

**PEMODELAN REGRESI *ELASTIC NET* PADA DATA ANGKA  
KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN**

**SKRIPSI**



**ARIEF RAHMAN NUR**

**H051191018**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2023**

**PEMODELAN REGRESI *ELASTIC NET* PADA DATA ANGKA  
KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN**

**SKRIPSI**

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains  
Pada Program Studi Statistika Departemen Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin**

**ARIEF RAHMAN NUR**

**H051191018**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**JULI 2023**

**LEMBAR PERNYATAAN KEOTENTIKAN**

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

**Pemodelan Regresi *Elastic Net* Pada Data Angka Kematian Bayi Di Provinsi Sulawesi Selatan**

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 18 Juli 2023




**ARIEF RAHMAN NUR**  
**H051191018**

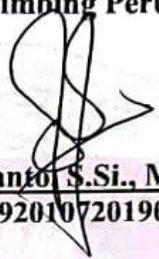
**PEMODELAN REGRESI *ELASTIC NET* PADA DATA ANGKA  
KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN**

**Disetujui Oleh**

**Pembimbing Utama**

  
**Andi Kresna Jaya, S.Si., M.Si.**  
NIP. 197312282000031001

**Pembimbing Pertama**

  
**Siswanto, S.Si., M.Si.**  
NIP. 199201072019031012

**Ketua Program Studi**

  
  
**Dr. Anna Islamiyati, S.Si., M.Si.**  
NIP. 197708082005012002

**Pada 18 Juli 2023**

**HALAMAN PENGESAHAN**

Skripsi ini diajukan oleh:

Nama : ARIEF RAHMAN NUR  
NIM : H051191018  
Program Studi : Statistika  
Judul Skripsi : Pemodelan Regresi *Elastic Net* Pada Data Angka  
Kematian Bayi Di Provinsi Sulawesi Selatan

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

**DEWAN PENGUJI**

1. Ketua : Andi Kresna Jaya, S.Si., M.Si. (.....)
2. Sekretaris : Siswanto, S.Si., M.Si. (.....)
3. Anggota : Dr. Nirwan, M.Si. (.....)
4. Anggota : Dr. Nurtiti Sunusi, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 18 Juli 2023

## KATA PENGANTAR

Segala puji senantiasa penulis panjatkan kepada Allah *subhanahu Wa Ta'ala* yang telah memberikan kekuatan dan rahmat-Nya sehingga penyusunan skripsi ini dapat terselesaikan dengan baik. Skripsi dengan judul “Pemodelan Regresi *Elastic Net* Pada Data Angka Kematian Bayi Di Provinsi Sulawesi Selatan” ini merupakan salah satu rangkaian syarat akademik yang harus dipenuhi untuk memperoleh Gelar Sarjana Sains pada Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.

Penulis dengan sungguh-sungguh menyadari bahwa penyelesaian skripsi ini tidak akan terwujud tanpa bantuan dan dorongan yang konsisten dari berbagai pihak yang telah memberikan dukungan moral dan materiil. Meskipun penulis memiliki keterbatasan dalam kemampuan dan pengetahuan, namun berkat bantuan dan dukungan yang diberikan, penulis berhasil menyelesaikan skripsi ini dengan baik. Oleh karena itu, penulis ingin mengungkapkan rasa terima kasih yang tak terhingga dan penghargaan yang tulus kepada semua pihak yang terlibat. Dengan penuh kesadaran dan rasa rendah hati, penulis ingin menggunakan kesempatan ini untuk mengungkapkan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada yang terhormat:

1. Terima kasih yang tak terhingga kepada ibunda **Emma** dan ayahanda **Sukri** tercinta atas dukungan, pengorbanan luar biasa, limpahan cinta dan kasih sayang, serta dengan ikhlas telah menemani setiap pilihan penulis dengan doa dan restu mulianya.
2. Terima kasih kepada **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh staf jajarannya.
3. Terima kasih kepada Bapak **Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh staf jajarannya.
4. Terima kasih kepada Ibu **Dr. Anna Islamiyati, S.Si., M.Si.**, sebagai Ketua Departemen Statistika atas dorongan semangat dan motivasi kepada penulis selama menjadi mahasiswa di Departemen Statistika.

5. Terima kasih kepada Pak **Andi Kresna Jaya, S.Si., M.Si.**, selaku Pembimbing Utama saya dengan penuh kesabaran telah memberikan arahan, dorongan semangat dan motivasi kepada penulis selama dalam menyelesaikan skripsi.
6. Terima kasih kepada pak **Siswanto, S.Si., M.Si.**, sebagai pembimbing pertama dan penasehat akademik penulis yang dengan penuh kesabaran membimbing dan memberikan dorongan semangat kepada penulis selama menjadi mahasiswa di Departemen Statistika.
7. Terima kasih kepada pak **Dr. Nirwan, M.Si.**, sebagai dosen penguji yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
8. Terima Kasih kepada Ibu **Dr. Nurtiti Sunusi S.Si., M.Si.**, sebagai dosen penguji yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
9. Terima kasih kepada segenap jajaran **Dosen Pengajar dan Staf Departemen Statistika** yang telah banyak membantu, memberikan ilmu-ilmunya, serta berbagai kemudahan lainnya yang diberikan selama menempuh pendidikan sarjana di Departemen Statistika.
10. Terima kasih kepada **Waode Sitti Amni** yang selalu memberikan semangat dan menemani perjalanan suka dan duka penulis selama ini dalam menyelesaikan proses perkuliahan di Universitas Hasanuddin.
11. Terima kasih yang kepada Kakak **Achmad Ik Tiara** yang telah senantiasa memberikan dukungan, motivasi dan telah membiayai selama perkuliahan penulis. Serta terimakasih juga kepada adik-adik tercinta penulis **Nurhakima, Nuralya dan Al Aqsa**.
12. Terima kasih kepada warga istana yang setia menemani penulis dari masa-masa sulit sampai akhir perkuliahan. Terima kasih kanda-kanda: **Agus Hermawan, Taufik Hidayat, Fadilah Amirul Adhel, Muhammad Syamsul Bahri, Muhammad Ferdiansyah, Rovino dan Sapriadi Rasyid.**

13. Terima kasih kepada **Muhammad Fathurahman** selaku sahabat terbaik penulis selama menjadi mahasiswa di Departemen Statistika.. Terima kasih, Ture.
14. Terima kasih kepada **Refa Joyce Semida, Mulliana, Nursyahfika, Nurazizah , dan Yasmin Pratiwi** selaku sahabat yang telah memberikan masukan dan bantuan dalam menyelesaikan perkuliahan.
15. Terima kasih kepada Teman **KKN Betao Riase 108** yang telah memberikan semangat dan motivasi dalam menyelesaikan skripsi ini.
16. Terima kasih kepada teman-teman sejawat di **BEM FMIPA Unhas dan Himastat** yang selalu memberikan pengalaman berharga dalam berorganisasi.
17. Terima kasih kepada teman-teman **Komunitas ANBISI P2JUH** yang telah menjadi teman bagi penulis di saat berkuliah di unhas sejak semester 2 hingga akhir.
18. Terima kasih kepada semua teman-teman di **angkatan Statistika 2019** yang telah menerima penulis yang merupakan orang pendatang di Makassar.
19. Terima kasih yang setinggi-tingginya kepada **seluruh pihak** yang mungkin tidak sempat penulis sebutkan satu persatu. Terima kasih atas segala dukungan, partisipasi, dan apresiasinya yang diberikan kepada penulis.

Penulis juga menyadari bahwa skripsi ini masih jauh dari kata sempurna, namun ini hasil terbaik yang dapat diberikan oleh penulis pada penelitian ini. Oleh karena dengan segala kerendahan hati penulis mengucapkan permohonan maaf yang sebesar-besarnya. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 18 Juni 2023



**ARIEF RAHMAN NUR**

**NIM. H051191018**



**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR**  
**UNTUK KEPENTINGAN AKADEMIK**

---

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : ARIEF RAHMAN NUR

NIM : H051191018

Program Studi : Statistika

Departemen : Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty-Free Right)** atas tugas akhir saya yang berjudul:

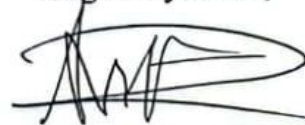
“Pemodelan Regresi *Elastic Net* Pada Data Angka Kematian Bayi Di Provinsi Sulawesi Selatan”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, 18 Juni 2023

Yang Menyatakan,



**ARIEF RAHMAN NUR**

**NIM. H051191018**

**ABSTRAK**

Angka kematian bayi merupakan indikator penting yang mencerminkan tingkat kesehatan dan kualitas pelayanan kesehatan bagi bayi di suatu wilayah. Dalam rangka menurunkan angka kematian bayi, analisis regresi dapat digunakan sebagai solusi untuk menentukan faktor-faktor yang berpengaruh terhadap angka kematian bayi. Masalah yang sering muncul pada analisis regresi berganda adalah multikolinieritas, yaitu adanya korelasi tinggi antara dua atau lebih variabel prediktor. Hal ini dapat menyebabkan masalah interpretasi yang sulit dan mengurangi keandalan hasil estimasi koefisien regresi. Salah satu metode yang efektif untuk menangani multikolinieritas adalah regresi *elastic net*. *Elastic net* adalah metode regresi yang menggabungkan regularisasi L1 (Lasso) dan L2 (*Ridge*), yang memungkinkan seleksi variabel yang efisien dan mengurangi pengaruh variabel yang berkorelasi tinggi. Pada penelitian ini, angka kematian bayi di Provinsi Sulawesi Selatan akan dimodelkan dengan menggunakan metode regresi *elastic net* berdasarkan variabel jumlah bayi yang diberi ASI eksklusif, jumlah pelayanan kesehatan bayi, jumlah pemberian vitamin A ke bayi, jumlah bayi berat badan lahir rendah, dan Jumlah ibu bersalin ditolong medis. Salah satu ukuran untuk melihat kebaikan model regresi adalah  $R^2$ . model regresi *elastic net* yang telah diuji memiliki nilai koefisien determinasi ( $R^2$ ) sebesar 39.26%.

Kata kunci: Multikolinearitas, Regresi *Elastic net*, angka kematian Bayi,  $R$ -squared, Optimasi *Coordinate Descent*.

**ABSTRACT**

*Infant mortality rate is a critical indicator reflecting the health level and quality of healthcare services for infants in a region. In order to reduce the infant mortality rate, regression analysis can be utilized as a solution to identify influential factors contributing to infant mortality. One common issue in multiple regression analysis is multicollinearity, which refers to high correlation between two or more predictor variables. This can lead to difficulties in interpretation and reduce the reliability of regression coefficient estimates. An effective method to address multicollinearity is the elastic net regression. The elastic net is a regression technique that combines L1 (Lasso) and L2 (Ridge) regularization, allowing for efficient variable selection and reducing the influence of highly correlated variables. In this study, the infant mortality rate in South Sulawesi Province will be modeled using the elastic net regression method based on variables such as the number of infants receiving exclusive breastfeeding, the number of healthcare services for infants, the number of vitamin A administrations to infants, the number of low birth weight infants, and the number of births assisted by medical personnel.  $R^2$  will be used as one of the metrics to assess the goodness of fit of the regression model. The tested elastic net regression model exhibits a coefficient of determination ( $R^2$ ) of 39.26%.*

**Keywords:** *Multicollinearity, Elastic Net regression, infant mortality rate, R-squared, Coordinate Descent optimization.*

## DAFTAR ISI

<b>HALAMAN SAMPUL</b> .....	<b>i</b>
<b>HALAMAN JUDUL</b> .....	<b>ii</b>
<b>LEMBAR PERNYATAAN KEOTENTIKAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>v</b>
<b>KATA PENGANTAR</b> .....	<b>vi</b>
<b>ABSTRAK</b> .....	<b>x</b>
<b>ABSTRACT</b> .....	<b>xi</b>
<b>DAFTAR ISI</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan penelitian .....	4
1.5 Manfaat Penelitian .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>6</b>
2.1 Analisis Regresi .....	6
2.2 Uji Multikolinearitas.....	7
2.3 Regresi <i>Ridge</i> .....	8
2.4 Regresi <i>Least Absolute Shrinkage And Selection Operator</i> .....	9
2.5 Regresi <i>Elastic Net</i> .....	10
2.6 Evaluasi Performa Model .....	11
2.7 <i>Coordinate Descent</i> .....	11
2.8 <i>Cross Validation</i> .....	12
2.9 Angka Kematian Bayi.....	13
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>14</b>
3.1 Data.....	14
3.2 Metode analisis .....	15
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>17</b>
4.1 Estimasi Parameter Pada Model <i>Elastic Net</i> Menggunakan Optimasi <i>Coordinate Descent</i> .....	17

4.2	Penerapan Metode <i>Elastic Net</i> Pada Data Yang Mengandung Multikolinieritas.....	19
4.2.1	Analisis Deskripsi.....	19
4.2.2	Standarisasi Data .....	22
4.2.3	Identifikasi Multikolinieritas .....	23
4.2.4	Pemilihan Nilai $\alpha$ dan $\lambda$ Yang Optimum .....	24
4.2.5	Pemodelan Data dengan Menggunakan Metode <i>Elastic Net</i> .....	24
4.2.6	Ukuran Kebaikan Model .....	25
4.2.7	Interpretasi Model <i>Elastic Net</i> .....	26
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>27</b>
5.1	Kesimpulan .....	27
5.2	Saran .....	27
<b>DAFTAR PUSTAKA.....</b>		<b>29</b>
<b>LAMPIRAN.....</b>		<b>32</b>

**DAFTAR TABEL**

<b>Tabel 3.1</b> Variabel Respon dan Prediktor .....	14
<b>Tabel 4.1</b> Analisis Deskripsi .....	20
<b>Tabel 4.2</b> Standarisasi Data .....	22
<b>Tabel 4.3</b> Uji Multikolinearitas .....	23
<b>Tabel 4.4</b> Hasil <i>Cross-validation</i> .....	24
<b>Tabel 4.5</b> Regresi <i>Elastic Net</i> .....	25

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Angka Kematian Bayi (AKB) atau *Infant Mortality Rate* (IMR) adalah mengukur jumlah bayi yang meninggal sebelum mencapai usia satu tahun per 1000 kelahiran hidup pada tahun yang sama. Hal ini mencerminkan tingkat kesehatan masyarakat. Selain itu, AKB merupakan salah satu target dari *Sustainable Development Goals* (SDGs) yang menunjukkan kesejahteraan bagi semua orang di segala usia (Puspita dkk., 2022). Semakin rendah AKB, semakin baik tingkat kesehatan masyarakat di wilayah tersebut. Indikator ini sangat erat kaitannya dengan target kelangsungan hidup anak karena mencerminkan kondisi sosial, ekonomi, dan lingkungan tempat tinggal anak-anak, termasuk aspek pemeliharaan kesehatannya. Secara umum, AKB lebih menggambarkan kondisi kesehatan reproduksi (Nurhafni dkk., 2021). Dalam rangka menurunkan AKB, analisis regresi dapat digunakan sebagai solusi untuk menentukan faktor-faktor yang berpengaruh terhadap AKB.

Analisis regresi merupakan metode statistik yang menjelaskan hubungan antara variabel respon dan variabel prediktor. Ada dua jenis regresi, yaitu regresi linier sederhana jika hanya terdapat satu variabel prediktor, dan regresi linier berganda jika terdapat lebih dari satu variabel prediktor. Menentukan model analisis regresi, diperlukan estimasi parameter yang akan menghasilkan nilai koefisien regresi. Metode *Ordinary Least Square* (OLS) atau Metode Kuadrat Terkecil (MKT) dapat digunakan untuk penaksiran koefisien regresi dengan meminimalkan jumlah kuadrat galatnya (*error*) (Mahalani dan Rifai, 2022). Menghasilkan estimasi parameter yang tidak bias pada regresi, terdapat beberapa asumsi yang harus dipenuhi. Salah satu asumsi penting adalah tidak terdapatnya multikolinearitas antar variabel prediktor. Jika asumsi ini tidak terpenuhi, maka estimasi parameter akan menjadi tidak akurat (Putri dan Sari, 2021). Dalam analisis regresi berganda, multikolinearitas sering terjadi ketika terdapat hubungan signifikan antara dua atau lebih variabel prediktor. Hal ini menyebabkan adanya penduga kuadrat terkecil yang memiliki variansi yang besar. Multikolinearitas dapat mempengaruhi keakuratan prediksi model dan menyebabkan kesalahan dalam pengambilan keputusan (Pardede dkk., 2022). Beberapa metode yang

sering digunakan untuk mengatasi masalah multikolinieritas antara lain metode regresi *ridge*, dan *Least Absolute Shrinkage and Selection Operator* (LASSO).

Regresi *ridge* dan LASSO merupakan dua jenis regresi regularisasi yang sering digunakan untuk mengatasi multikolinearitas. Regresi regularisasi (*regularization method*) adalah metode penyusutan estimasi atau prediksi pada regresi yang digunakan ketika terjadi pelanggaran asumsi multikolinearitas. Metode ini memberikan batasan atau penalisasi pada koefisien regresi. Regresi LASSO melakukan pembatasan koefisien variabel menjadi nol, sehingga dapat memilih variabel yang penting dan mencegah *overfitting* (Racharla dkk., 2020). Meskipun demikian, pemilihan variabel yang salah dan masalah multikolinearitas menjadi kelemahan dari regresi LASSO. Sementara itu, regresi *ridge* membatasi ukuran koefisien semua variabel, sehingga mampu mengatasi masalah multikolinearitas. Namun, regresi *ridge* kurang efektif dalam membatasi *overfitting* (Tsigler dan Bartlett, 2020). Oleh karena itu, untuk menutupi kelemahan masing-masing antara LASSO dan *ridge*, penggabungan kedua teknik tersebut dapat dilakukan dengan menggunakan teknik yang dikenal dengan regresi *elastic net*.

Regresi *elastic net* adalah salah satu metode yang menjadi alternatif dalam membatasi ukuran koefisien dan memilih variabel yang penting dalam model regresi linier. Metode ini memadukan kelebihan dari regresi LASSO dan *ridge* dengan memadukan *loss function* dari kedua metode tersebut. Regresi *elastic net* memberikan fleksibilitas yang lebih besar dalam mengontrol parameter alpha yang mempengaruhi kombinasi penalti *ridge* dan LASSO (De Leone dkk., 2020). Dalam regresi *elastic net*, ketika alpha mendekati 0, model akan cenderung ke arah regresi LASSO dan ketika alpha mendekati tak hingga, model akan cenderung ke arah regresi *ridge*. Regresi *elastic net* memiliki beberapa keunggulan dibandingkan regresi LASSO dan *ridge*, salah satunya adalah kemampuannya dalam mengatasi masalah *underfitting* dan *overfitting* secara bersamaan. Selain itu, regresi *elastic net* memiliki kontrol yang lebih baik dalam memilih variabel yang penting. Regresi *elastic net* juga dapat mengatasi masalah multikolinieritas dan memperbaiki akurasi model regresi linier. Oleh karena itu,



dalam memilih metode penyusutan, regresi *elastic net* bisa menjadi pilihan yang lebih baik dibandingkan regresi LASSO atau *ridge* (Yanke dkk., 2022).

Altalbany (2021) telah melakukan penelitian tentang perbandingan antara metode regresi *ridge*, *Least Absolute Shrinkage and Selection Operator* (LASSO), dan regresi *elastic Net* dengan menggunakan data simulasi dengan ukuran sampel 200, 1.000, 10.000, 50.000, dan 100.000. Hasil penelitian menunjukkan bahwa metode *elastic net* lebih unggul daripada metode *Ridge* dan LASSO dalam memperkirakan koefisien regresi pada tingkat multikolinearitas yang rendah, sedang, dan tinggi untuk semua ukuran sampel. Namun, ketika data mengandung multikolinearitas yang parah pada ukuran sampel kurang dari 10.000 observasi, metode LASSO lebih baik dibandingkan dengan metode Elastic Net. Selain itu, *elastic net* juga diaplikasikan pada angka kemiskinan dan kerentanan kemiskinan di provinsi Jawa Tengah dengan membandingkan metode *ridge*, LASSO, dan *elastic net*. Hasilnya, *elastic net* memiliki nilai *Mean Squared Error* (MSE) yang paling kecil dibandingkan dengan metode *ridge*, dan LASSO (Kusuma dan Wulansari, 2019). Penelitian ini memfokuskan penerapan *elastic Net* pada data simulasi dan bidang ekonomi dan sosial. Namun, dalam penelitian ini, akan mengaplikasikan metode tersebut pada data kesehatan, khususnya tentang AKB.

Angka kematian bayi yang tinggi di Indonesia mendorong perlunya membahas peran faktor-faktor kunci dalam kesehatan bayi dan dampaknya. Dalam konteks ini, terdapat beberapa indikator kesehatan bayi yang perlu diperhatikan. Jumlah bayi yang diberi ASI eksklusif, jumlah pelayanan kesehatan bayi, dan jumlah pemberian vitamin A ke bayi memberikan gambaran tentang ketersediaan perawatan kesehatan dan gizi bayi, yang merupakan faktor penting dalam kesehatan bayi. Selain itu, data mengenai jumlah bayi dengan berat badan lahir rendah (BBLR) dan jumlah ibu yang melahirkan dengan bantuan medis juga sangat penting karena dapat memberikan gambaran tentang tingkat kesehatan ibu selama kehamilan dan persalinan, yang berpotensi mempengaruhi kesehatan bayi (Kemenkes, 2022). Data-data ini menjadi sangat penting dalam menganalisis angka kematian bayi di Indonesia dan menemukan cara untuk menguranginya.

Namun, adanya faktor-faktor yang saling berkaitan dan berpengaruh terhadap angka kematian bayi, seperti jumlah pelayanan kesehatan bayi dan

jumlah ibu bersalin ditolong medis, dapat menyebabkan adanya korelasi antar variabel tersebut. Oleh karena itu, data pada angka kematian bayi di Sulawesi Selatan memiliki potensi untuk terjadi multikolinieritas antar variabel yang mempengaruhi angka kematian bayi. Untuk mengatasi masalah multikolinieritas antar variabel dalam data angka kematian bayi di Sulawesi Selatan, diperlukan metode analisis data yang tepat. Salah satu metode yang cocok untuk mengatasi masalah ini adalah metode *elastic net*, yang memadukan antara Lasso dan *ridge regression*. Dengan menggunakan metode *elastic net*, kita dapat menemukan variabel yang paling berpengaruh terhadap angka kematian bayi dengan menghilangkan variabel yang tidak signifikan dan mengurangi efek dari variabel yang saling berkaitan. Dengan demikian, metode ini dapat memberikan hasil yang lebih akurat dan efektif dalam menganalisis angka kematian bayi di Sulawesi Selatan.

### **1.2 Rumusan Masalah**

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi parameter model regresi *elastic net* dengan optimasi *coordinate descent* pada kasus angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019?
2. Bagaimana model pada data angka kematian bayi di Provinsi Sulawesi Selatan melalui regresi *elastic net* dengan optimasi *coordinate descent*?

### **1.3 Batasan Masalah**

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019.
2. Penentuan model regresi terbaik berdasarkan nilai *Root Mean Square Error* (RMSE) dan  $R^2$ .

### **1.4 Tujuan penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan estimasi parameter model regresi *elastic net* dengan optimasi *coordinate descent* pada kasus angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019.

2. Mendapatkan model pada data angka kematian bayi di Provinsi Sulawesi Selatan melalui regresi *elastic net* dengan optimasi *coordinate descent*.

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Memberikan informasi tentang pemodelan regresi *elastic net* dengan menggunakan metode *cross validation* pada kasus angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019.
2. Memberikan informasi tentang analisis regresi *elastic net* sebagai salah satu metode regresi regularisasi untuk menganalisis data yang mengandung Multikolinearitas.
3. Sebagai sumber informasi bagi Dinas Kesehatan Provinsi Sulawesi selatan terkait faktor-faktor yang mempengaruhi jumlah kasus angka kematian bayi di Provinsi Sulawesi Selatan sehingga dapat mengambil langkah antisipatif untuk menekan angka kematian bayi.

## BAB II TINJAUAN PUSTAKA

### 2.1 Analisis Regresi

Regresi linier adalah metode statistik yang digunakan untuk menentukan hubungan antara satu atau lebih variabel X (variabel prediktor) dengan sebuah variabel Y (variabel respon) melalui pembentukan model. Analisis regresi merupakan salah satu metode analisis hipotesis yang digunakan untuk mengetahui pengaruh antara variabel satu dengan yang lain dan dapat dinyatakan dalam persamaan regresi (Ningsih dan Dukalang, 2019). Dalam perkembangannya, regresi linier dibagi menjadi dua jenis yaitu regresi linier sederhana dan regresi linier berganda. Regresi linier sederhana digunakan untuk menggambarkan hubungan antara satu variabel respon dengan satu variabel terikat (Ningsih dan Dukalang, 2019). Model regresi linier berganda adalah pengembangan dari model regresi linier sederhana, dengan jumlah variabel responnya lebih dari satu dan hanya ada satu variabel terikat. Persamaan regresi linier berganda yang mencakup dua atau lebih variabel respon memiliki bentuk umum yang berbeda dibandingkan dengan model regresi linier sederhana yang hanya memiliki satu variabel respon. bentuk umum dari persamaan regresi linier berganda sebagai berikut:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon \quad (2.1)$$

Apabila menggunakan data sampel, maka dapat menggunakan model estimasi dari Persamaan (2.1) adalah sebagai berikut:

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

Berdasarkan Persamaan (2.2) maka bentuk matriks dari model regresi linier berganda dapat dituliskan sebagai berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{2n} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan:

$\mathbf{Y}$  : Vektor pengamatan berukuran  $n \times 1$

$\mathbf{X}$  : Matriks variabel respon ukuran  $n \times (p + 1)$

$\boldsymbol{\beta}$  : Vektor parameter yang akan ditaksir berukuran  $(p + 1) \times 1$

$\varepsilon$  : Vektor random error berukuran  $n \times 1$

Menurut Sulistianingsih dkk (2022), estimasi parameter dapat diperoleh dengan menggunakan metode kuadrat terkecil atau *Ordinary Least Squares* (OLS) yaitu dengan meminimumkan jumlah kuadrat error.

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = y'y - 2\beta'X'y + \beta'X'X\beta \quad (2.3)$$

Untuk mendapatkan estimator metode kuadrat terkecil atau OLS ( $\hat{\beta}$ ) yang meminimumkan  $S(\beta)$  disyaratkan bahwa:

$$\frac{\partial S(\beta)}{\partial \beta} \Big|_{\beta=\hat{\beta}} = 0$$

Maka

$$\begin{aligned} \frac{\partial S(\beta)}{\partial \beta} \Big|_{\beta=\hat{\beta}} &= 0 \\ -2X'y + 2X'X\hat{\beta} &= 0 \\ X'X\hat{\beta} &= X'y \\ \hat{\beta} &= (X'X)^{-1}(X'y) \end{aligned} \quad (2.4)$$

perbedaan unit satuan pada model regresi yang tidak distandarkan dapat menyebabkan koefisien regresi tidak bisa dibandingkan. Oleh karena itu, perlu dilakukan standarisasi menggunakan rumus sebagai berikut:

$$Y_i^* = \frac{Y_i - \bar{y}}{S_y} \text{ dan } X_{ji}^* = \frac{X_{ji} - \bar{X}_j}{S_j} \quad (2.5)$$

Dengan  $S_y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{y})^2}{n-1}}$  dan  $S_j = \sqrt{\frac{\sum_{i=1}^n (Y_{ji} - \bar{X}_j)^2}{n-1}}$

Sehingga diperoleh model regresi standar sebagai berikut:

$$Y_1^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \dots + \beta_k^* X_{ki}^* + \varepsilon_i^*$$

## 2.2 Uji Multikolinearitas

Multikolinearitas merupakan keadaan dimana terjadi hubungan linear yang sempurna atau mendekati antar variabel independen dalam model regresi. Suatu model regresi dikatakan mengalami multikolinearitas jika ada fungsi linear yang sempurna pada beberapa atau semua independen variabel dalam fungsi linear. Gejala adanya multikolinearitas antara lain dengan melihat nilai *Variance Inflation Factor* (VIF) dan Tolerance nya. Jika nilai  $VIF < 10$  dan

$Tolerance > 0,1$  maka dinyatakan tidak terjadi multikolinearitas (Mardiatmoko, 2020).

Deteksi adanya multikolinearitas pada model regresi dapat dilakukan dengan menggunakan nilai *Variance Inflation Factory* (VIF). *Variance Inflation Factory* (VIF) yaitu faktor perubahan variansi dalam variabel respon ke- $i$ . Pada VIF besarnya nilai bergantung pada nilai koefisien determinasi ( $R^2$ ) yang dihasilkan. Berikut persamaan yang digunakan dalam menghitung VIF:

$$VIF = \frac{1}{(1 - R_i^2)} = \frac{1}{Tolerance} \quad (2.6)$$

Dengan  $R_i^2$  adalah koefisien determinasi ke- $i$ , di mana  $i = 1,2,3, \dots, k$ . Tolerance (TOL) dalam perhitungan VIF memiliki ketentuan semakin rendah nilai yang diperolehnya maka semakin besar kemungkinan terjadinya multikolinieritas antar variabel. Batas nilai yang digunakan untuk VIF adalah antara 5 sampai 10 dan batas nilai dari tolerance adalah 0.1. Apabila nilai yang diperoleh  $VIF \geq 10$  dan  $tolerance < 0.1$  maka terjadi multikolinearitas yang kuat di antara variabel respon dan sebaliknya.

### 2.3 Regresi Ridge

Regresi *ridge* dikembangkan untuk menstabilkan nilai koefisien regresi yang mungkin terpengaruh oleh multikolinieritas. Ini pertama kali diperkenalkan oleh A. E. Hoerl pada tahun 1962. Regresi *ridge* merupakan variasi dari metode OLS yang menghasilkan penduga yang sedikit terpengaruh oleh multikolinieritas dengan mengurangi varians penduga. Meskipun penduga koefisien regresi bias, metode ini dapat mendekati nilai parameter yang sebenarnya (Patil dkk., 2021). Regresi *ridge* meminimumkan JKG dengan batasan L2-norm kuadrat dari koefisien regresi. Sehingga estimasi parameter pada model regresi *Ridge* adalah (Yanke dkk., 2022):

$$\hat{\beta}_{ridge} = \arg \min \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (2.7)$$

Dengan meminimumkan JKG regresi *ridge* diperoleh persamaan:

$$X'y = (X'X + \lambda I)^{-1} \hat{\beta} \quad (2.8)$$

Dalam proses estimasi koefisien regresi, nilai bias  $\lambda$  dimasukkan ke diagonal matriks  $X'X$ . Estimator regresi *ridge* yang dipengaruhi oleh besarnya tetapan bias  $\lambda$  dengan besarnya nilai  $\lambda$  berada pada selang  $0 \leq \lambda \leq 1$  adalah (Lestari dkk., 2022):

$$\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1} X'y, \lambda \geq 0 \quad (2.9)$$

#### 2.4 Regresi *Least Absolute Shrinkage And Selection Operator*

Metode LASSO merupakan hasil pengembangan dari Tibshirani dan diperkenalkan pertama kali pada tahun 1996 sebagai penyeleksi variabel dengan pendugaan parameter. Metode LASSO melakukan penyusutan koefisien regresi dari variabel yang memiliki korelasi tinggi terhadap galat, dengan tujuan koefisien regresi tersebut mendekati nol atau bahkan sama dengan nol. Sehingga metode LASSO dapat melakukan peran sebagai seleksi variabel sekaligus mengatasi multikolinearitas (Sartika dkk., 2020)

$$\hat{\beta}_{lasso} = \arg \min \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2.10)$$

Dengan fungsi kendala  $\sum_{j=1}^p |\beta_j| \leq t$ .

Menurut Robbani dkk {2019) nilai  $t$  adalah suatu besaran yang mengontrol besarnya penyusutan koefisien regresi LASSO. Nilai  $t$  dinamakan parameter tuning dengan ketentuan nilai  $t \geq 0$ . Misalkan diketahui  $\beta_j$  merupakan penaksir OLS, dengan nilai  $t_0$  didefinisikan  $t_0 = \sum_{j=1}^p |\beta_j|$ , maka:

- 1) Jika nilai  $t < t_0$ , maka koefisien OLS akan menyusut ke arah nol, dan memungkinkan untuk menjadi tepat nol.
- 2) Jika nilai  $t \geq t_0$ , maka koefisien regresi LASSO memberikan nilai yang sama dengan koefisien OLS.

## 2.5 Regresi *Elastic Net*

Metode *elastic net* pertama kali diperkenalkan oleh Zou dan Hastie pada tahun 2005. *elastic net* adalah metode regresi yang menggabungkan *ridge regression* dan *LASSO regression*. Metode ini menambahkan penalti L1 dan L2 pada fungsi loss, sehingga dapat mengatasi masalah multikolinearitas dan menyeleksi fitur yang penting secara bersamaan (Kusuma dan Wulansari, 2019). Persamaan *elastic net* sebagai berikut :

$$L(\lambda_1, \lambda_2, \beta) = |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda_2|\beta|^2 + \lambda_1|\beta| \quad (2.11)$$

Dengan :

$$|\mathbf{y} - \mathbf{X}\beta|^2 = |\mathbf{y} - \mathbf{X}\beta'| |\mathbf{y} - \mathbf{X}\beta|$$

$$|\beta|^2 = |\beta'| |\beta|$$

Menurut (Altelbany, 2021) *elastic net* adalah kelanjutan dari regresi LASSO yang kuat dalam menangani korelasi antar variabel prediktor. Untuk mencegah ketidakseimbangan saat variabel prediktor sangat berkorelasi. Zou dan Hastie (2005) merekomendasikan penggunaan estimator *elastic net* yang merupakan gabungan dari regresi *ridge* dan regresi LASSO dengan persamaan :

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

Parameter Regularisasi  $\lambda$  adalah jumlah dari dua penalti *nonnegative*  $\lambda = \lambda_1 + \lambda_2$ , maka nilai dari  $a = \frac{\lambda_1}{\lambda_1 + \lambda_2}$  dan  $1 - a = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ , dengan  $0 < a < 1$ . Maka diperoleh persamaan:

$$\hat{\beta}_{EN} = \arg \min \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \left( a \sum_{j=1}^p |\beta_j| + (1 - a) \sum_{j=1}^p \beta_j^2 \right) \right\} \quad (2.13)$$

Jika  $\alpha = 0$ , maka *elastic net* menjadi penalti gulud dan jika  $\alpha = 1$  maka *elastic net* menjadi penalti LASSO. Pada regularisasi *elastic net* terdapat penyusutan koefisien bersama dari peubah-peubah penjelas yang berkorelasi seperti gulud dan seleksi peubah seperti LASSO. Jika  $\lambda = 0$ , maka model *elastic net* berubah menjadi OLS (Racharla dkk., 2020).



## 2.6 Evaluasi Performa Model

*Root Mean Square Error* (RMSE) adalah matriks evaluasi yang mengukur seberapa dekat prediksi model dengan nilai yang sebenarnya. Metrik ini dihitung dengan cara menghitung selisih antara nilai prediksi dan nilai sebenarnya, kemudian megelevasikan hasilnya ke kuadrat, menjumlahkan semua hasil tersebut, dan kemudian mengambil akar kuadrat dari jumlah tersebut. RMSE mengukur besarnya kesalahan prediksi secara keseluruhan, dengan semakin kecil nilai RMSE, semakin baik kinerja model (Hodson, 2022). Adapun untuk persamaan RMSE :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^k (\tilde{y}_i - y_i)^2} \quad (2.14)$$

Selain metode RMSE, untuk pemilihan model terbaik dapat digunakan uji kebaikan model, yakni menggunakan *R-squared* atau juga disebut *Coefficient of Determination*, adalah matriks evaluasi yang mengukur seberapa baik model cocok dengan data yang diamati. Matriks ini mengukur proporsi variasi dalam data yang dapat dijelaskan oleh model. *R-squared* dihitung dengan membagi variasi yang dijelaskan oleh model dengan total variasi dalam data.  $R^2$  memiliki rentang nilai dari 0 hingga 1, dengan nilai yang lebih tinggi menunjukkan bahwa model lebih baik dalam menjelaskan data (Gujarati, 2021). Kebanyakan penelitian lebih sering menggunakan koefisien determinasi yang di sesuaikan (*Adjusted R<sup>2</sup>*), rumusnya sebagai berikut:

$$R_{adj}^2 = 1 - \frac{SS_E / (n - k - 1)}{SS_R / (n - 1)} \quad (2.15)$$

Dengan SSR adalah jumlah kuadrat regresi, SSE adalah jumlah kuadrat residual,  $k$  adalah variabel prediktor, dan  $n$  adalah jumlah sampel.

## 2.7 Coordinate Descent

*Coordinate Descent* (CD) adalah sebuah algoritma optimasi yang digunakan untuk menyelesaikan berbagai macam masalah seperti regresi linear, regresi logistik, regresi Poisson, serta beberapa masalah optimisasi lainnya (Khan dan

Byram, 2022). *coordinate descent* merupakan algoritma optimisasi yang mengambil langkah optimasi dalam satu dimensi pada suatu waktu untuk mengurangi kompleksitas perhitungan pada setiap iterasi. Pada setiap iterasi, CD hanya mengubah satu variabel dan variabel lainnya tetap atau konstan. Pada konteks regresi, CD memperbarui koefisien regresi dengan memilih satu variabel independen pada suatu waktu. CD akan mengiterasi melalui semua variabel independen dalam urutan tertentu, dan kemudian memulai lagi dari awal setelah melakukan satu iterasi lengkap melalui semua variabel independen (Hazimeh dan Mazumder, 2020).

## 2.8 Cross Validation

Validasi silang (*Cross Validation-CV*) merupakan metode yang digunakan untuk menentukan ketepatan model dengan cara membagi data menjadi beberapa bagian. Idealnya, ketika data yang dimiliki cukup banyak, data validasi dapat ditentukan dengan mudah. Namun, ketika data yang dimiliki terbatas, solusinya adalah melakukan validasi silang dengan menggunakan sebagian data untuk mengepas model dan sebagian data lainnya sebagai data pengujian model (Fanny dkk., 2018). Proses validasi silang melibatkan pembagian data menjadi dua kelompok, yaitu kelompok data pemodelan dan kelompok data validasi. Penentuan kelompok data dilakukan secara acak agar setiap pengamatan memiliki kesempatan yang sama untuk menjadi data validasi (Kusuma dan Wulansari, 2019).

Salah satu jenis validasi silang adalah validasi silang lipat-k. Teknik ini sangat efektif digunakan ketika jumlah sampel yang tersedia tidak banyak. Dalam metode validasi silang lipat-k, sampel dikelompokkan ke dalam k grup secara acak untuk diuji. (Fanny dkk., 2018).

$$CVE_{(k)} = \frac{1}{k} \sum_{i=1}^k RMSE_i \quad (2.16)$$

Metode *cross validation* yang paling efektif adalah *5-fold* atau *10-fold*. Hal ini dikarenakan metode ini memberikan dugaan sisaan prediksi dengan bias yang rendah, namun memiliki RMSE yang kecil dan variansi yang lebih rendah. Hal ini sangat penting dalam meningkatkan akurasi prediksi dan mengurangi resiko

kesalahan dalam pembuatan keputusan berdasarkan hasil prediksi (Mayapada dkk., 2018).

## 2.9 Angka Kematian Bayi

Indikator kematian bayi, yang juga dikenal sebagai *Infant Mortality Rate* (IMR) adalah mengukur jumlah bayi yang meninggal sebelum mencapai usia satu tahun per 1000 kelahiran hidup pada tahun yang sama. IMR dapat memberikan gambaran tentang kondisi sosial ekonomi masyarakat karena bayi merupakan kelompok usia yang sangat rentan terhadap dampak dari perubahan lingkungan dan kondisi sosial ekonomi (Wardani dan Wulandari, 2020). Adapun untuk menghitung angka kematian bayi menggunakan rumus sebagai berikut:

$$AKB = \frac{\text{jumlah kematian bayi}}{\text{jumlah bayi lahir}} \times 1000 \quad (2.15)$$

AKB dapat menjadi indikator kesehatan masyarakat yang penting karena dapat mencerminkan kondisi kesehatan umum dan kemajuan pembangunan di suatu negara. Selain itu, angka ini juga dapat memberikan informasi tentang masalah-masalah kesehatan yang mungkin terjadi pada ibu dan bayi selama kehamilan dan persalinan. Misalnya, AKB yang tinggi dapat menunjukkan adanya masalah dalam akses terhadap perawatan medis dan pendidikan kesehatan, kurangnya akses terhadap air bersih dan sanitasi, atau masalah dalam status gizi ibu dan bayi. Oleh karena itu, pengurangan angka kematian bayi menjadi tujuan penting bagi pemerintah dan organisasi kesehatan di seluruh dunia, dan upaya-upaya terus dilakukan untuk mencapai tujuan tersebut melalui program-program kesehatan masyarakat yang tepat dan efektif serta peningkatan akses terhadap perawatan medis yang memadai (Lestari dkk., 2022).