Persentase Balita *Stunting* (Persen)
X_1	Persentase Balita Gizi Kurang (Persen)
X_2	Persentase Balita Mendapatkan Asi Eksklusif (Persen)
X_3	Balita Mendapatkan Imunisasi Dasar Lengkap (Jiwa)
X_4	Persentase Balita Dipantau Pertumbuhannya (Persen)

Tabel 1. (Lanjutan)

Variabel	Keterangan
X_5	Balita Mendapatkan Imunisasi DPT-HB-Hib3 (Jiwa)
X_6	Cakupan Pelayanan Kesehatan Ibu Hamil (Jiwa)
X_7	Jumlah Tenaga Gizi (Jiwa)
X_8	Jumlah Rumah Sakit
X_9	Persentase Rumah Tangga dengan Air Minum Layak (Persen)
X_{10}	Persentase Rumah Tangga dengan Sanitasi Layak (Persen)
X_{11}	Indeks Pembangunan Manusia (IPM)
X_{12}	Jumlah Penduduk Miskin (Ribu Orang)
X_{13}	Kepadatan Penduduk (Per km^2)

3.3 Metode Penelitian

Penelitian ini dilakukan secara studi pustaka dengan memahami buku-buku pendukung serta melalui karya ilmiah yang disusun dalam bentuk jurnal. Pada penelitian ini, penulis menggunakan *software* RStudio versi 4.2.1 dalam mempercepat perhitungan yang menyajikan hasil yang akurat.

Dalam penelitian ini dilakukan analisis menggunakan *package glm.net*. Langkah-langkah yang dilakukan yaitu:

1. Mendeteksi multikolinearitas dengan memperhatikan korelasi antarvariabel independen dan nilai *variance inflation factor* (VIF).
2. Melakukan analisis model regresi dengan MKT.
3. Melakukan analisis dengan Regresi Ridge.
   a. Melakukan *center and scale* data.
   b. Memilih nilai λ optimal dengan validasi silang.
   c. Mendapat nilai duga β Regresi Ridge berdasarkan nilai λ optimal.
   d. Analisis model regresi yang diperoleh menggunakan Regresi Ridge.
4. Melakukan analisis dengan metode LASSO.
   a. Melakukan center and *scale data.*
   b. Memilih nilai λ optimal dengan validasi silang.

c. Mendapat nilai duga β Regresi LASSO berdasarkan nilai λ optimal.

d. Analisis model regresi yang diperoleh menggunakan Regresi LASSO.

5. Melakukan analisis dengan metode *Elastic-Net.*

    a. Melakukan *center and scale* data.

    b. Memilih nilai λ optimal dengan validasi silang.

    c. Mendapat nilai duga β metode *Elastic-Net* berdasarkan nilai λ optimal.

    d. Analisis model regresi yang diperoleh menggunakan metode *Elastic-Net.*

6. Membandingkan nilai koefisien dari MKT, Regresi Ridge, LASSO dan *Elastic-Net.*

7. Menghitung nilai MSE dan AIC dari model yang diperoleh dengan MKT, Regresi Ridge, LASSO dan *Elastic-Net.*

8. Perbandingan Nilai *Standard Error* Parameter MKT, Regresi Ridge, LASSO, dan *Elastic-Net.*

9. Pemilihan model terbaik dengan membandingkan nilai MSE dan AIC dari model yang diperoleh dengan MKT, Regresi Ridge, LASSO dan *Elastic-Net.*

V.KESIMPULAN

Berdasarkan hasil penelitian dalam menangani masalah multikolinearitas dengan menggunakan Regresi Ridge, LASSO dan *Elastic-Net* pada kasus balita *stunting* di Indonesia pada tahun 2021, dapat disimpulkan bahwa berdasarkan nilai MSE dan AIC, model regresi yang diperoleh dengan Regresi *Elastic-Net* lebih baik dalam menduga variabel dependen daripada model Ridge dan LASSO. Sehingga pada analisis menggunakan data kasus balita *stunting* di Indonesia pada tahun 2021, didapat model terbaik dengan Regresi *Elastic-Net* yaitu:

$$\begin{aligned}\hat{Y} \quad = \quad & -0{,}34 \times 10^{-15} + 24{,}201480\ X_1 + 4{,}142324\ X_2 + 4{,}091625\ X_4 \\ & 0{,}811826\ X_6 - 3{,}120010\ X_7 + 2{,}349594\ X_9 + 2{,}196391\ X_{10} \\ & -7{,}270900\ X_{11} - 1{,}158410\ X_{13}\end{aligned}$$

Berdasarkan model yang didapat, kasus *stunting* pada balita dipengaruhi oleh persentase balita gizi kurang (X_1), persentase balita mendapatkan asi eksklusif (X_2), persentase balita dipantau pertumbuhannya (X_4),cakupan pelayanan kesehatan ibu hamil (X_6), jumlah tenaga gizi (X_7), persentase rumah tangga dengan air minum layak (X_9), persentase rumah tangga dengan sanitasi layak (X_{10}), indeks pembangunan manusia (X_{11}), dan kepadatan penduduk (X_{13}).

DAFTAR PUSTAKA

Draper, N. R. & Smith, H. 1998. *Applied Regression Analysis.* Third Edition. John Wiley & Sons, New York.

Efron, B. & Tibshirani, R. 1993. *An Introduction to the Bootstrap*. 1st ed. Chapman & Hall, New York.

Fadilah, P.M. 2018. Analisis Regresi dengan Least Absolute Shrinkage and Selection Operator (LASSO) dalam Menangani Multikolinearitas (Skripsi). Jurusan Matematika FMIPA Unila, Bandar Lampung.

Ghozali, Imam. 2009. *Aplikasi Analisis Multivariate Dengan Program SPSS*. Edisi Keempat. Universitas Diponegoro, Semarang.

Gujarati, D. N. & Porter, D. C. 2009. *Basic Econometrics*. McGraw-Hill Irwin, Boston.

Hastie, T., Tibshirani, R, & Friedman, J. 2008. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, New York.

Herawati, N., Nisa, K., Azis, D., & Nabila, S.U. 2018. Ridge Regression For Handling Different Levels Of Multicollinearity. *Sci.Int.(Lahore)*. **30**(4): 597-600.

Herawati, N., Nisa, K., Setiawan, E., Nusyirwan, & Tiryono. 2018. Regularized Multiple Regression Methods to Deal with Severe Multicollinearity. *International Journal of Statistics and Applications*. **8**(4): 167-172.

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. *Profil Kesehatan Indonesia Tahun 2021*. Kemenkes RI, Jakarta.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. 2005. *Applied linear Statistical Models.* Five Edition. McGraw-Hill Irwin, Boston.

Montgomery, D. C., Peck, E. A., & Vining, G.G. 2012. *Introduction to Linear Regression Analysis.* Fourth Edition. John Willey and Sons, New York.

Nabila, S.U. 2018. Perbandingan Metode Regresi Ridge dan Metode Regresi Komponen Utama Dalam Menangani Multikolinearitas (Skripsi). Jurusan Matematika FMIPA Unila, Bandar Lampung.

Nandari, G. 2020. Analisis Regresi Berganda dengan Metode Elastic-Net Dalam Mengatasi Multikolinearitas (Skripsi). Jurusan Matematika FMIPA Unila, Bandar Lampung.

Tibshirani, R. 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. **58**(1): 267-288.

Zhao, P. & Yu, B. 2006. On Model Selection Consistency of Lasso. *Journal of Machine Learning Research.* 7: 2541-2562.

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.