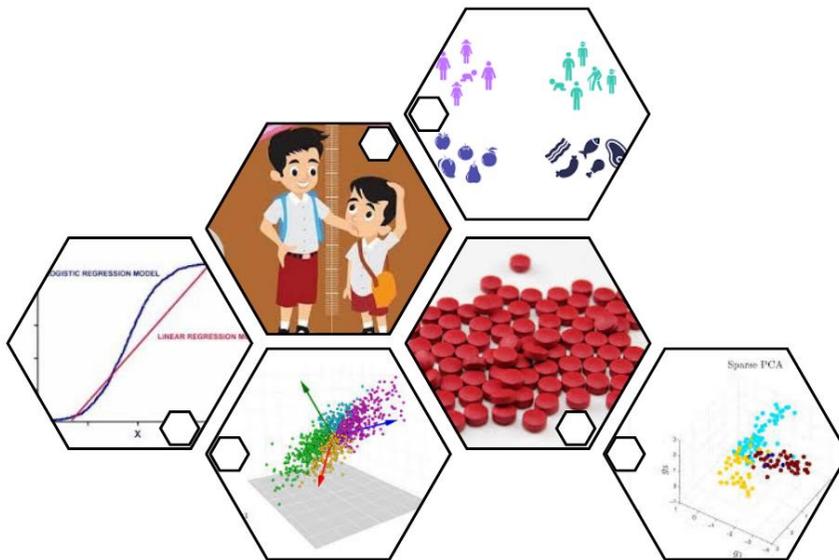


**MODEL REGRESI LOGISTIK *SPARSE PRINCIPAL COMPONENT*
ANALYSIS PADA PREDIKTOR KATEGORIK**

**REGRESSION LOGISTIC *SPARSE PRINCIPAL COMPONENT* ANALYSIS
ON CATEGORY PREDICTORS MODEL**



DWI AULIYAH

H062212010



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

**MODEL REGRESI LOGISTIK SPARSE PRINCIPAL COMPONENT
ANALYSIS PADA PREDIKTOR KATEGORIK**

DWI AULIYAH

H062212010



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

**MODEL REGRESI LOGISTIK *SPARSE PRINCIPAL COMPONENT*
ANALYSIS PADA PREDIKTOR KATEGORIK**

Tesis
sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

DWI AULIYAH
H062212010

kepada

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

TESIS

**MODEL REGRESI LOGISTIK *SPARSE PRINCIPAL COMPONENT ANALYSIS*
PADA PREDIKTOR KATEGORIK**

**DWI AULIYAH
H062212010**

telah dipertahankan di depan Panitia Ujian Magister pada 16 Agustus
2024

dan dinyatakan telah memenuhi syarat kelulusan

pada

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

Mengesahkan:

Pembimbing Utama



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

Pembimbing Pendamping



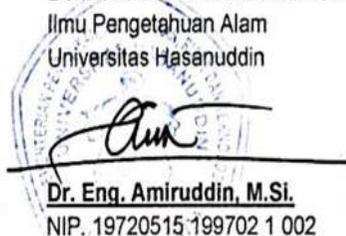
Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.
NIP. 196209261987022001

Ketua Program Studi
Magister Statistika



Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam
Universitas Hasanuddin



Dr. Eng. Amiruddin, M.Si.
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul " Model Regresi Logistik *Sparse Principal Component Analysis* Pada Prediktor Kategorik" adalah benar karya saya dengan arahan dari tim pembimbing (Dr. Anna Islamiyati, S.Si., M.Si. dan Prof. Dr. Dr Georgina Maria Tinungki M.Si.). Karya Ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. 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 tesis ini. Sebagian dari tesis ini akan dipublikasikan di *Journal of Sciences, Islamic Republic of Iran* sebagai artikel dengan judul "*Regression Logistic Analysis Model with Categorical Principal Component Analysis in Stunting Cases*".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 16 Agustus 2024




DWI AULIYAH
NIM. H062212010

UCAPAN TERIMA KASIH

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas limpahan rahmat dan hidayah-Nya kepada penulis. Shalawat dan salam tercurahkan kepada Rasulullah *Shallallahu 'Alaihi Wa sallam*, keluarganya, *tabi'in, tabi'ut tabi'in*, serta orang-orang sholeh yang haq hingga kadar Allah berlaku atas diri mereka. *Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan dari Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tesis berjudul "Model Regresi Logistik *Sparse Principal Component Analysis* Pada Prediktor Kategorik" sebagai salah satu syarat memperoleh gelar magister pada Program Studi Magister Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.

Terima Kasih yang tak terhingga kepada kedua orang tuaku Bapak **Demi Yazis** dan Ibu **Yetti Silviyandri** yang selalu ada memberikan kasih sayang tak terhingga, semangat, motivasi dan doa yang tak pernah putus, juga kepada saudara-saudaraku, kakak **Hayyu Wulandari** dan adikku **Yuki Hilmi Yazis** serta kemenakan kecilku **Latheef Tsaqib Muzammil** yang selalu memberikan dukungan dan semangat dalam penyelesaian tesis ini. Ucapan rasa hormat dan juga terima kasih yang tulus kepada:

1. Yth. **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin.
2. Yth. **Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruhnya jajarannya.
3. Yth. **Dr. Anna Islamiyati, S.Si., M.Si.** selaku Ketua Departemen Statistika sekaligus pembimbing utama penulis yang senantiasa sangat baik dan ikhlas meluangkan waktu dalam memberikan arahan, motivasi, pengetahuan, dan bimbingan untuk penulis dalam menyelesaikan tesis.
4. Yth. **Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku Ketua Program Studi Magister Statistika sekaligus sebagai penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis.
5. Yth. **Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.** sebagai pembimbing yang senantiasa sangat baik meluangkan waktu dalam membimbing dan mengarahkan penulis dalam menyelesaikan tesis.
6. Yth. **Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis.
7. Yth. **Dr. Nirwan Ilyas, M.Si.** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis.
8. Sahabat terbaik penulis **Cici Pulcerima** dan **Yeni Novita** terima kasih telah dengan tulus selalu ada untuk mendukung penulis agar terus berjuang untuk setiap proses penulis termasuk dalam penyelesaian tesis ini.
9. Sahabatku **Dila, Haura, Gading, Lia, Hana, Delawahyuni, Ali Sadikin, Aslan Nonti** dan **Jasrin Padindi** terima kasih selalu memberi semangat dan bersedia selalu menemani setiap proses penulis.
10. Teman-teman seperjuangan **Kak Muhammad Fadil, Kak Mahrani, Kak Irwan Usman, Kak Muammar Ashari** dan teman-teman ruang diskusi terima kasih sudah selalu memberikan semangat dan bantuan dalam proses penyelesaian tesis ini.
11. Semua pihak yang telah membantu penulis yang tidak bisa disebutkan satu per satu terimakasih atas doa serta dukungannya.

Semoga Tuhan Yang Maha Esa memberikan balasan yang berlipat ganda, kasih dan hikmat-Nya atas segala kebaikan yang telah diberikan kepada penulis. Penulis menyadari bahwa masih banyak kekurangan dalam tesis ini, untuk itu dengan segala kerendahan hati, penulis memohon maaf.

Makassar, 16 Agustus 2024

Dwi Auliyah

ABSTRAK

DWI AULIYAH. **Model Regresi Logistik *Sparse Principal Component Analysis* Pada Prediktor Kategorik** (dibimbing oleh Dr. Anna Islamiyati, S.Si., M.Si. dan Prof Dr. Dr. Georgina Maria Tinungki, M.Si.).

Latar Belakang. Analisis regresi logistik biner digunakan untuk mengetahui hubungan antara variabel prediktor dengan variabel respon yang mengandung dua kategorik. Variabel prediktor dengan skala kategorik yang memiliki masalah multikolinieritas diatasi dengan metode *categorical principal component analysis*. Komponen utama yang terbentuk dengan metode *categorical principal component analysis* memiliki kombinasi linear dari setiap variabel asli, sehingga sering sulit untuk diinterpretasikan. *Sparse principal component analysis* diperlukan untuk menghasilkan komponen utama yang *sparse*. Parameter model *sparse categorical principal component logistic regression* diestimasi dengan metode maksimum likelihood dan metode iterasi Newton Raphson. Penelitian ini diterapkan pada data kejadian stunting di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara. **Tujuan.** Penelitian ini bertujuan untuk memperoleh model hubungan antara kondisi pertumbuhan balita (stunting) di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara dengan faktor risiko stunting berdasarkan metode regresi logistik *sparse principal component analysis* pada prediktor kategorik. **Metode.** Adapun metode yang digunakan pada penelitian ini adalah regresi logistik *sparse principal component analysis* pada prediktor kategorik. **Hasil.** Hasil penelitian ini diperoleh model regresi logistik *sparse principal component analysis* pada prediktor kategorik yang bersesuaian dengan data kejadian stunting di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara. **Kesimpulan.** Dari hasil penelitian, diperoleh hasil estimasi parameter model regresi logistik *sparse principal component analysis* pada prediktor kategorik.

Kata Kunci: Regresi Logistik; *Principal Component Analysis*; *Sparse Principal Component Analysis*; Prediktor Kategorik; Stunting

ABSTRACT

DWI AULIYAH. **Regression Logistic Sparse Principal Component Analysis On Category Predictors Model** (supervised by Dr. Anna Islamiyati, S.Si., M.Si. and Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.).

Background. Binary logistic regression analysis is used to determine the relationship between predictor variables and response variables containing two categories. Predictor variables with a categorical scale that have multicollinearity problems are overcome by the categorical principal component analysis method. The principal components formed by the categorical principal component analysis method have a linear combination of each original variable, so they are often difficult to interpret. Sparse principal component analysis is needed to produce sparse principal components. The parameters of the sparse categorical principal component logistic regression model are estimated using the maximum likelihood method and the Newton Raphson iteration method. This study was applied to stunting incident data in Maginti District, West Muna Regency, Southeast Sulawesi Province. **Purpose.** This study aims to obtain a model of the relationship between toddler growth conditions (stunting) in Maginti District, West Muna Regency, Southeast Sulawesi Province with stunting risk factors based on the sparse principal component analysis logistic regression method on categorical predictors. **Method.** The method used in this study is sparse principal component analysis logistic regression on categorical predictors. **Results.** The results of this study obtained a sparse principal component analysis logistic regression model on categorical predictors that correspond to stunting incidence data in Maginti District, West Muna Regency, Southeast Sulawesi Province. **Conclusion.** From the results of the study, the results of the parameter estimation of the sparse principal component analysis logistic regression model on categorical predictors were obtained.

Keywords: Logistic Regression; Principal Component Analysis; Sparse Principal Component Analysis; Category Predictor; Stunting

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	ii
PERNYATAAN PENGAJUAN.....	iii
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
PERNYATAAN KEASLIAN TESIS.....	v
UCAPAN TERIMA KASIH.....	iii
ABSTRAK	v
ABSTRACT.....	vi
DAFTAR ISI.....	iii
DAFTAR TABEL	v
DAFTAR LAMPIRAN	vi
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Kajian Teori.....	3
1.6.1 Multikolinieritas.....	3
1.6.2 Vektor Eigen dan Nilai Eigen	3
1.6.3 <i>Principal Component Analysis</i>	4
1.6.4 <i>Sparse Principal Component Analysis</i>	4
1.6.5 <i>Categorical Principal Component Analysis</i>	6
1.6.6 Regresi Logistik Biner	7
1.6.7 <i>Principal Component Logistic Regression</i>	7
1.6.8 Newton Raphson.....	8
1.6.9 Pengujian Parameter	9
1.6.9 Stunting Pada Balita.....	9
1.6.10 Kerangka Konseptual.....	10
BAB II. METODE PENELITIAN.....	11
2.1 Sumber Data dan Variabel Penelitian	11
2.2 Definisi Operasional Variabel.....	11
2.3 Metode Analisis.....	13
BAB III. HASIL DAN PEMBAHASAN	15

3.1 Estimasi Parameter Model Regresi Logistik <i>Sparse Principal Component</i> <i>Analysis</i> Pada Prediktor Kategorik.....	15
3.2 Model Hubungan Antara Kejadian Stunting Di Kecamatan Maginti Dengan Faktor Resikonya Berdasarkan Metode Regresi Logistik <i>Sparse Principal Component</i> <i>Analysis</i> Pada Prediktor Kategorik.....	19
3.2.1 Deskripsi Data.....	19
3.2.2 Menentukan Variabel X Optimal	24
3.2.3 Identifikasi Matriks Korelasi.....	26
3.2.4 Menghitung Nilai Eigen dan Vektor Eigen.....	26
3.2.5 Membentuk Komponen Utama.....	27
3.2.6 Menentukan Nilai Loading yang Sparse.....	28
3.2.7 Menentukan Model Regresi Logistik Biner <i>Sparse Principal Component</i> <i>Analysis</i> Pada Prediktor Kategorik	30
BAB IV. PENUTUP	34
4.1 Kesimpulan	34
4.2 Saran	34
DAFTAR PUSTAKA.....	35
LAMPIRAN	37

DAFTAR TABEL

Tabel 1. Variabel Penelitian.....	11
Tabel 2. Definisi Operasional Variabel	11
Tabel 3. Sebaran Frekuensi Kejadian Stunting	19
Tabel 4. Sebaran Frekuensi Jenis Kelamin Berdasarkan Kejadian Stunting.....	19
Tabel 5. Sebaran Frekuensi Tinggi Badan Bayi Lahir Berdasarkan Kejadian Stunting	20
Tabel 6. Sebaran Frekuensi Berat Badan Bayi Lahir Berdasarkan Kejadian Stunting	20
Tabel 7. Sebaran Frekuensi Memiliki Riwayat Penyakit Berdasarkan Kejadian Stunting	20
Tabel 8. Sebaran Frekuensi Usia Pemberian MPASI Berdasarkan Kejadian Stunting.	21
Tabel 9. Sebaran Frekuensi Mengonsumsi Susu Formula Berdasarkan Kejadian Stunting	21
Tabel 10. Sebaran Frekuensi Ibu Hamil Mendapatkan PMT Berdasarkan Kejadian Stunting	21
Tabel 11. Sebaran Frekuensi Ibu Hamil Mendapatkan TTD Berdasarkan Kejadian Stunting	22
Tabel 12. Sebaran Frekuensi Usia Ayah Berdasarkan Kejadian Stunting.....	22
Tabel 13. Sebaran Frekuensi Usia Ibu Berdasarkan Kejadian Stunting	23
Tabel 14. Sebaran Frekuensi Pendidikan Terakhir Ayah Berdasarkan Kejadian Stunting	23
Tabel 15. Sebaran Frekuensi Pendidikan Terakhir Ibu Berdasarkan Kejadian Stunting	24
Tabel 16. Sebaran Frekuensi Air Minum Yang Digunakan Berdasarkan Kejadian Stunting	24
Tabel 17. Kuantifikasi Kategori Optimal C_j^* Untuk Setiap Kategori Pada X^*	25
Table 18. Proporsi Total Varians Komponen Utama	27
Table 19. Nilai Loading Optimal.....	28
Table 20. Nilai Proporsi Varians Kumulatif Berdasarkan Nilai ℓ_1 dan ℓ_2	28
Tabel 21. Nilai Loading <i>Sparse</i> Optimal	29
Tabel 22. Hasil Estimasi Regresi Logistik <i>Sparse Principal Component Analysis</i> Pada Prediktor Kategorik	30
Tabel 23. Uji Rasio <i>Likelihood</i>	31

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian Kejadian Stunting di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara	37
Lampiran 2. Variabel Prediktor Hasil Kuantifikasi Optimal.....	38
Lampiran 3. Nilai Komponen Utama Sparse Optimal	40

BAB I

PENDAHULUAN

1.1 Latar Belakang

Regresi logistik merupakan analisis regresi yang bertujuan untuk memprediksi suatu peubah respon kategorik berdasarkan satu atau beberapa peubah prediktor. Analisis regresi logistik biner merupakan salah satu analisis regresi logistik yang digunakan untuk menganalisis hubungan antara peubah prediktor dengan peubah respon berskala dikotomi. Skala dikotomi adalah skala data nominal dengan dua kategori (Schober & Vetter, 2019). Salah satu asumsi dalam analisis regresi logistik adalah tidak terjadi multikolinieritas pada peubah prediktor (Bayman & Dexter, 2021).

Multikolinieritas yang terjadi pada peubah prediktor adalah terdapat hubungan yang kuat atau korelasi yang kuat diantara beberapa atau seluruh peubah prediktor. Analisis regresi logistik memiliki kelemahan dalam prediksi pada saat terjadi multikolinieritas pada peubah prediktor. Jika ada multikolinieritas diantara peubah prediktor, maka koefisien regresi yang dihasilkan dalam suatu analisis menjadi sangat lemah sehingga tidak dapat memberikan hasil yang mewakili sifat atau pengaruh dari peubah prediktor. Estimasi parameter model regresi logistik tidak akurat dan interpretasi *odds* rasio dapat keliru ketika ada multikolinieritas diantara peubah prediktor (Bayman & Dexter, 2021). Metode yang dapat digunakan pada peubah prediktor untuk mengatasi multikolinieritas tanpa harus mengeluarkan peubah prediktor yang saling berkorelasi tinggi yaitu dengan metode *principal component analysis* (Agarwal et al., 2021).

Metode *principal component analysis* bertujuan mereduksi dimensi untuk mengatasi masalah multikolinieritas pada data. Pada dasarnya *principal component analysis* mentransformasi secara linier variabel prediktor yang umumnya saling berkorelasi menjadi sejumlah variabel yang lebih sedikit dan tidak saling berkorelasi yang disebut komponen utama. Setelah terbentuk beberapa komponen hasil *principal component analysis* yang bebas multikolinieritas, komponen-komponen tersebut menjadi variabel prediktor baru yang selanjutnya diregresikan pengaruhnya terhadap variabel respon (Agarwal et al., 2021).

Pan et al. (2024) telah menggabungkan *principal component analysis* dan regresi logistik untuk mengatasi multikolinieritas pada regresi logistik. Islamiyati (2015) telah mengembangkan *principal component analysis* untuk data berskala campuran pada regresi logistik biner. Penelitian-penelitian tersebut melibatkan variabel prediktor yang berskala kuantitatif dan campuran. Namun pada data riil, sering ditemukan data dengan prediktor yang berjenis kategorik, misalnya data kuesioner. Kemalbay dan Korkmazoglu (2014) telah mengenalkan metode *categorical principal component analysis* untuk data prediktor kategorik pada regresi logistik.

Principal component analysis juga memiliki kekurangan yaitu setiap komponen utama yang terbentuk merupakan kombinasi linier dari semua variabel sehingga menimbulkan nilai *loading* yang dihasilkan tidak nol untuk setiap komponen utama. Hal ini dapat menyebabkan hasil komponen utama yang diperoleh akan sulit untuk diinterpretasikan. Salah satu perkembangan metode *principal component analysis* adalah *sparse principal component analysis*. Metode *sparse principal component*

analysis yang menggabungkan kekuatan *principal component analysis* klasik, reduksi data, dan pemodelan *sparseness* mengeluarkan variabel tidak efektif dari setiap komponen utama pada model *principal component analysis* yang terbentuk dengan mengecilkan beban variabel tersebut menjadi nol (Erichson et al., 2020).

Berdasarkan uraian tersebut, penulis menyusunnya dalam sebuah penelitian pada data kondisi pertumbuhan balita (Kejadian Stunting) di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara. Stunting adalah masalah kurang gizi kronis yang disebabkan oleh asupan gizi yang kurang dalam waktu cukup lama akibat pemberian makanan yang tidak sesuai dengan kebutuhan gizi. Masalah stunting merupakan masalah kesehatan masyarakat yang berhubungan dengan meningkatnya risiko kesakitan, kematian, dan hambatan pada pertumbuhan baik motorik maupun mental. Hasil Survei Status Gizi Indonesia (SSGI) prevalensi stunting di Indonesia pada tahun 2022 adalah 21,6%. Berdasarkan data dari Dinas Kesehatan Kabupaten Muna Barat Tahun 2022, jumlah balita stunting di Kecamatan Maginti mencapai sebanyak 45 jiwa dari total jumlah balita stunting di Kabupaten Muna Barat sebanyak 201 jiwa. Faktor resiko yang dapat menyebabkan kondisi stunting diantaranya, jenis kelamin, tinggi badan bayi lahir, berat badan bayi lahir, memiliki riwayat penyakit, usia pemberian MPASI, mengonsumsi susu formula, Ibu hamil mendapatkan Pemberian Makanan Tambahan (PMT), Ibu mendapatkan Tablet Tambah Darah (TTD), usia Ayah, usia Ibu, Pendidikan terakhir Ayah, Pendidikan terakhir Ibu, dan air minum yang dikonsumsi (Dayuningsih dan Tria, 2020). Faktor resiko kejadian stunting ini dapat digunakan sebagai peubah prediktor pada penerapan model regresi logistik *sparse principal component analysis* pada prediktor kategorik dengan peubah respon dalam penelitian ini yaitu kondisi pertumbuhan pada balita yang terdiri dari dua kategori yaitu 1 (mewakili balita yang mengalami kondisi stunting) dan 0 (mewakili balita yang tidak mengalami kondisi stunting).

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini berdasarkan latar belakang yang telah diuraikan adalah:

1. Bagaimana estimasi parameter model regresi logistik *sparse principal component analysis* pada prediktor kategorik?
2. Bagaimana model hubungan antara kondisi pertumbuhan balita (*stunting*) di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara dengan faktor resiko stunting berdasarkan metode regresi logistik *sparse principal component analysis* pada prediktor kategorik?

1.3 Batasan Masalah

Data yang digunakan pada penelitian ini adalah data kondisi pertumbuhan balita di Kecamatan Maginti Kabupaten Muna Barat, Provinsi Sulawesi Tenggara dengan variabel prediktor yang meliputi jenis kelamin, tinggi badan bayi lahir, berat badan bayi lahir, memiliki riwayat penyakit, usia pemberian MPASI, mengonsumsi susu formula, Ibu hamil mendapatkan Pemberian Makanan Tambahan (PMT), Ibu mendapatkan Tablet

Tambah Darah (TTD), usia Ayah, usia Ibu, Pendidikan terakhir Ayah, Pendidikan terakhir Ibu, dan air minum yang dikonsumsi. Selain itu, proporsi kumulatif keragaman total yang mampu dijelaskan oleh komponen-komponen utama yang dipilih dari hasil metode *sparse principal component analysis* pada prediktor kategorik yaitu minimal 80%.

1.4 Tujuan Penelitian

Berikut tujuan penelitian berdasarkan rumusan masalah yang telah diuraikan:

1. Memperoleh estimasi parameter model regresi logistik *sparse principal component analysis* pada prediktor kategorik.
2. Memperoleh model hubungan antara kondisi pertumbuhan balita (*stunting*) di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara dengan faktor resiko *stunting* berdasarkan metode regresi logistik *sparse principal component analysis* pada prediktor kategorik.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Penulis memperoleh pemahaman yang lebih baik mengenai model regresi logistik *sparse principal component analysis* pada prediktor kategorik.
2. Pembaca memperoleh informasi mengenai model regresi logistik *sparse principal component analysis* pada prediktor kategorik yang bersesuaian dengan data kejadian *stunting* di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara.

1.6 Kajian Teori

1.6.1 Multikolinieritas

Multikolinieritas merupakan suatu kondisi yang terjadi ketika terdapat korelasi diantara variabel prediktor atau dapat dikatakan antar variabel prediktor tidak bersifat saling bebas. Mendeteksi adanya kasus multikolinieritas di dalam model regresi dapat dilakukan dengan cara menganalisis matriks korelasi (Yanke et al., 2022). Mendeteksi adanya multikolinieritas pada variabel prediktor dilakukan dengan menghitung nilai koefisien korelasi sederhana (*simple correlation*) antar variabel prediktor berdasarkan matriks korelasi R sebagai berikut:

$$R = \frac{1}{n-1} X^T X$$

dengan X adalah variabel prediktor yang telah distandarisasi dan n adalah jumlah data. Jika terdapat koefisien korelasi sederhana yang mencapai atau melebihi 0,8 dan mencapai atau kurang dari $-0,8$ maka hal tersebut menunjukkan terjadinya masalah multikolinieritas dalam regresi (Ridwan & Sunendiari, 2021).

1.6.2 Vektor Eigen dan Nilai Eigen

Jika R adalah sebuah matriks berukuran $n \times n$, maka terdapat suatu skalar λ vektor tak nol a sehingga memenuhi persamaan sebagai berikut (Nasution et al., 2019):

$$\begin{aligned} Ra &= \lambda a \\ Ra - \lambda a &= 0 \end{aligned}$$

Skalar λ disebut nilai eigen dari R dan \mathbf{a} disebut sebagai vektor eigen dari R yang bersesuaian dengan λ . Vektor eigen \mathbf{a} menyatakan matriks kolom yang apabila dikalikan dengan sebuah matriks $n \times n$ menghasilkan vektor lain yang merupakan kelipatan vektor itu sendiri. Operasi $R\mathbf{a} = \lambda\mathbf{a}$ menyebabkan vektor \mathbf{a} menyusut atau memanjang dengan faktor λ dengan arah yang sama jika λ positif dan arah berkebalikan jika λ negative.

Untuk memperoleh nilai eigen matriks R yang berukuran $n \times n$, maka $R\mathbf{a} = \lambda\mathbf{a}$ dapat ditulis sebagai $R\mathbf{a} = \lambda\mathbf{I}\mathbf{a}$ atau $(R - \lambda\mathbf{I})\mathbf{a} = 0$. Agar λ menjadi nilai eigen, maka harus ada pemecahan tak nol dari persamaan $(R - \lambda\mathbf{I})\mathbf{a} = 0$. Persamaan $(R - \lambda\mathbf{I})\mathbf{a} = 0$ akan mempunyai persamaan tak nol jika dan hanya jika (Nasution et al., 2019):

$$|R - \lambda\mathbf{I}| = 0$$

1.6.3 Principal Component Analysis

Principal component analysis adalah suatu teknik statistik yang secara linear mengubah bentuk sekumpulan variabel asli menjadi kumpulan variabel yang lebih kecil dan tidak saling berkorelasi yang dapat mewakili informasi dari kumpulan variabel asli. *Principal component analysis* adalah teknik analisis statistik untuk mentransformasi peubah-peubah asli yang masih saling berkorelasi satu dengan yang lain menjadi satu set peubah baru yang tidak berkorelasi lagi. Peubah-peubah baru tersebut disebut sebagai Komponen Utama (Greenacre et al., 2022).

Principal component analysis menjelaskan bagian dari variasi dalam kumpulan variabel yang diamati atas dasar beberapa dimensi. Tujuan khusus *principal component analysis* yaitu untuk menghilangkan multikolinieritas antar variabel prediktor dan mereduksi sejumlah besar variabel menjadi sejumlah kecil faktor. Reduksi data pengamatan menggunakan *principal component analysis* dapat dilakukan tanpa mengurangi informasi dari semua data. Oleh karena itu, *principal component analysis* dipandang sebagai transformasi dari X_1, X_2, \dots, X_p (Pan et al., 2024).

Misal R merupakan matriks korelasi dari variabel-variabel $X = [X_1, X_2, \dots, X_p]$ dengan pasangan nilai eigen dan vektor eigen yaitu $(\lambda_1, \mathbf{a}_1), (\lambda_2, \mathbf{a}_2), \dots, (\lambda_p, \mathbf{a}_p)$ dengan $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Komponen utama yang dibentuk sebagai kombinasi linier dapat didefinisikan pada persamaan berikut (Greenacre et al., 2022):

$$\begin{aligned} Z_1 &= \mathbf{a}'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Z_2 &= \mathbf{a}'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Z_r &= \mathbf{a}'_r X = a_{r1}X_1 + a_{r2}X_2 + \dots + a_{rp}X_p \end{aligned}$$

atau

$$\mathbf{a}'_r X = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & a_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \dots & a_{rp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

dengan r adalah jumlah komponen utama yang terbentuk.

1.6.4 Sparse Principal Component Analysis

Salah satu bentuk pengembangan terbaru dari *principal component analysis* adalah *sparse principal component analysis*. *Sparse principal component analysis*

menggabungkan kelebihan *principal component analysis* klasik, reduksi data, dengan pemodelan *sparseness*, yang mengeluarkan variabel yang tidak efektif dari model *principal component analysis* dengan mengecilkan nilai loading dari variabel-variabel prediktor menjadi nol. *Sparse principal component analysis* memiliki kelebihan dalam membuat interpretasi komponen utama menjadi lebih mudah (Erichson et al., 2020).

Zou et al., (2006) memperkenalkan *sparse principal component analysis* menggunakan metode *elastic net* yang dikembangkan dari metode *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk menghasilkan komponen utama yang dimodifikasi dari nilai-nilai loading yang *sparse*. Zou et al., (2006) mengemukakan metode *principal component analysis* dapat diformulasikan sebagai masalah optimasi pada regresi, sehingga nilai-nilai *loading* dapat diperoleh dengan menerapkan batasan *elastic net* pada koefisien regresi δ .

Elastic net merupakan suatu metode seleksi dengan menggabungkan batasan L_1 -norm dan L_2 -norm kuadrat pada δ . Batasan L_1 -norm dapat menghasilkan model yang lebih sederhana karena terjadi penyusutan beberapa δ yang tepat nol dan batasan L_2 -norm kuadrat menghasilkan model yang tidak menyeleksi variabel namun meningkatkan efek pengelompokan dan penyusutan δ . Kombinasi batasan L_1 -norm dan L_2 -norm kuadrat memberikan hasil beberapa koefisien regresi tepat nol tetapi tidak sebanyak dengan hanya menggunakan batasan L_1 -norm dan persentase varians yang lebih tinggi (Erichson et al., 2020). Keuntungan yang lain yang ditawarkan oleh metode *elastic net* adalah dari segi pengelompokan variabel yaitu *elastic net* cenderung memilih sekelompok variabel yang berkolerasi tinggi (Zou et al., 2006).

Zou et al., (2006) menyusun suatu algoritma untuk meminimumkan kriteria *sparse principal component analysis*. Misal $\hat{\delta} = [\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_k]$ merupakan matriks berukuran $p \times k$, p adalah jumlah variabel bebas dan k adalah jumlah komponen utama yang terpilih, $\hat{\delta}_k$ adalah suatu vektor penduga *naive elastic net*, Z_k adalah komponen utama yang terbentuk, dan A_k adalah nilai loading untuk setiap komponen utama berdasarkan metode *principal component analysis*, serta $Z_k = XA_k$ untuk setiap $k = 1, \dots, r$, sehingga diperoleh Persamaan berikut.

$$\begin{aligned}\hat{\delta}_k &= \arg \min_{\delta_k} \{ \|Z_k - X\delta_k\|^2 + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1 \} \\ \hat{\delta}_k &= \arg \min_{\delta_k} \{ \|XA_k - X\delta_k\|^2 + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1 \} \\ \hat{\delta}_k &= \arg \min_{\delta_k} \{ \|X(A_k - \delta_k)\|^2 + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1 \} \\ \hat{\delta}_k &= \arg \min_{\delta_k} \{ (\|X(A_k - \delta_k)\|)(\|X(A_k - \delta_k)\|) + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1 \} \\ \hat{\delta}_k &= \arg \min_{\delta_k} \{ (A_k - \delta_k)^T (\|X\| \|X\|) (A_k - \delta_k) + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1 \}\end{aligned}$$

Dengan mensubstitusikan $\|X\| = \sqrt{X^T X}$, maka diperoleh Persamaan berikut

$$\hat{\delta}_k = \arg \min_{\delta_k} \{ (A_k - \delta_k)^T X^T X (A_k - \delta_k) + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1 \} \quad (1)$$

dengan $\|\delta_k\|_2^2 = \sum_{i=1}^p \delta_{ik}^2$ dan $\|\delta_k\|_1 = \sum_{i=1}^p |\delta_{ik}|$.

Zou et al. (2006) merumuskan algoritma untuk *sparse principal component analysis* pada data dengan $n > p$ yang disebut sebagai algoritma *general sparse*

principal component analysis untuk mereduksi dimensi data menggunakan penduga *naive elastic net* dengan langkah-langkah sebagai berikut.

1. Misalkan $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_r]$, merupakan nilai-nilai *loading* dari setiap komponen utama berdasarkan metode *principal component analysis*.
2. Penduga *naive elastic net* untuk $k = 1, 2, \dots, r$ dihitung dengan persamaan sebagai berikut.

$$\hat{\delta}_k = \arg \min_{\delta_k} \{(A_k - \delta_k)^T X^T X (A_k - \delta_k) + \ell_2 \|\delta_k\|_2^2 + \ell_1 \|\delta_k\|_1\}$$

3. Untuk setiap δ yang diperoleh dari langkah 2, hitung SVD dari $\mathbf{X}^T \mathbf{X} \mathbf{B}$ dengan $\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{U} \mathbf{D} \mathbf{V}^T$, lalu perbaharui $\mathbf{A} = \mathbf{U} \mathbf{V}^T$.
4. Langkah 2 dan 3 diulang sampai δ konvergen.
5. Melakukan normalisasi: $\hat{v}_k = \frac{\delta_k}{\|\delta_k\|}$; $k = 1, \dots, r$.

1.6.5 Categorical Principal Component Analysis

Categorical principal component analysis merupakan salah satu metode yang dilakukan untuk mengatasi multikolinieritas pada data berskala kategorik dengan menggunakan penskalaan optimal yang mengubah label kategorik ke nilai-nilai numerik dengan memaksimalkan keragaman antar peubah (Linting & Van Der Kooij, 2012).

Terdapat n individu dengan p peubah diberikan dengan $n \times p$ pengamatan skor matriks \mathbf{X} dimana masing-masing peubah didefinisikan oleh \mathbf{X}_j dengan $j = 1, 2, \dots, p$. Jika peubah \mathbf{X}_j merupakan skala pengukuran nominal atau ordinal, maka transformasi linier skala optimal diamati pada masing-masing skor dengan mengubahnya menjadi kuantifikasi kategori. Misalkan \mathbf{C}_j adalah matriks berukuran $c_j \times 1$ dengan c_j adalah jumlah kategori untuk setiap variabel \mathbf{X}_j dan nilai \mathbf{C}_j adalah bilangan bulat berurutan. Untuk menemukan solusi masalah \mathbf{X}_j dan nilai \mathbf{C}_j dapat dirumuskan dengan meminimalkan fungsi sebagai berikut (Linting & Van Der Kooij, 2012):

$$\sigma(\bar{\mathbf{X}}; \mathbf{C}_j) = \frac{1}{n} \sum_{j=1}^p \frac{1}{p} \left\{ \text{tr}(\bar{\mathbf{X}} - \mathbf{G}_j \mathbf{C}_j)^T (\bar{\mathbf{X}} - \mathbf{G}_j \mathbf{C}_j) \right\} \quad (2)$$

dengan $\bar{\mathbf{X}}$ adalah rata-rata \mathbf{X}_j .

Sebagai variabel numerik kontinu, variabel hasil kuantifikasi juga memiliki varians seperti pada umumnya variabel kontinu. Varians *categorical principal component analysis* dihitung dari memaksimalkan varians variabel kuantitatif hasil kuantifikasi. Kuantifikasi optimal mengubah variabel kategorik menjadi variabel numerik karena varians hanya dimiliki variabel numerik. Pada *categorical principal component analysis*, korelasi dihitung di antara variabel hasil kuantifikasi. Kuantifikasi optimal bertujuan mengoptimalkan matriks korelasi dari variabel terkuantifikasi dan untuk memaksimalkan varians pada variabel terkuantifikasi (Linting & Van Der Kooij, 2012).

Minimalisasi fungsi pada Persamaan (2) diberikan oleh algoritma *Alternating Least Square* (ALS) yaitu algoritma komputasi untuk meminimalkan fungsi kuadrat terkecil. Algoritma ALS menemukan perkiraan kuadrat terkecil dari setiap parameter dengan memperbarui setiap matriks parameter secara bergantian. Untuk meminimalkan fungsi $\sigma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ parameter matriks $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, dan $\boldsymbol{\theta}_3$, dengan $\boldsymbol{\theta}^{(t)}$ yaitu t estimasi dari $\boldsymbol{\theta}$ maka algoritma ALS memperbarui perkiraan $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, dan $\boldsymbol{\theta}_3$ dengan memecahkan

masalah kuadrat terkecil untuk setiap parameter berdasarkan Persamaan berikut (Kuroda et al., 2013):

$$\begin{aligned}\theta_1^{(t+1)} &= \arg \min_{\theta_1} \sigma(\theta_1, \theta_2^{(t)}, \theta_3^{(t)}), \\ \theta_2^{(t+1)} &= \arg \min_{\theta_2} \sigma(\theta_1^{(t+1)}, \theta_2, \theta_3^{(t)}), \\ \theta_3^{(t+1)} &= \arg \min_{\theta_3} \sigma(\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3).\end{aligned}\tag{3}$$

1.6.6 Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode statistika yang digunakan untuk menggambarkan hubungan antara variabel respon Y yang bersifat biner dengan variabel prediktor X yang bersifat kualitatif, kuantitatif ataupun kombinasi keduanya. Variabel respon Y terdiri dari 2 kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan $Y = 1$ (sukses) dan $Y = 0$ (gagal). Dalam keadaan demikian, variabel Y mengikuti distribusi Bernoulli untuk setiap observasi dengan distribusi peluang sebagai berikut (Maalouf, 2011):

$$f(Y_i) = \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i}, Y_i = 0, 1$$

dengan $1 - \pi(X_i)$ adalah probabilitas $f(Y_i) = 0$ dan $\pi(X_i)$ adalah probabilitas $f(Y_i) = 1$. Model dari regresi logistik biner adalah sebagai berikut:

$$\pi(X) = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$

dengan $\pi(X)$ adalah probabilitas sukses dan $\beta_0, \beta_1, \dots, \beta_p$ adalah parameter regresi. Pada regresi logistik, $\pi(X)$ adalah fungsi yang nonlinier sehingga untuk mempermudah dalam pendugaan parameter, $\pi(X)$ ditransformasi dengan menggunakan transformasi logit sebagai berikut (Maalouf, 2011):

$$\text{Logit } \pi(X) = g(X) = \ln \left[\frac{\pi(X)}{1 - \pi(X)} \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Metode estimasi parameter regresi logistik dilakukan dengan metode *maximum likelihood estimation* (Maalouf, 2011). Metode tersebut mengestimasi parameter β untuk memaksimalkan fungsi *likelihood*. Fungsi *likelihoodnya* dapat dilihat pada persamaan sebagai berikut:

$$L(\beta) = \prod_{i=1}^n f(Y_i) = \prod_{i=1}^n \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i}$$

Untuk mempermudah perhitungan, fungsi *likelihood* dimaksimalkan dalam bentuk $\ln L(\beta)$ sebagai berikut (Maalouf, 2011):

$$\ln L(\beta) = \ln \left(\prod_{i=1}^n \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i} \right)$$

1.6.7 Principal Component Logistic Regression

Principal component logistic regression merupakan teknik multivariat yang dapat dipakai untuk mereduksi dimensi dari variabel-variabel prediktor. *Principal component logistic regression* bertujuan meningkatkan estimasi parameter model logistik yang

memiliki multikolinearitas dengan menggunakan komponen utama dari variabel prediktor. Secara umum bentuk persamaan dari model *principal component logistic regression* yaitu (Pan et al., 2024):

$$\pi(Z) = \frac{\exp \{\beta_0 + \sum_{k=1}^r \sum_{j=1}^p Z_k a_{jk} \beta_j\}}{1 + \exp \{\beta_0 + \sum_{k=1}^r \sum_{j=1}^p Z_k a_{jk} \beta_j\}} = \frac{\exp \{\beta_0 + \sum_{k=1}^r Z_k \gamma_k\}}{1 + \exp \{\beta_0 + \sum_{k=1}^r Z_k \gamma_k\}}$$

dengan $\pi(Z)$ adalah probabilitas sukses dan $\beta_0, \gamma_1, \dots, \gamma_r$ adalah parameter regresi. Model *principal component logistic regression* tersebut dapat diformulasikan dalam bentuk matriks dengan transformasi fungsi logit $g = \ln \left(\frac{\pi(Z)}{1-\pi(Z)} \right)$ sebagai berikut (Pan et al., 2024):

$$g = \beta_0 + \sum_{k=1}^r Z_k \gamma_k + \varepsilon \quad (4)$$

dengan :

g = Probabilitas kejadian sukses pada $Y = 1$

β_0 = Konstanta

γ_k = Koefisien regresi logistik berdasarkan komponen utama yang terbentuk

Z_k = Komponen utama yang terbentuk

k = Banyaknya komponen utama yang terbentuk dari 1 hingga r

ε = Error

Pendugaan parameter $(\beta_0, \gamma_1, \dots, \gamma_r)$ dapat diperoleh dengan metode *maksimum likelihood estimation* dimana metode ini ditaksir dengan memaksimalkan fungsi *likelihood*. Fungsi *likelihoodnya* dapat dilihat pada persamaan sebagai berikut (Pan et al., 2024):

$$\begin{aligned} f((\beta_0, \gamma_1, \gamma_2, \dots, \gamma_r); Y_i) &= \prod_{i=1}^n f(Y_i) \\ &= \prod_{i=1}^n (\pi(Z_i))^{Y_i} (1 - \pi(Z_i))^{1-Y_i} \\ &= \prod_{i=1}^n \left(\frac{\pi(Z_i)}{1-\pi(Z_i)} \right)^{Y_i} (1 - \pi(Z_i)) \end{aligned}$$

dengan demikian fungsi ln *likelihoodnya* adalah:

$$\ln L((\beta_0, \gamma_1, \gamma_2, \dots, \gamma_r); Y_i) = \ln \prod_{i=1}^n \left(\frac{\pi(Z_i)}{1-\pi(Z_i)} \right)^{Y_i} (1 - \pi(Z_i))$$

1.6.8 Newton Raphson

Metode *newton raphson* adalah salah satu metode untuk menyelesaikan persamaan $f(x) = 0$. Ciri-ciri metode *newton raphson* yaitu memerlukan sebuah hampiran awal dan memerlukan perhitungan turunan fungsi $f(x)$ dalam setiap iterasi. Kedua ciri metode *newton raphson* menyatakan bahwa hampiran berikutnya diperoleh dengan cara menarik garis singgung kurva $y = f(x)$ pada titik yang mempunyai absis hampiran sebelumnya hingga meotong sumbu x . Titik potong garis singgung tersebut dengan sumbu x merupakan hampiran berikutnya. Proses berlanjut sampai hampiran yang diperoleh memenuhi syarat keakuratan yang ditentukan (Huang & He, 2022).

Misalkan g adalah suatu fungsi dengan bilangan x pada domain g merupakan titik tetap g jika memenuhi $x = g(x)$. Iterasi $x_{t+1} = g(x)$ disebut iterasi titik tetap.

Misalkan fungsi f mempunyai turunan pertama f' , barisan x_0, x_1, x_2, \dots yang diperoleh dari iterasi $x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}$ disebut barisan iterasi *newton*. Fungsi g yang didefinisikan sebagai $g(x) = x - \frac{f'(x)}{f''(x)}$ disebut fungsi iterasi *newton Raphson* (Huang & He, 2022).

1.6.9 Pengujian Parameter

Model yang telah diperoleh perlu diuji signifikansi pada koefisien $\beta_0, \gamma_1, \dots, \gamma_r$ terhadap variabel respon, yaitu dengan uji serentak. Pengujian ini dilakukan untuk memeriksa kemaknaan koefisien $\beta_0, \gamma_1, \dots, \gamma_r$ terhadap variabel respon secara bersama-sama dengan menggunakan statistik uji (Hosmer & Lemeshow, 2000).

Hipotesis:

H_0 : $\gamma_1 = \gamma_2 = \dots = \gamma_r = 0$ (semua komponen utama dalam model regresi logistik tidak mempengaruhi variabel respon).

H_1 : Paling sedikit ada satu $\gamma_k \neq 0$; $k = 1, 2, \dots, r$ (paling sedikit ada satu komponen utama dalam model regresi logistik yang berpengaruh terhadap variabel respon).

Statistik uji yang digunakan adalah statistik uji G atau *likelihood ratio test*.

$$G^2 = -2 \ln \frac{l_0}{l_1}$$

dengan:

l_0 = nilai *likelihood* untuk model yang tidak mengandung variabel prediktor.

l_1 = nilai *likelihood* untuk model yang mengandung variabel prediktor.

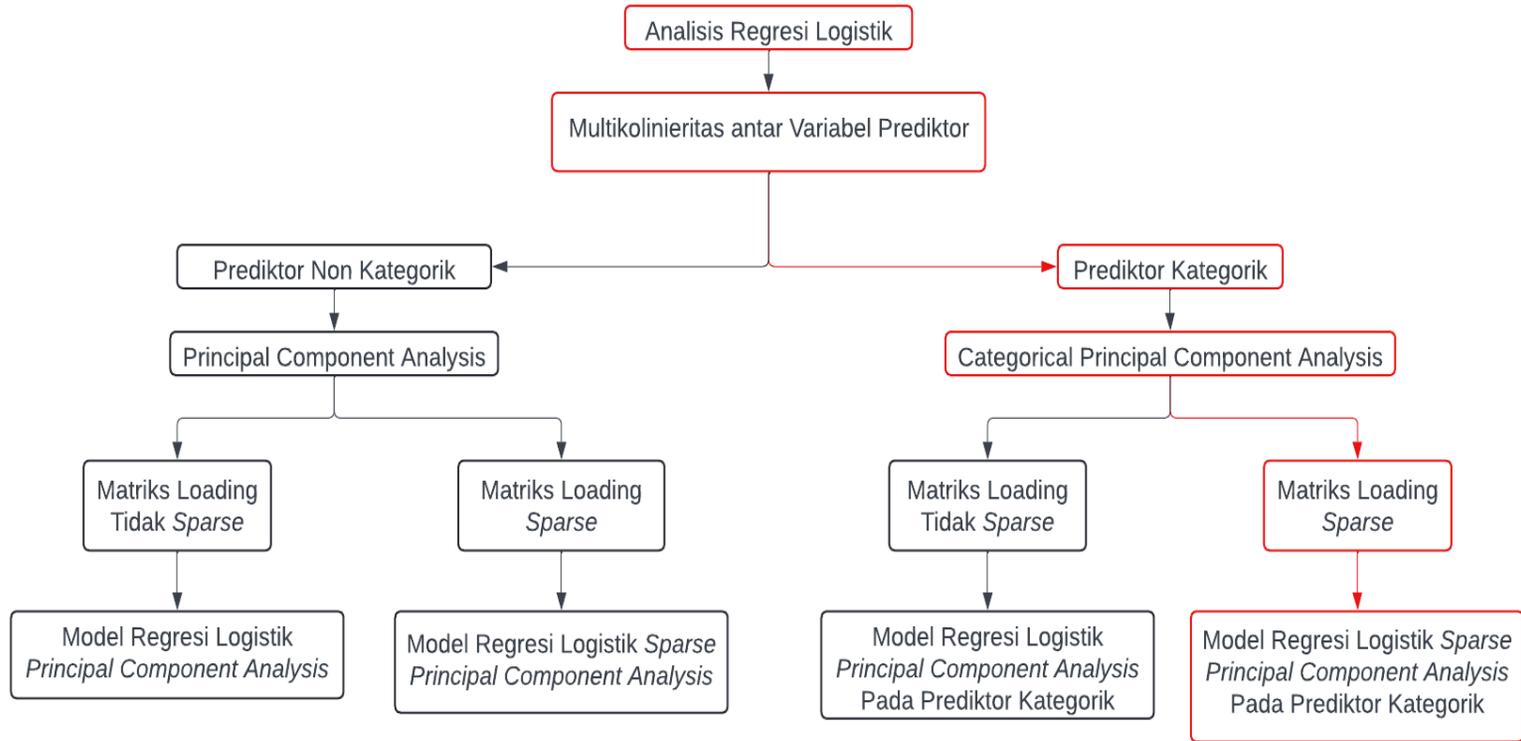
Statistik uji G mengikuti distribusi chi kuadrat (χ^2) dengan daerah penolakan H_0 adalah $G > \chi^2_{(\alpha;r)}$ atau $p\text{-value} < \alpha$, $\alpha = 0,05$ dan r = jumlah komponen utama (Hosmer & Lemeshow, 2000).

1.6.9 Stunting Pada Balita

Stunting atau keterlambatan pertumbuhan merupakan masalah gizi kronis yang sering terjadi pada anak-anak di dunia, termasuk Indonesia. Stunting dapat terlihat ketika anak memiliki tinggi badan lebih pendek dari tinggi badan normal yang seharusnya dimiliki oleh anak pada usia yang sama (Dayuningsih & Tria, 2020). Dampak yang ditimbulkan oleh stunting dalam jangka pendek yaitu terganggunya perkembangan otak, gangguan pertumbuhan fisik, dan gangguan metabolisme dalam tubuh. Dalam jangka panjang dampak yang dapat ditimbulkan yaitu menurunnya kemampuan kognitif dan prestasi belajar, menurunnya kekebalan tubuh, risiko munculnya berbagai penyakit, disabilitas pada usia tua yang berakibat pada rendahnya produktivitas (Mardlatilla & Ratih, 2022). Berdasarkan Survei Status Gizi Indonesia (SSGI) Kementerian Kesehatan, prevalensi balita stunting di Indonesia mencapai 21,6% pada 2022. Standard *World Health Organization* (WHO) terkait prevalensi stunting harus di angka kurang dari 20%.

Faktor penyebab stunting diantaranya dipengaruhi oleh rendahnya akses terhadap makanan bergizi, Ibu yang masa remajanya dan masa kehamilannya kurang nutrisi, rendahnya akses terhadap pelayanan kesehatan, kurangnya pengetahuan orang tua mengenai kesehatan dan gizi, pola asuh, pemberian ASI eksklusif, serta riwayat penyakit infeksi serta faktor genetik. Multi faktor yang sangat beragam tersebut membutuhkan intervensi yang paling menentukan yaitu pada 1000 hari pertama kehidupan (Dayuningsih & Tria, 2020).

1.6.10 Kerangka Konseptual



Gambar 1. Kerangka Konseptual

BAB II METODE PENELITIAN

2.1 Sumber Data dan Variabel Penelitian

Data yang akan digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kabupaten Muna Barat, Provinsi Sulawesi Tenggara yaitu data kondisi pertumbuhan pada balita yang mengikuti Posyandu pada bulan Desember 2022 di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara. Variabel yang digunakan dalam penelitian ini terdiri dari dua variabel yaitu variabel respon (Y) dan variabel prediktor (X) yang memiliki satuan kategorik. Variabel prediktor (X) dalam penelitian ini terdiri dari 13 variabel yang merupakan faktor-faktor yang diduga dapat mempengaruhi kondisi pertumbuhan pada balita. Adapun variabel tersebut sebagai berikut.

Tabel 1. Variabel Penelitian

Jenis Variabel	Nama Variabel
Variabel Respon (Y)	Kondisi Pertumbuhan Pada Balita (Kejadian <i>Stunting</i>)
Variabel Prediktor (X_1)	Jenis Kelamin
Variabel Prediktor (X_2)	Tinggi Badan Bayi Lahir
Variabel Prediktor (X_3)	Berat Badan Bayi Lahir
Variabel Prediktor (X_4)	Memiliki Riwayat Penyakit
Variabel Prediktor (X_5)	Usia Pemberian MPASI
Variabel Prediktor (X_6)	Mengonsumsi Susu Formula
Variabel Prediktor (X_7)	Ibu Hamil Mendapatkan Pemberian Makanan Tambahan (PMT)
Variabel Prediktor (X_8)	Ibu Mendapatkan Tablet Tambah Darah (TTD)
Variabel Prediktor (X_9)	Usia Ayah
Variabel Prediktor (X_{10})	Usia Ibu
Variabel Prediktor (X_{11})	Pendidikan Terakhir Ayah
Variabel Prediktor (X_{12})	Pendidikan Terakhir Ibu
Variabel Prediktor (X_{13})	Air Minum Yang dikonsumsi

2.2 Definisi Operasional Variabel

Adapun definisi dari variabel–variabel penelitian adalah sebagai berikut:

Tabel 2. Definisi Operasional Variabel

Variabel	Definisi Operasional	Kategori
Kejadian <i>Stunting</i> (Y)	Kondisi pertumbuhan pada balita berdasarkan tinggi badan menurut usia yang kurang dari -2 Standar Deviasi pada kurva pertumbuhan WHO.	0 = Tidak <i>Stunting</i> 1 = <i>Stunting</i>
Jenis Kelamin (X_1)	Identitas fisik balita berdasarkan jawaban ibu balita.	1 = Perempuan 2 = Laki-laki

Tinggi Badan Bayi Lahir (X_2)	Tinggi badan balita saat dilahirkan dengan satuan <i>centi meter</i> (cm).	1 = Rendah (Laki-laki < 48 cm, Perempuan < 47,3 cm) 2 = Normal (Laki-laki \geq 48 cm, Perempuan \geq 47,3 cm)
Berat Badan Bayi Lahir (X_3)	Berat badan balita saat dilahirkan dengan satuan <i>kilo gram</i> (kg).	1 = Rendah (< 2,5 kg) 2 = Normal (\geq 2,5 kg)
Memiliki Riwayat Penyakit (X_4)	Riwayat penyakit infeksi yang berhubungan dengan <i>stunting</i> yang dimiliki oleh balita yaitu Diare, ISPA, Kecacingan, dan TBC.	1 = Tidak 2 = Ya
Usia Pemberian MPASI (X_5)	Usia pertama pemberian MPASI pada balita.	1 = Sesuai (6 bulan) 2 = Tidak Sesuai (< 6 bulan & > 6 bulan)
Mengonsumsi Susu Formula (X_6)	Pengakuan ibu balita dalam memberikan susu formula sebagai tambahan ASI kepada balita.	1 = Ya 2 = Tidak
Ibu Hamil Mendapatkan Pemberian Makanan Tambahan (PMT) (X_7)	Pengakuan ibu pada saat hamil apakah sudah mendapatkan PMT.	1 = Ya 2 = Tidak
Ibu Mendapatkan Tablet Tambah Darah (TTD) (X_8)	Pengakuan ibu pada saat hamil apakah sudah mendapatkan TTD.	1 = Ya 2 = Tidak
Usia Ayah (X_9)	Usia ayah balita saat penelitian dilakukan.	1 = < 25 Tahun 2 = 25 – 40 Tahun 3 = > 40 Tahun
Usia Ibu (X_{10})	Usia ibu balita saat penelitian dilakukan.	1 = < 21 Tahun 2 = 21-35 Tahun 3 = > 35 Tahun
Pendidikan Terakhir Ayah (X_{11})	Jenjang pendidikan formal yang dicapai oleh ayah balita.	1 = SD 2 = SMP 3 = SMA 4 = Perguruan Tinggi
Pendidikan Terakhir Ibu (X_{12})	Jenjang pendidikan formal yang dicapai oleh ibu balita.	1 = SD 2 = SMP 3 = SMA 4 = Perguruan Tinggi
Air Minum Yang Dikonsumsi (X_{13})	Air minum yang dikonsumsi oleh keluarga.	1 = Galon kemasan 2 = Galon isi ulang 3 = Air masak

2.3 Metode Analisis

Adapun langkah-langkah analisis yang dilakukan berdasarkan tujuan penelitian adalah sebagai berikut:

- a. Memperoleh estimasi parameter model regresi logistik *sparse principal component analysis* pada prediktor kategorik dengan tahapan sebagai berikut :
 1. Mentransformasi variabel prediktor kategorik menjadi variabel prediktor numerik kontinu X^* dengan kuantifikasi optimal.
 2. Menghitung matriks korelasi pada variabel hasil kuantifikasi X^* .
 3. Menghitung nilai eigen dan vektor eigen dari matriks korelasi.
 4. Memilih komponen utama optimal Z_k^* dengan $k < j$ berdasarkan nilai eigen dan proporsi total varians komponen utama optimal yang dihasilkan oleh nilai eigen dari variabel hasil kuantifikasi.
 5. Menghitung matriks nilai loading berdasarkan komponen utama optimal yang telah dihasilkan.
 6. Memilih parameter tuning ℓ_1 dan ℓ_2 yang menghasilkan model matriks *loading* yang *sparse* dan proporsi keragaman kumulatif yang dihasilkan $\geq 80\%$.
 7. Menghitung penduga *naive elastic net* δ_k menggunakan Persamaan (2.1)
 8. Untuk setiap δ_k yang diperoleh dari langkah 7, dihitung SVD dari $X^T X B$ dengan $X^T X B = U D V^T$, lalu dihitung $A = U V^T$
 9. Mengulangi langkah 7 dan 8 sampai diperoleh δ konvergen.
 10. Melakukan normalisasi: $v_k = \frac{\delta_k}{\|\delta_k\|}$, dengan k adalah jumlah komponen utama optimal yang terbentuk.
 11. Membentuk komponen utama *sparse* optimal berdasarkan v_k yang telah diperoleh.
 12. Menyatakan model regresi logistik *sparse principal component analysis* pada prediktor kategorik berdasarkan komponen utama *sparse* yang terbentuk dengan menggunakan persamaan *principal component logistic regression* sebagai berikut :

$$g = \beta_0 + \sum_{k=1}^r Z_k \gamma_k + \varepsilon$$

13. Mencari estimasi parameter β_0 dan γ_k dengan menggunakan metode *maximum likelihood*.
- b. Memperoleh estimasi model kondisi pertumbuhan balita (stunting) di Kecamatan Maginti, Kabupaten Muna Barat, Provinsi Sulawesi Tenggara berdasarkan regresi logistik *sparse principal component analysis* pada prediktor karegorik dengan tahapan sebagai berikut :
 1. Menguji adanya multikolinearitas pada peubah prediktor melalui matriks korelasi.
 2. Membuat analisis deskriptif data untuk memberikan gambaran mengenai karakteristik data.
 3. Menentukan nilai-nilai variabel prediktor numerik kontinu berdasarkan proses kuantifikasi optimal.
 4. Menghitung matriks korelasi pada variabel prediktor hasil kuantifikasi.

5. Menghitung nilai eigen dan vektor eigen dari matriks korelasi variabel prediktor hasil kuantifikasi.
6. Memilih komponen utama optimal \mathbf{Z}_k^* dengan $k < j$ berdasarkan nilai eigen dan proporsi total varians komponen utama yang dihasilkan oleh nilai eigen dari variabel hasil kuantifikasi.
7. Membentuk komponen utama *sparse* optimal dari hasil komponen utama optimal.
8. Membentuk model regresi logistik pada komponen utama *sparse* optimal yang dihasilkan yaitu \mathbf{Z}_k^* , $k = 1, 2, \dots, r$ berdasarkan model umum *principal component logistic regression* seperti yang diberikan pada persamaan berikut :

$$g = \beta_0 + \sum_{k=1}^r Z_k \gamma_k + \varepsilon$$

9. Menguji parameter secara serentak dengan statistik uji G atau *Likelihood Ratio Test* berdasarkan persamaan sebagai berikut:

$$G^2 = -2 \ln \frac{l_0}{l_1}$$

10. Menginterpretasikan model.
11. Menarik kesimpulan.