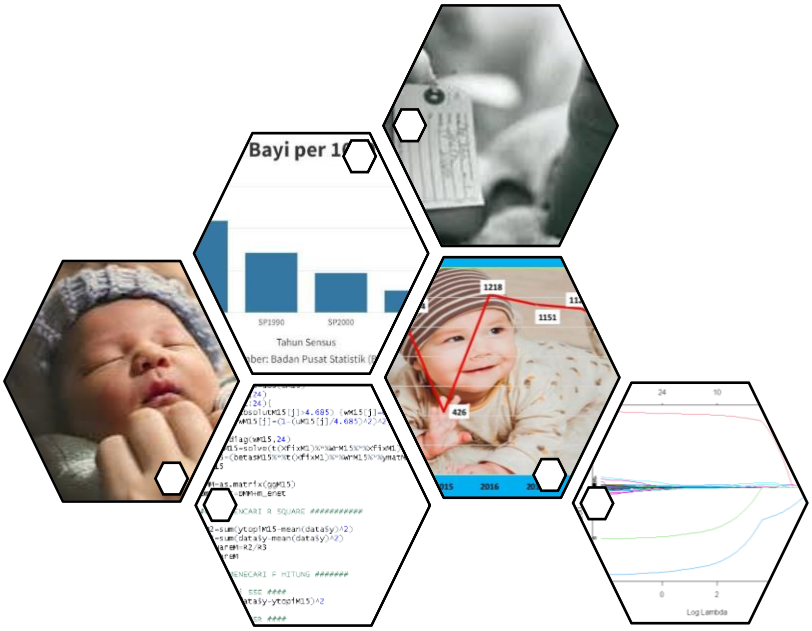
PEMODELAN REGRESI *ELASTIC NET ROBUST MM-ESTIMATION* PADA DATA ANGKA KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN TAHUN 2020



NUR AVIATUL ZAHRA
H051201052

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

PEMODELAN REGRESI *ELASTIC NET ROBUST MM-ESTIMATION* PADA DATA ANGKA KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN TAHUN 2020

NUR AVIATUL ZAHRA
H051201052

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

PEMODELAN REGRESI *ELASTIC NET ROBUST MM-ESTIMATION* PADA DATA ANGKA KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN TAHUN 2020

NUR AVIATUL ZAHRA
H051201052

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Statistika

pada

PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

SKRIPSI

PEMODELAN REGRESI *ELASTIC NET ROBUST MM-ESTIMATION* PADA DATA ANGKA KEMATIAN BAYI DI PROVINSI SULAWESI SELATAN TAHUN 2020

NUR AVIATUL ZAHRA
H051201052

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada 05 Agustus 2024
dan dinyatakan telah memenuhi syarat kelulusan
pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing Tugas Akhir,

Sitti Sahriman, S.Si., M.Si.
NIP. 19881018 201504 2 002

Mengetahui:
Ketua Program Studi,

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

iv

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Pemodelan Regresi *Elastic Net Robust MM-Estimation* Pada Data Angka Kematian Bayi di Provinsi Sulawesi Selatan Tahun 2020" adalah benar karya saya dengan arahan dari pembimbing Sitti Sahriman, S.Si., M.Si. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 2 Agustus 2024

METERAI TEMPEL
10000
3023EALX293038776

NUR AVIATUL ZAHRA
NIM H051201052

v

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul "Penggunaan Pemodelan Regresi *Elastic Net Robust MM-Estimation* Pada Data Angka Kematian Bayi di Provinsi Sulawesi Selatan Tahun 2020" yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak sehingga penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada **Ibu Sitti Sahriman S.Si., M.Si**. selaku dosen pembimbing yang telah memberikan arahan, bimbingan, dan motivasi kepada penulis dari awal hingga selesainya skripsi ini. Ucapan terima kasih juga penulis sampaikan kepada **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.** dan **Ibu Dr. Anna Islamiyati, S.Si., M.Si**. selaku dosen penguji yang telah memberikan masukan dan kritik dalam penyempurnaan skripsi ini. Tidak lupa pula penulis menyampaikan terima kasih kepada Rektor Universitas Hasanuddin dan seluruh jajaran Fakultas Matematika dan Ilmu Pengetahuan Alam serta Departemen Statistika karena telah memberikan fasilitas yang sangat baik kepada penulis selama menempuh pendidikan sarjana.

Kepada cinta pertama dan panutan penulis **Bapak Syarifuddin** dan **Ibu Nurdiana** yang selalu menjadi penyemangat dan sandaran dari segala macam masalah. Yang tak henti-hentinya mencurahkan kasih sayang, nasihat, perhatian, serta senantiasa memberikan dukungan baik secara moral maupun finansial dan tak kenal lelah mendoakan hingga penulis mampu menyelesaikan studinya sampai meraih gelar sarjana. Semoga Bapak dan Ibu sehat, panjang umur dan bahagia selalu. Ucapan terima kasih juga penulis haturkan kepada saudara tersayang penulis **Nur Isnawati**, yang selalu memberikan motivasi dan semangat kepada penulis.Terima kasih juga kepada sahabat saya **Dilla** dan **Melvi** yang selalu mendengar setiap keluh kesah dan memberikan respon positif serta nasehat yang mendukung. Terima kasih sudah menjadi sahabat yang baik bahkan seperti saudara.Teman-teman Departemen Statistika Angkatan 2020 khususnya **Ira, Dwi, Ayu, Dania, Aliyah, Nurfa, Tiwi, Isra** yang senantiasa menjadi partner terbaik dalam proses penulisan tugas akhir ini hingga selesai.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga skripsi ini dapat memberikan manfaat untuk berbagai pihak.

Penulis,

Nur Aviatul Zahra

ABSTRAK

NUR AVIATUL ZAHRA. **Pemodelan Regresi *Elastic Net Robust MM-Estimation* Pada Data Angka Kematian Bayi di Provinsi Sulawesi Selatan Tahun 2020** (dibimbing oleh Sitti Sahriman, S.Si., M.Si.).

Latar belakang. Suatu data yang akan dianalisis menggunakan regresi linier berganda harus dipastikan memenuhi asumsi-asumsi klasik. Namun pada kenyataannya seringkali terdapat pelanggaran pada satu atau lebih asumsi seperti terjadinya multikolinieritas dan pencilan. Penanganan multikolinearitas dapat dilakukan dengan metode regresi *elastic net*, sementara pencilan dapat diatasi menggunakan metode *robust MM-estimation*. Kedua metode tersebut dapat dikombinasikan menjadi metode regresi *elastic net robust* yang dapat mengatasi masalah multikolinearitas dan pencilan. Kelebihan dari metode ini adalah mampu mencegah *overvitting* dengan penalti yang diterapkan sehingga menghasilkan model yang lebih stabil karena varians yang ekstrem dalam data berkurang.**Tujuan**. Penelitian ini bertujuan untuk memperoleh estimasi parameter model *elastic net robust MM-estimation* dan memperoleh faktor apa saja yang mempengaruhi data. **Metode**. Penelitian ini menggunakan data angka kematian bayi di Provinsi Sulawesi Selatan Tahun 2020. **Hasil**. Hasil penelitian ini menunjukkan bahwa model *elastic net robust MM-estimation* memiliki nilai R^2 yang lebih tinggi sebesar 97,23% dibandingkan dengan model *elastic net* (61,28%). Pada model *elastic net robust MM-estimation* nilai *mean absolute error* (MAE) yang dihasilkan sebesar 0,664851 dan nilai *mean absolute percentage error* (MAPE) sebesar 0,901648. Nilai tersebut lebih kecil dibandingkan pada metode *elastic net* yang memiliki nilai MAE sebesar 0,721729 dan MAPE sebesar 0,955319. **Kesimpulan**. Pada tingkat signifikansi sebesar 5% faktor-faktor yang berpengaruh signifikan terhadap angka kematian bayi di Provinsi Sulawesi Selatan Tahun 2020 adalah jumlah pelayanan kesehatan bayi, jumlah pemberianVitamin A, jumlah bayi BBLR, dan jumlah ibu bersalin di tolong medis.

Kata kunci: Angka Kematian Bayi; *MM-estimation* ; Multikolinieritas; Pencilan; Regresi *Elastic Net Robus*

ABSTRACT

NUR AVIATUL ZAHRA. **Robust Elastic Net Regression Modeling MM-Estimation on Infant Mortality Data in South Sulawesi Province for the Year 2020** (supervised by Sitti Sahriman, S.Si., M.Si, S.Si., M.Si.)

Background. A dataset that will be analyzed using multiple linear regression must meet classical assumptions. However, in reality, there are often violations of one or more assumptions, such as multicollinearity and outliers. Handling multicollinearity can be achieved with the elastic net regression method, while outliers can be addressed using robust MM-estimation. These two methods can be combined into a robust elastic net regression method that addresses both multicollinearity and outliers. The advantages of this method include its ability to prevent overfitting with penalties applied to coefficient magnitudes and its ability to produce a more stable model due to the reduction in extreme variance in the data. **Aim.** This study aims to obtain parameter estimates for the robust elastic net MM-estimation model and to identify the factors that influence the data. **Method.** This study applied t mortality data in South Sulawesi Province for the year 2020. **Results**. The results of this study showed that the robust elastic net MM-estimation model had an R^2 value of 97,23%, which was higher compared to the elastic net model (61,28%). In the robust elastic net MM-estimation model, the mean absolute error (MAE) was 0,664851 and the mean absolute percentage error (MAPE) was 0,901648. These values were lower compared to the elastic net method, which had an MAE of 0,721729 and a MAPE of 0,955319. **Conclusion.** At a 5% significance level, the factors that significantly affected the infant mortality rate in South Sulawesi Province in 2020 were the number of baby health services, the amount of Vitamin A administration, the number of low birth weight babies, and the number of medically assisted deliveries.

Keywords: Infant Mortality Rate; MM-Estimation; Multicollinearity; Outliers; Elastic Net Regression Robust

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Bias	Perbedaan antara estimasi yang dihasilkan oleh suatu metode dan nilai sebenarnya dari parameter yang diestimasi
Breakdown Point	Ukuran kekekaran suatu estimasi
Cross validation	Teknik dalam melakukan validasi model terbaik.
Elastic net	Gabungan dari regresi *ridge* dan regresi LASSO
Estimasi	Proses menggunakan data sampel untuk menilai atau mengira parameter populasi yang tidak diketahui.
Galat	Perbedaan antara nilai yang diobservasi dan nilai yg sebenarnya.
Iterasi	Proses pengulangan suatu langkah untuk mencapai hasil yang diinginkan.
Korelasi	Ukuran yang menunjukkan sejauh mana dua variabel atau lebih saling berkaitan atau berhubungan.
Least Absolute Shrinkage	Metode untuk pemilihan variabel dan regularisasi dengan menambahkan jumlah nilai absolut.
Matriks	Struktur yang terdiri dari elemen-elemen yang disusun dalam bentuk tabel atau grid
Multikolinieritas	Kondisi dimana dua atau lebih variabel prediktor saling berkorelasi satu sama lain.
Overfitting	Kondisi model terlalu kompleks dan menangkap noise dalam data pelatihan
Outlier	Nilai atau pengamatan yang berbeda dari nilai-nilai lainnya
Underfitting	Kondisi dimana model terlalu sederhana untuk menangkap pola dalam data.
Parameter	Nilai yang digunakan untuk mendefinisikan fungsi atau model.
Regresi Ridge	Metode untuk pemilihan variabel dan regularisasi dengan menambahkan jumlah kuadrat dari koefisiennya.
Regulasi	Teknik untuk mencegah *overfitting* dengan menambahkan penalti pada model
Robust	Teknik untuk mengurangi pengaruh dari gangguan data dan memastikan bahwa model atau analisis tetap akurat dan konsisten.

DAFTAR SIMBOL DAN SINGKATAN

Simbol/Singkatan	Arti dan Penjelasan
α	Parameter kontrol terhadap penalti
β_0	Konstanta
$\hat{\beta}_{lasso}$	Nilai estimasi parameter regresi LASSO
$\hat{\beta}_{MM}$	*Robust MM Estimation*
β_j	Koefisien regresi variabel prediktor ke- $j(j = 1,2, \dots p)$
$\hat{\beta}_{ridge}$	Nilai estimasi parameter regresi *ridge*
e_i	Variabel residual ke-*i*
H_0	Hipotesis nol
H_1	Hipotesis alternatif
h_{ii}	Nilai *laverage*
I	Matriks Identitas
λ_1	Penalti regresi LASSO
λ_2	Penalti regresi *ridge*
n	Banyaknya sampel
p	Banyaknya variabel prediktor
R_j^2	Nilai koefisien determinasi
r_{jk}	Korelasi antara variabel X_j dan X_k
ρ	Fungsi objektif
S_{jj}	Standar deviasi variabel X_j
S_{kk}	Standar deviasi variabel X_k
$\hat{\sigma}$	Estimasi skalara *robust*
t	Nilai fungsi kendala
$\bar{X}_j$	Rata rata variabel X_j
$\bar{X}_k$	Rata rata variabel X_k
X_{ij}	Variabel prediktor ke-*j* untuk observasi ke-*i*
y_i	Variabel respon untuk observasi ke-$i(i = 1,2, \dots n)$
c	Nilai *tunning constant*
AKB	Angka Kematian Bayi
DFFITS	*Difference Fitted Value FITS*
LASSO	*Least absolute shrinkage and selection operator*
MKT	Metode Kuadrat Terkecil
OLS	*Ordinary least square*
RMSE	*Root Mean Square Error*
SSE	*Sum Square Error*
VIF	*Variance Inflation Factor*

DAFTAR ISI

Halaman

DAFTAR TABEL

Nomor Urut **Halaman**

DAFTAR LAMPIRAN

Nomor Urut **Halaman**

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan adalah salah satu aspek penting dalam menilai kualitas sumber daya manusia yang tidak hanya penting di tingkat global tetapi juga merupakan upaya memenuhi komitmen internasional. Salah satu indikator penting untuk mengetahui derajat kesehatan di suatu negara serta untuk mengukur tingkat kemajuan suatu bangsa adalah angka kematian bayi. Angka Kematian Bayi (AKB) adalah jumlah bayi yang meninggal sebelum mencapai umur 1 tahun pada waktu tertentu per 1000 kelahiran hidup pada kurun waktu yang sama (Dinas Kesehatan, 2021). AKB rentan terhadap keadaan lingkungan tempat tinggal. Oleh karena itu, AKB merupakan tolak ukur yang sensitif untuk setiap tindakan yang dilakukan pemerintah, terutama di bidang Kesehatan (Marizal & Monalisa, 2022)

Berdasarkan data Profil Kesehatan tahun 2019 Dinas Kesehatan Provinsi Sulawesi Selatan, tercatat jumlah AKB pada tahun 2018 mencapai angka 1039 kematian. Berbeda halnya pada tahun 2019 dan 2020 AKB mengalami penurunan. Pada tahun 2019 jumlah kematian bayi menjadi 919 bayi atau 6,02 per 1000 kelahiran hidup sedangkan kematian bayi pada tahun 2020 mencapai angka 754 bayi atau 4,87 per 1000 kelahiran hidup. Ada banyak faktor pemicu yang mempengaruhi tingkat AKB diantaranya adalah berat badan lahir rendah (BBLR) yang terjadi karena asupan gizi ibu yang kurang saat sedang hamil dan bayi lahir secara premature. Selain itu faktor lain yang mempengaruhi AKB adalah keterbatasan jumlah sarana kesehatan, ketidakrutinan dalam pemeriksaan kesehatan bayi, kekurangan pemberian vitamin A pada bayi usia 6-11 bulan, serta kekurangan jumlah tenaga medis (Dinas Kesehatan, 2021). Untuk mengetahui faktor-faktor apa saja yang mempengaruhi AKB metode statistika yang umum digunakan adalah analisis regresi.

Analisis regresi merupakan suatu metode statistika yang digunakan untuk memeriksa hubungan antara variabel respon dan prediktor dalam suatu penelitian. Hubungan pada metode tersebut sebenarnya tidak dapat diketahui secara pasti tetapi model hubungan tersebut dapat di estimasi berdasarkan data pengamatan. Metode yang paling umum digunakan untuk mengestimasi parameter regresi adalah metode kuadrat terkecil (MKT). Metode ini dapat digunakan untuk penaksiran koefisien regresi dengan meminimalkan jumlah kuadrat galatnya (Mahalani dan Rifai, 2022). Selain itu, dapat menghasilkan pendugaan yang efisien apabila asumsi klasik terpenuhi. Namun, jika terjadi pelanggaran asumsi klasik, maka estimasi yang dihasilkan bersifat bias dan hasil estimasi model regresinya menjadi tidak tepat atau tidak akurat (Aflakhah dkk., 2020). Pelanggaran asumsi klasik yang umum terjadi pada pemodelan regresi yang melibatkan banyak variabel prediktor adalah multikolinearitas.

Multikolinieritas merupakan suatu kondisi adanya hubungan atau korelasi antara dua atau lebih variabel prediktor yang mengakibatkan model regresi yang diperoleh jauh dari akurat sehingga tidak memungkinkan melihat pengaruh variabel prediktor terhadap variabel responnya (Gujarati, 1992). Multikolinearitas dapat mengakibatkan beberapa masalah dalam analisis regresi, seperti membuat estimasi koefisien regresi menjadi tidak stabil, menyebabkan standar residual koefisien menjadi besar, dan menyebabkan koefisien regresi menjadi tidak signifikan. Terdapat beberapa cara untuk mengatasi multikolinearitas dalam regresi berganda seperti regresi *ridge*, *least absolute shrinkage and selection operator* (LASSO) dan *elastic net*.

Regresi *ridge* merupakan metode yang mengestimasi parameter menggunakan jumlah kuadrat eror dengan bergantung pada jumlahan kuadrat koefisien-koefisiennya. Sementara itu metode ini membatasi koefisien semua variabel, sehingga masalah multikolinearitas mampu diatasi. Kelemahan metode ini adalah tidak efektif dalam mencegah *overfitting* (Tsigler dan Bartlett, 2020). Berbeda dengan regresi *ridge*, regresi LASSO mampu menghilangkan variabel-variabel pada model regresi dengan menyusutkan koefisiennya menjadi tepat nol. Kelebihan dari regresi ini, yaitu mampu menyeleksi variabel prediktor dalam model (Pardede, 2022). Meskipun demikian, pemilihan variabel yang tidak stabil menjadi kelemahan dari regresi LASSO. Oleh karena itu, untuk menutupi kelemahan dari metode LASSO dan *ridge*, pendekatan yang digunakan untuk menggabungan kedua teknik tersebut yaitu dengan regresi *elastic net*.

Metode regresi *elastic net* merupakan sebuah metode regularisasi yang menggabungkan regresi *ridge* dan regresi LASSO. Metode ini menggunakan kombinasi dari penalti L1 (regresi LASSO) dan penalti L2 (regresi *ridge*) untuk melakukan seleksi variabel dan mengontrol multikolinearitas. Metode ini dapat melakukan kontrol yang lebih baik dalam memilih variabel yang penting, serta dapat mengidentifikasi kelompok variabel yang saling berkorelasi (Sinaga dkk., 2021). Regresi *elastic net* memberikan fleksibilitas yang lebih besar dalam mengontrol parameter alpha yang mempengaruhi kombinasi penalti *ridge* dan LASSO. Dalam regresi *elastic net*, ketika alpha mendekati 0, model akan cenderung ke arah regresi *ridge* dan ketika alpha mendekati 1, model akan cenderung ke arah regresi LASSO. Regresi *elastic net* memiliki beberapa keunggulan dibandingkan regresi LASSO dan *ridge*, salah satunya adalah kemampuannya dalam mengatasi masalah *underfitting* dan *overfitting* secara bersamaan (De Leone dkk., 2020).

Pelanggaran asumsi yang sering terjadi setelah multikolinieritas dalam regresi yaitu adanya pencilan dalam data. Pencilan merupakan pengamatan yang jauh dari pusat data yang dapat berpengaruh besar terhadap koefisien regresi. Pencilan dapat menyebabkan varians pada data menjadi lebih besar, *mean* tidak dapat menunjukkan nilai sebenarnya (bias) serta error yang besar dari model yang terbentuk. Oleh karena itu, perlu dilakukan analisis untuk mengatasi pencilan. Salah satu caranya yaitu menggunakan regresi *robust*. Regresi *robust* merupakan metode yang digunakan untuk menganalisis data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang *robust* atau resistansi terhadap pencilan (Al Amin & Juniati, 2021). Regresi *robust* memiliki beberapa estimasi salah satunya adalah *MM-estimation*. Metode estimasi tersebut gabungan dari *S-estimation* dan *M-estimation* yang memiliki *breakdown point* dan efisiensi tinggi (Setyowati dkk., 2021).

Beberapa penelitian terdahulu telah menggunakan regresi *elastic net* diantaranya (Handayani & Wachidah, 2023) menggunakan regresi *elastic net* untuk mengatasi masalah multikolinearitas pada kasus tigkat pengangguran terbuka di Provinsi Jawa Barat. (Nur dkk., 2024) menerapkan perbandingan metode regresi *ridge*, LASSO dan *elastic-net* pada kasus AKB di Sulawesi Selatan. Selain itu, (Guntur &

Wulansari, 2019) membandingkan regresi *ridge*, lasso, dan *elastic-net* dalam menganalisis kemiskinan dan kerentanan di Provinsi Jawa Tengah Tahun 2017. Serta Fikri dkk. (2023) melakukan penelitian dengan melihat performa teknik regularisasi dalam penanganan masalah multikolinieritas menggunakan regresi *ridge*, LASSO dan *elastic net*. Penggunaan metode *elastic net* pada masalah multikolinearitas tidak efisien jika data tidak berdistribusi normal yang seringkali disebabkan karena adanya pencilan pada data yang digunakan. Akan tetapi, belum banyak penelitian yang menganalisis data terindikasi multikolinieritas sekaligus pencilan dengan menggunakan metode *elastic net robust*. Berdasarkan uraian di atas, penulis akan mengkaji tentang penggunaan regresi *elastic net robust* pada data yang mengandung multikolinearitas dan pencilan yang dituliskan dalam sebuah tugas akhir dengan judul **"Pemodelan Regresi *Elastic Net Robust MM-Estimation* pada Data Angka Kematian Bayi Di Provinsi Sulawesi Selatan Tahun 2020"**

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi parameter model *elastic net robust* dengan *MM-estimation* pada data yang mengandung multikolinearitas dan *outlier* pada data Angka Kematian Bayi di Provinsi Sulawesi Selatan tahun 2020?
2. Bagaimana faktor yang mempengaruhi Angka Kematian Bayi di Provinsi Sulawesi Selatan tahun 2020?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penggunaan *robust MM-estimation* pada regresi *elastic net*
2. Data yang digunakan adalah data Angka Kematian Bayi (AKB) Provinsi Sulawesi Selatan tahun 2020.
3. Pemilihan model terbaik berdasarkan nilai MAE, MAPE dan *R-Squared*

1.4 Tujuan Penelitian

Berdasarkan rumusan yang telah diuraikan, maka tujuan penelitian ini adalah sebagai berikut:

1. Untuk memperoleh estimasi parameter model *elastic net robust MM-estimation* pada data Angka Kematian Bayi di Provinsi Sulawesi Selatan tahun 2020.
2. Untuk memperoleh faktor apa saja yang Angka Kematian Bayi di Provinsi Sulawesi Selatan tahun 2020.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat menambah wawasan pengetahuan mengenai model *elastic net robust* pada data yang mengandung multikolinearitas dan *outlier*.. Selain itu, memberikan informasi mengenai faktor-faktor yang mempengaruhi angka kematian bayi di Provinsi Sulawesi Selatan dengan menggunakan model *elastic net* *robust*, sehingga dapat menjadi acuan bagi pemerintah dan masyarakat dalam upaya menurunkan angka kematian bayi di Indonesia khususnya di Provinsi Sulawesi Selatan.

1.6 Kajian Teori

1.6.1 Regresi Linear Berganda

Model regresi linier berganda adalah pengembangan dari model regresi linier sederhana, dengan jumlah variabel prediktornya (X) lebih dari satu dan hanya ada satu variabel respon (Y). Regresi linier berganda juga dapat dikatakan sebagai model persamaan yang menjelaskan hubungan satu variabel respon dengan satu atau lebih variabel prediktornya. Regresi linier berganda bertujuan untuk memprediksi nilai variabel respon apabila nilai dari variabel prediktornya diketahui. Adapun model yang digunakan untuk analisis regresi berganda sebagai berikut (Sengupta, 2001):

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \cdots + \beta_p X_{ip} + e_i \tag{1}$$

Berdasarkan Persamaan (1) maka dapat ditulis bentuk matriks dari model regresi linier berganda dapat dituliskan sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{2}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

dengan,

$\boldsymbol{y}$: Vektor pada variabel respon berukuran $(n \times 1)$

$\boldsymbol{X}$: Matriks dari variabel prediktor berukuran $(n \times (p+1))$

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

$\boldsymbol{e}$: Vektor sisaan yang berukuran $(n \times 1)$

1.6.2 Metode Kuadrat Terkecil

Untuk mengestimasi parameter β salah satu metode yang paling sering digunakan dalam teknik analisis regresi adalah metode kuadrat terkecil atau *ordinary least square* (OLS). Metode ini sama dengan meminimumkan $e'e$ (jumlah kuadrat residual) sehingga nilai regresinya akan mendekati nilai yang sebenarnya (Candraningtyas dkk., 2023). Dari Persamaan (2) fungsi kuadrat residual dapat ditulis:

$$\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{X\beta}) \tag{3}$$

$$\begin{aligned} S(\beta) &= \sum_{i=1}^{n} e_i^2 = e'e \\ &= (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= (\boldsymbol{y}' - \boldsymbol{\beta}'\boldsymbol{X}')(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} \end{aligned} \tag{4}$$

Dari Persamaan (4) di atas kemudian diturunkan secara parsial terhadap $\boldsymbol{\beta}$ dan disamakan dengan nol untuk mendapatkan $\widehat{\boldsymbol{\beta}}$. sehingga diperoleh estimasi kuadrat terkecil untuk parameter $\widehat{\boldsymbol{\beta}}$ sebagai berikut:

$$\frac{\partial(\boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}}{\partial\boldsymbol{\beta}}=\mathbf{0}$$

$$-2\boldsymbol{X}'\boldsymbol{y}+2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\mathbf{0}$$

$$2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}=2\boldsymbol{X}'\boldsymbol{y}$$

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{X}'\boldsymbol{y}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\widehat{\boldsymbol{\beta}}_{MKT}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{5}$$

1.6.3 Multikolinieritas

Pada tahun 1934 istilah multikolinearitas pertama kali diperkenalkan oleh Ragnar Frisch, yang berarti adanya korelasi atau hubungan di antara variabel prediktor dalam model regresi (Corlett & Aigner, 1972). Jika variabel prediktor saling berkolerasi (berhubungan) maka dinyatakan model tersebut tidak baik. Sebaliknya jika terbebas dari multikolinearitas maka model regresi linier dinyatakan baik. Adanya multikolinearitas pada model regresi tentu dapat memberikan dampak terhadap model tersebut. Dampak tersebut yaitu sebagai berikut (Neter, 1997):

1. Mengakibatkan variansi penduga kuadrat terkecil menjadi besar yang tentunya akan menghasilkan galat baku yang lebih besar sehingga berakibat terhadap selang kepercayaan untuk parameter model regresi menjadi lebih besar.
2. Adanya variabel prediktor yang saling berkolerasi berakibat pada penjelasan yang diberikan terhadap variabel respon menjadi tidak akurat (memberikan pengaruh yang sama terhadap variabel respon).
3. Hasil pengujian hipotesis parameter berdasarkan metode kuadrat terkecil menjadi tidak valid.

1.6.3.1 Variance Inflation Factor

Nilai *Variance Inflation Factor* (VIF) adalah cara yang dapat digunakan untuk mengidentifikasi multikolinieritas pada data. VIF untuk menghitung seberapa besar pengaruh korelasi antar variabel prediktor pada model regresi. Multikolinearitas muncul jika nilai $VIF > 10$. VIF dapat dihitung menggunaka rumus berikut (Auqino dkk., 2019):

$$VIF_j=\frac{1}{1-R_j^2} \tag{6}$$

dimana $1-R_j^2$ adalah nilai toleransi dan R_j^2 adalah nilai koefisien determinasi.

Besarnya nilai VIF ini bergantung pada nilai koefisien determinasi (R_j^2) yang dihasilkan. Batas nilai toleransi adalah 0,10 dan nilai VIF adalah 10. Jika $VIF > 10$ dan nilai toleransi < 0.10, maka tejadi multikolinearitas tinggi antar variabel prediktor dengan variabel prediktor lainnya. Jika $VIF < 10$ dan nilai toleransi > 0.10, maka dapat diartikan tidak terdapat multikolinearitas (Ridwan & Sunendiari, 2021).

1.6.3.2 Nilai Korelasi

Matriks dengan elemen diagonal utama berjumlah 1 disebut matriks simetris. Korelasi (r) bertujuan untuk mengukuran seberapa kuat hubungan antara dua variabel. Nilai korelasi dapat dilambangkan dengan persamaan berikut (Sarwono, 2006):

$$\rho = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}$$

$$r_{jk} = \frac{1}{n-2}\sum_{r=1}^{n}\left(\frac{x_{jk}-\bar{X}_j}{\sqrt{S_{jj}}}\right)\left(\frac{x_{jk}-\bar{X}_k}{\sqrt{S_{kk}}}\right) \tag{7}$$

Untuk $j = k$ menghasilkan $r = 1$

Hipotesis yang digunakan adalah sebagai berikut:

H_0 = Tidak ada hubungan yang signifikan

H_1 = Terdapat hubungan yang signifikan

Dasar pengambilan keputusannya adalah jika nilai *p-value* < 0.05 maka berkorelasi dan sebaliknya jika nilai *p-value* > 0.05 maka tidak berkorelasi. Niai korelasi juga dapat dibandingkan dengan r_{tabel} dengan ketentuan, jika nilai korelasi $> r_{tabel}$ maka berkorelasi dan sebaliknya jika nilai korelasi$< r_{tabel}$ maka tidak berkorelasi (Vusvitasari dkk., 2019).

Adapun pedoman derajat hubungan dijelaskan pada Tabel 2.1 berikut:

Tabel 1. Pedoman Derajat Hubungan

Interval koefisien	Tingkat hubungan
0.00-0.199	Sangat Lemah
0.20-0.399	Lemah
0.40-0.599	Sedang
0.60-0.799	Kuat
0.80-1.000	Sangat Kuat

1.6.4 Pencilan

Pencilan adalah pengamatan yang jauh dari pusat data yang dapat memengaruhi koefesien regresi secara signifikan. Pencilan dapat muncul karena kesalahan dalam memasukkan data, kesalahan pengukuran, analisis, atau kesalahan-kesalahan lain. Dalam analisis data, pengaruh pencilan dapat dibedakan berdasarkan asal pencilan tersebut yaitu apakah pencilan berasal dari peubah respons atau dari peubah prediktornya.

1.6.4.1 Metode *Difference Fitted Value* FITS

Difference Fitted Value FITS (*DFFITS*) adalah salah satu teknik yang dapat digunakan untuk mengidentifikasi pencilan. Metode *DFFITS* merupakan suatu ukuran berpengaruh yang ditimbulkan oleh pengamatan ke-I terhadap nilai taksiran $\hat{y}_i$ (Nurdin dkk., 2019) Metode ini merupakan gabungan antara metode *laverage* (hii) dan ti . ti

adalah R-student (*studentized deleted residual*) untuk kasus ke-i. Rumus untuk menghitung nilai *DFFITS* yaitu (Siburian dkk., 2019):

$$DFFITS = t_i \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}} \quad (8)$$

dengan

$$t_i = e_i \sqrt{\frac{n - p - 1}{SSE(1 - h_{ii}) - e_i^2}} \quad (9)$$

dimana e_i adalah residual ke -*i* dan SSE adalah jumlah kuadrat residual. h_{ii} merupakan elemen diagonal ke-i dari matriks H dengan persamaan sebagai berikut:

$$H = X_i (X^T X)^{-1} X_i^T \qquad i = 1,2, \ldots, n \quad (10)$$

Suatu data akan dikatakan *pencilan* apabila:

$|DFFITS| > 1$ *untuk n*≤ 30

$|DFFITS| > 2\sqrt{\frac{p}{n}}$ *untuk n > 30*

dengan *p* adalah banyaknya parameter dalam model dan *n* adalah banyaknya pengamatan.

1.6.4.2 Nilai Pengaruh (*Laverage Value)*

Nilai pengaruh digunakan untuk menginformasikan seberapa jauh kasus tersebut dari nilai *mean* himpunan data variabel prediktor. Dengan kata lain, nilai pengaruh dapat mendeteksi pencilan. Nilai pengaruh h_{ii} dari pengamatan menunjukkan besarnya peranan X_i terhadap Y_i dan didefinisikan pada Persamaan (10), dengan $X_i = [X_0, X_1, X_2, \ldots, X_n]$ adalah vektor baris yang berisi nilai-nilai dari peubah variabel bebas dalam pengamatan ke-i. Nilai h_{ii} berada diantara 0 dan 1 dengan $k = d - 1$. Sehingga dapat ditulis menjadi:

$$\frac{2k}{n} = \frac{2(p-1)}{n} \quad (11)$$

Suatu data pengamatan ke-i, diidentifikasi sebagai pencilan apabila nilai $h_{ii} > \frac{2p}{n}$. Sehingga data pengamatan ke-i dikatakan pencilan terhadap X (Nurdin dkk., 2019).

1.6.5 Regresi *Ridge*

Regresi *ridge* dikembangkan untuk menstabilkan nilai koefisien regresi yang 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. Varians tersebut dilakukan dengan menambahkan tetapan bias c yang dalam hal ini adalah λ pada diagonal matriks yang mempengaruhi besarnya koefisien penduga *ridge* dan penduga yang dihasilkan adalah penduga yang bias (Fauzan & Soehardjoepri, 2023).

Meskipun penduga koefisien regresi bias, metode ini dapat mendekati nilai parameter yang sebenarnya. Regresi *ridge* serupa dengan OLS yang meminimumkan jumlah kuadrat resiudal pada pendugaan koefisien regresi. Pada metode ini menambahkan penalti penyusutan untuk mengurangi residual (Kusuma & Wulansari, 2019). Regresi *ridge* meminimumkan jumlah kuadrat galat dengan batasan L2-norm kuadrat dari koefisien regresi. Sehingga estimasi parameter pada model regresi *ridge* adalah (Yanke dkk., 2022):

$$\hat{\beta}_{ridge} = \arg min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} {\beta_j}^2 \right\} \quad (12)$$

Pada Persamaan (12) suku pertama menunjukkan jumlah kuadrat yang sisaan, sedangkan suku kedua menunjukkan penalti penyusutan dalam meminimumkan jumlah kuadrat galat. Nilai estimasi parameter β diperoleh dengan menghitung nilai turunan parsialnya terhadap β, kemudian di samakan dengan nol, sehingga penaksir model regresi *ridge* yang diperoleh adalah (Sulistianingsih dkk., 2023):

$$\boldsymbol{X'y} = (\boldsymbol{X'X} + \boldsymbol{\lambda I})^{-1}\widehat{\boldsymbol{\beta}} \quad (13)$$

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

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ridge}} = (\boldsymbol{X'X} + \boldsymbol{\lambda I})^{-1}\boldsymbol{X'y}\,, \lambda \geq 0 \quad (14)$$

1.6.6 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. Tujuan dari metode ini adalah agar koefisien regresi tersebut mendekati nol atau bahkan sama dengan nol untuk seleksi variabel prediktor. Oleh karena itu metode LASSO dapat melakukan peran sebagai seleksi variabel sekaligus mengatasi multikolinearitas. Adapun persamaan dari metode ini sebagai berikut (Rahmawati & Suratman, 2022):

$$\hat{\beta}_{lasso} = \arg min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \quad (15)$$

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

Fungsi kendala (nilai t) merupakan suatu besaran yang mengontrol besarnya penyusutan koefisien regresi LASSO hingga 0. Hasil *cross validasi* digunakan dalam pemilihan nilai t. Nilai t tersebut dinamakan sebagai parameter tuning dengan ketentuan nilai $t \geq 0$. Misalkan diketahui β_j merupakan penaksir OLS, dengan nilai t_0 didefinisikan $t_0 = \sum_{j=1}^{p} |\beta_j|$, maka (Robbani, 2019).

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

1.6.7 Regresi *Elastic Net*

Metode *elastic net* pertama kali diperkenalkan oleh Zou dan Hastie pada tahun 2005. *Elastic net* adalah metode regresi yang menggabungkan regresi *ridge* dan regresi LASSO . Metode Ini memiliki kemampuan menyusutkan koefisien regresi menjadi tepat nol. Selain itu, metode ini juga mampu melakukan seleksi pada peubah secara simultan dan memilih kelompok peubah yang saling berkorelasi (Handayani & Wachidah, 2023). Selain itu, 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 :

$$L(\lambda_1, \lambda_2, \beta) = |\boldsymbol{y} - \boldsymbol{X\beta}|^2 + \lambda_1|\boldsymbol{\beta}| + \lambda_2|\boldsymbol{\beta}|^2 \tag{16}$$

dengan :

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

$|\boldsymbol{\beta}|^2 = |\boldsymbol{\beta}|'||\boldsymbol{\beta}|$

$\boldsymbol{\beta}' = \boldsymbol{\beta}$ transpose

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 estimasi *elastic net* yang merupakan gabungan dari regresi *ridge* dan regresi LASSO dengan persamaan :

$$\hat{\beta}_{EN} = \arg min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2 \right\} \tag{17}$$

Parameter Regularisasi λ 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:

$$\hat{\beta}_{EN} = \arg min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \left(a \sum_{j=1}^{p} |\beta_j| + (1-a) \sum_{j=1}^{p} {\beta_j}^2 \right) \right\} \tag{18}$$

dengan α adalah parameter gabungan antara regresi *ridge* (α = 0) dan LASSO (α = 1).

Penalti penyusutan pada *elastic net* menjadi penalti penyusutan regresi *ridge* Jika $a = 0$. Sedangkan ketika $a = 1$, maka penalti penyusutan pada *elastic net* menjadi penalti penyusutan LASSO. Pada regularisasi *elastic net* terdapat penyusutan koefisien bersama dari peubah penjelas yang berkorelasi seperti *ridge* dan seleksi peubah seperti LASSO. Jika $\lambda = 0$, maka model *elastic net* berubah menjadi OLS (Sinaga dkk., 2021).

1.6.8 *Cross Validation*

Validasi silang (*Cross Validation*) merupakan metode dalam melakukan validasi model terbaik. Metode ini akan menguji kekuatan model yang dibuat dengan melakukan penyusunan ulang (*resampling*) serta membagi data menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* akan digunakan untuk melatih model sehingga model dapat memahami pola pada data, dan data *testing* akan digunakan sebagai pengujian untuk validasi latihan model (Widyaningsih dkk., 2021). Penentuan kelompok data dilakukan secara acak agar setiap pengamatan memiliki kesempatan yang sama untuk menjadi data validasi (Kusuma dan Wulansari, 2019).

Salah satu metode dari *cross validation* yang sering digunakan adalah *k-fold cross validation*. 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., 2019).

$$CVE_{(k)} = \frac{1}{k}\sum_{i=1}^{k} RMSE_i \tag{19}$$

Metode ini akan membagi data menjadi k bagian (*folds*) yang sama besar. Nilai k dibebaskan kepada peneliti, tetapi disarankan tidak terlalu besar dan tidak terlalu kecil. Nilai k yang terlalu besar akan menghasilkan model yang tidak bias, tetapi dapat membuat variansi menjadi besar sehingga dapat memicu terjadinya overfit. Nilai k yang terlalu kecil akan menghasilkan model yang serupa dengan metode *cross validation* biasa yang hanya membagi data menjadi *train* – *test* saja (dapat memicu terjadinya bias) (Widyaningsih dkk., 2021). Nilai 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).

1.6.9 Regresi *Robust*

Regresi *robust* merupakan metode penting untuk menganalisis data yang dipengaruhi oleh pencilan sehingga dihasilkan model *robust* atau kekar terhadap pencilan (Nurbaroqah dkk., 2022). Menurut Chen (2002) metode-metode estimasi dalam regresi *robust* diantaranya adalah:

1. Estimasi LTS (*least trimmed squares*), merupakan metode yang menggunakan konsep pengepasan metode kuadrat terkecil untuk meminimumkan jumlah kuadrat galat dan memiliki *high breakdown point*.
2. Estimasi S (*Scale*), merupakan salah satu metode penduga regresi *robust* yang memiliki *high breakdown point* yang diperlukan dalam mengidentifikasi *bad observation* untuk membedakan *good leverage point* dan *bad leverage point*. Meskipun memiliki *high breakdown point* yang sama dengan estimasi LTS, tetapi estimasi S memiliki efisiensi yang lebih tinggi.
3. Estimasi M (*Maximum likelihood type*), merupakan metode yang menganalisis data dengan mengasumsikan bahwa sebagian besar yang terdeteksi pencilan pada variabel independen.
4. Estimasi MM (*Method of Moment*), merupakan metode yang menggabungkan antara estimasi S yang memiliki high breakdown point dan estimasi M.

1.6.9.1 *Robust* S Estimasi

Rousseeuw dan Yohai memperkenalkan metode estimasi S pada tahun 1984 yang merupakan metode dengan *breakdown* point tinggi. S-estimasi memiliki kemampuan untuk mencapai *breakdown point* sebesar 50%, yang menunjukkan bahwa setengah dari pencilan dapat dihindari serta memberikan pengaruh baik pada pengamatan lainnya. *Breakdown point* ini dicapai memnggunakan fungsi pembobot *tukey bisquare* dan *tunning constan* yang nilainya sebesar 1,547 (Astuti dkk., 2023). Estimasi S dirumuskan persamaan sebagai berikut:

$$\hat{\beta}_s = \min \beta \; \hat{\sigma}_s(e_1, e_2, \dots, e_n) \tag{20}$$

Dengan menentukan skala minimum estimasi robust $\hat{\sigma}_s$:

$$\min \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{j=0}^{p} X_{ij}\beta_j}{\hat{\sigma}} \right)$$

Dengan:

$$\hat{\sigma}_s = \sqrt{\frac{\sum_{i=1}^{n}(e_i^2) - (\sum_{i=1}^{n} e_i)^2}{n(n-1)}} \tag{21}$$

1.6.9.2 *Robust* M Estimasi

Estimasi M digunakan untuk meminimumkan fungsi objektif (ρ) dari residualnya. Fungsi objektif merupakan fungsi yang digunakan untuk mencari fungsi pembobot pada regresi. Estimasi M memiliki efisiensi penduga yang tinggi yaitu mencapai 95%. Efisiensi yang tinggi akan memberikan hasil estimasi lebih stabil dan konsisten karena kurang dipengaruhi oleh fluktuasi dari data sampel yang menunjukkan bahwa. Persamaan dari estimasi tersebut sebagai berikut (Wardani dkk., 2021):

$$\hat{\beta}_M = min_\beta \sum_{i=1}^{n} \rho \left(y_i - \sum_{j=0}^{p} X_{ij}\beta_j \right) \tag{22}$$

$$min_\beta \sum_{i=1}^{n} \rho(u_i) = min_\beta \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = min_\beta \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{j=0}^{p} X_{ij}\beta_j}{\hat{\sigma}} \right)$$

dengan $\rho(u_i)$ merupakan fungsi simetris dari residual dan $\hat{\sigma}$ adalah skala estimasi robust. Untuk mencari $\hat{\sigma}$ pada regresi robust digunakan persamaan:

$$\hat{\sigma}_{\boldsymbol{M}} = \frac{median\,|e_i - median(e_i)|}{0{,}6745} \tag{23}$$

1.6.9.3 *Robust* MM Estimasi

Robust MM-estimasi adalah estimasi yang menggabungkan estimasi *high breakdown point* (50%) dan efisiensi penduga (mencapai 95%) yang dikenalkan oleh Yohai pada tahun 1987. Langkah pertama dalam estimasi ini adalah mencari *S-estimasi*, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Estimasi S menjamin nilai *breakdown point* tinggi dan estimasi M membuat estimasi mempunyai efisiensi tinggi. *MM-estimasi* didefinisikan (Nurdin dkk., 2019):

$$\hat{\beta}_{MM} = \min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_s}\right) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{p} X_{ij}\beta_j}{\hat{\sigma}}\right) \tag{24}$$

dimana ρ adalah fungsi objektif dan $\hat{\sigma}$ adalah estimasi skala robust

1.6.10 Fungsi Objektif dan Fungsi Pembobot

Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust*. Fungsi pembobot yang digunakan yaitu fungsi pembobot *Tukey Bisquare*. Diberikan suatu fungsi obyektif sebagai berikut (Nurdin dkk., 2019):

$$\rho = \begin{cases} \frac{c^2}{6}\left\{1 - \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2\right\}, & |u_i| \leq c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \tag{25}$$

Untuk fungsi pembobot di gunakan rumus sebagai berikut:

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \tag{26}$$

Permasalahan dalam estimasi regresi robust adalah perlu dilakukan pemilihan *tunning constant* agar estimasi yang diperoleh lebih spesifik dan memimimumkan jumlah kuadrat galat (Nurdin dkk., 2019). Dengan $u_i = \frac{e_i}{\hat{\sigma}}$ dan c adalah *tunning constatnt*. Diketahui nilai *tuning constant* pada metode *tukey bisquare* untuk *S-Estimasi* yaitu 1,547 dan untuk *M-Estimasi* yaitu 4,685 (Astuti dkk., 2023).

1.6.11 Koefisien Determinasi

Koefisien determinasi atau dikenal dengan *R-squared*, adalah salah satu ukuran tentang kecocokan data dengan model. Rumus koefisien determinasi sebagai berikut (Handayani & Wachidah, 2023):

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{27}$$

Semakin dekat R^2 dengan1 makin baik kecocokan data dengan model. Dan sebaliknya makin dekat R^2 dengan 0 makin jelek kecocokan model tersebut.

1.6.12 Pengujian Signifikansi Parameter

1.6.12.1 Uji Simultan

Uji hipotesis yang digunakan untuk pengaruh secara bersama-sama atau secara simultan adalah uji F yang dilakukan dengan membandingkan nilai F hitung dengan F tabel. Pengujian ini dilakukan untuk mengetahui ada atau tidaknya hubungan linier antara variabel respon dengan variabel prediktor. Adapun rumusan hipotesis yang digunakan yaitu:

$H_0 :\ \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1: terdapat $\beta_j \neq 0$, dengan $\mathrm{j} = 1,2,\ldots,p$

Maka statistik uji *F* yaitu:

$$F_{hitung} = \frac{SSR/p}{SSE/(n-p-1)} \tag{28}$$

dengan SSR adalah jumlah kuadrat regresi, SSE adalah jumlah kuadrat residual, p adalah banyaknya parameter independen, dan n adalah jumlah observasi.

Kriteria pengujiannya yaitu apabila $F_{hitung} > F_{tabel}$ maka H_0 ditolak. Cara lain untuk mengambil Keputusan adalah dengan membandingkan nilai $P - value$ dengan a, jika $P - value < a$ maka H_0 di tolak (Siburian dkk., 2019).

1.6.12.2 Uji Parsial

Pengujian ini digunakan untuk menguji ada atau tidaknya pengaruh yang signifikan antara masing-masing variabel prediktor terhadap model regresi linier. Adapun hipotesis yang digunakan yaitu:

$H_0 : \beta_1 = 0,$ variabel X_j tidak berpengaruh signifikan terhadap variabel y.

$H_1 : \beta_1 \neq 0$ variabel X_j tidak berpengaruh signifikan terhadap variabel y

Adapun statistik ujinya yaitu:

$$T_{hitung} = \frac{\beta_j}{s_{\beta j}} \tag{29}$$

dengan

$$s_{\beta j} = \sqrt{var\beta_j} \tag{30}$$

Kriteria ujinya yaitu H_0 ditolak jika $T_{hitung} > T_{tabel}$ atau dengan pengujian $p - value < a$ (Deria, 2019)

1.6.13 Angka Kematian Bayi

Kematian bayi adalah kematian yang terjadi sesaat setelah bayi dilahirkan sampai bayi belum berusia kurang dari satu tahun. Angka Kematian Bayi (AKB) adalah jumlah kematian bayi usia 0-11 bulan yang dinyatakan dalam 1000 kelahiran hidup pada tahun yang sama (Dinas Kesehatan, 2021). Secara matematis dapat ditulis sebagai berikut:

$$\text{AKB} = \frac{\text{Jumlah Kematian Bayi}}{\text{Jumlah Kelahiran Bayi}} \times 1000 \tag{31}$$

Angka kematian bayi merupakan salah satu indikator penting dalam menentukan tingkat kesehatan masyarakat karena dapat menggambarkan kesehatan penduduk secara umum dan termasuk di dalam salah satu target *Sustainable Development Goals* (SDGS) (Wulandari & Utomo, 2021).

Badan Pusat Statistik (BPS) Pada tahun 2020 ada 16,85 kematian bayi per 1.000 kelahiran bayi Indonesia, menurut Badan Pusat Statistik (BPS). Meskipun angka kematian bayi di Indonesia telah menurun setiap tahunnya, angka tersebut masih di bawah standar yang ditentukan. Dibandingkan dengan negara berkembang lainnya seperti Malaysia, yang hanya memiliki 10 kematian per 1.000 kelahiran, AKB Indonesia masih tergolong tinggi (Permata Sari dkk., 2023).

Dari sisi penyebabnya, kematian bayi ada dua macam yaitu kematian bayi endogen dan kematian bayi eksogen. Kematian bayi endogen atau biasa disebut dengan kematian neonatal adalah kematian bayi yang disebabkan oleh faktor faktor yang dibawa anak sejak lahir, yang diperoleh dari orang tuanya selama dalam kandungan dan terjadi pada bulan pertama setelah dilahirkan. Sedangkan kematian bayi eksogen atau kematian post-neonatal adalah kematian bayi yang disebabkan oleh faktor-faktor yang berhubungan dengan pengaruh lingkungan sekitar dan terjadi setelah satu bulan sampai menjelang usia satu tahun (Yanti dkk., 2023).

BAB II

METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari publikasi Dinas Kesehatan Provinsi Sulawesi Selatan Tahun 2020 pada website dinkes.sulselprov.go.id mengenai Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan.

2.2 Variabel Penelitian

Variabel dalam penelitian ini terdiri dari variabel dependen (Y) dan variabel independen (X). Adapun variabel yang digunakan dalam penelitian ini adalah.

Tabel 2. Variabel respon dan variabel prediktor

Variabel	Keterangan	Definisi Operasional
Y	Angka kematian bayi	Jumlah kematian bayi (usia di bawah satu tahun) per 1.000 kelahiran hidup dalam suatu wilayah dan periode tertentu.
X_1	Jumlah pelayanan kesehatan bayi	Jumlah pelayanan kesehatan yang diberikan kepada bayi (usia di bawah satu tahun)
X_2	Jumlah pemberian vitamin A	Cakupan pemberian vitamin A pada bayi usia 6-11 bulan
X_3	Jumlah bayi berat badan lahir rendah	Jumlah bayi lahir dengan berat badan di bawah batas normal (< 2,5 kg)
X_4	Jumlah ibu bersalin ditolong medis	Jumlah ibu yang melahirkan anaknya dengan bantuan tenaga medis

2.3 Tahapan Analisis Data

Dalam penelitian ini pengolahan data menggunakan *software* R Studio versi 4.2.1 dalam menyelesaikan proses perhitungan. Adapun tahapan analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mengidentifikasi adanya multikolinearitas pada data dengan menghitung nilai VIF dan nilai korelasi sesuai dengan persamaan (6) dan (7).

$$VIF_j = \frac{1}{1 - R_j^2}$$

$$r_{jk} = \frac{1}{n-2}\sum_{r=1}^{n}\left(\frac{x_{jk} - \bar{X}_j}{\sqrt{S_{jj}}}\right)\left(\frac{x_{jk} - \bar{X}_k}{\sqrt{S_{kk}}}\right)$$

2. Mengidentifikasi adanya pencilan (outlier) pada data menggunakan metode $DFFITS$ sesuai persamaan (8).

$$DFFITS = t_i\left(\frac{h_{ii}}{1 - h_{ii}}\right)^{\frac{1}{2}}$$

3. Mengestimasi parameter dengan metode *elastic net*. Adapun tahapan estimasinya yaitu:
   a) Melakukan standarisasi data pada variabel respon dan variabel prediktor

b) Membagi data menjadi 2 bagian, 80% data training dan 20% data testing. Data training berisi variabel respon (Y) dan variabel prediktor (X) yang akan digunakan untuk membuat model regresi. Data testing berisi variabel prediktor (X) yang akan digunakan untuk melakukan prediksi pada model regresi yang telah dibuat.

c) Membagi data menjadi k subset atau 5-*fold* yang sama besar pada dataset *training*.

d) Menentukan nilai alpha dan lambda yang paling optimum yang akan digunakan dalam model *elastic net* melalui proses *cross validation* pada data *training*.

e) Menghitung dugaan galat *cross validation* dari semua subet untuk model *elastic net* terbaik. Pemilihan dilakukan dengan memilih dugaan galat terendah. Dugaan galat validasi silang (*cross validation error*-CVE) dapat diperoleh dengan menghitung rata-rata dari persamaan (19)

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

f) Melakukan estimasi parameter model regresi *elastic net* sesuai persamaan (18)

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

4. Memodelkan angka kematian bayi dengan metode regresi *elastic net robust* (estimasi MM) melalui tahapan berikut :

a. Mengestimasi parameter dengan metode robust S-Estimasi

a) Menghitung nilai residual $e_i = y_i - \hat{y}_i$

b) Menghitung $\hat{\sigma}_s$ dengan menggunakan persamaan (21)

$$\hat{\sigma}_s = \sqrt{\frac{\sum_{i=1}^{n}(e_i^2) - (\sum_{i=1}^{n} e_i)^2}{n(n-1)}}$$

c) Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_s}$

d) Menghitung pembobot awal w_i menggunakan pembobbot *tukey bisquare* sesuai dengan persamaan (26)

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |u_i| \leq 1.547 \\ 0, & |u_i| > 1.547 \end{cases}$$

e) Mengestimasi koefisien $\hat{\beta}_S$ menggunakan pembobot w_i.

f) Menghitung bobot baru w_i dengan skala estimas dari iterasi awal.

g) Mengulangi langkah a) hingga g) sampai nilai $\hat{\beta}_S$ konvergen

b. Menghitung nilai residual $e_i = y_i - \hat{y}_i$ dari $\hat{\beta}_S$

c. Menghitung $\hat{\sigma}_M$ dengan menggunakan persamaan (23)

$$\hat{\sigma}_{\boldsymbol{M}} = \frac{median\ |e_i - median(e_i)|}{0{,}6745}$$

d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_s}$

e. Menghitung pembobot awal w_i menggunakan pembobbot *tukey bisquare* sesuai dengan persamaan (26)

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |u_i| \leq 4.685 \\ 0, & |u_i| > 4.685 \end{cases}$$

f. Mengestimasi koefisien $\hat{\beta}_{MM}$ menggunakan pembobot w_i.

g. Mengulangi langkah b hingga g sampai nilai $\hat{\beta}_{MM}$ konvergen

h. Menghitung parameter regresi *elastic net robust* dengan persamaan

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ENMM}} = ((\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}) + \widehat{\boldsymbol{\beta}}_{\boldsymbol{EN}}$$

5. Melakukan uji signifikansi parameter model regresi *elastic net robust* dengan menggunakan uji F sesuai persamaan (28) dan uji t sesuai persamaan (29)

$$F_{hitung} = \frac{SSR/p}{SSE/(n-p-1)}, T_{hitung} = \frac{\beta_j}{s_{\beta j}}$$

6. Menghitung koefisien determinasi sesuai persamaan (27)

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

7. Menentukan ukuran kebaikan model yang diperoleh berdasarkan nilai MAE dan MAPE.

8. Menginterpretasikan model yang telah di peroleh dan menarik kesimpulan.