

# PCR DAN PLSR ALGORITMA NIPALS DALAM MENANGANI MULTIKOLINIERITAS PADA PREVALENSI STUNTING DI NUSA TENGGARA TIMUR

Natalie Efrata Susanti<sup>1</sup>, Vera Maya Santi<sup>2§</sup>, Devi Eka Wardani<sup>3</sup>

<sup>1</sup>Program Studi Matematika, FMIPA, Universitas Negeri Jakarta [Email: natalie.efrata1812@gmail.com]

<sup>2</sup>Program Studi Statistika, FMIPA, Universitas Negeri Jakarta [Email: vmsanti@unj.ac.id]

<sup>3</sup>Program Studi Matematika, Universitas Negeri Jakarta [Email: deviekawm@gmail.com]

<sup>§</sup>Corresponding Author

## ABSTRACT

*Nutritional problems contribute to 50% of deaths among children under five, particularly in low- and middle-income countries. One of the most common issues in Indonesia is stunting, a condition where a child's height falls below the standard for their age. In 2022, East Nusa Tenggara (NTT) recorded the highest stunting prevalence in Indonesia at 35.3%. However, quantitative statistical analyses of its contributing factors in NTT remain limited. This study aims to compare partial least squares regression (PLSR) using the NIPALS algorithm with principal component regression (PCR) in addressing multicollinearity. The secondary data were obtained from the 2022 Indonesian Nutrition Status Survey (SSGI), published by the Ministry of Health and BPS NTT, consisting of one response variable and ten predictor variables. Results showed that the PLSR model outperforms PCR, with an adjusted  $R^2$  of 0.741 compared to 0.322. The superiority of PLSR is also evident from its lower RMSE and MAE values (2.783 and 1.910) compared to PCR (4.742 and 3.346). PLSR identified five significant predictors: average daily protein consumption per capita, number of children receiving DPT and HB immunizations, Human Development Index, percentage of households with access to safe drinking water, and number of people living in poverty.*

**Keywords:** PLSR, NIPALS, PCR, multicollinearity, stunting

## 1. PENDAHULUAN

Stunting adalah persoalan gizi anak yang utama di Indonesia, ditandai dengan tinggi badan yang jauh lebih rendah dibanding anak pada usianya akibat kekurangan gizi dalam jangka panjang, terutama selama 1.000 hari awal kehidupan. Kondisi ini berdampak pada pertumbuhan fisik, gangguan perkembangan kognitif, dan meningkatkan risiko penyakit kronis yang berpotensi menimbulkan kerugian ekonomi Indonesia. Faktor penyebabnya meliputi kurangnya konsumsi gizi ibu hamil, pola pemberian ASI, status ekonomi, pendidikan, dan usia ibu.

Pada tahun 2022, diperkirakan ada 149 juta balita yang mengalami stunting (WHO, 2024). Berdasarkan Survei Status Gizi Indonesia (SSGI) 2021 yang dijalankan Kementerian Kesehatan RI,

prevalensi stunting di Indonesia mencapai 24,4% (2021) dan turun hingga 21,6% (2022), tetapi menurut WHO, angka ini masih tinggi. Stunting dengan tingkat prevalensi tertinggi di Indonesia pada 2021 terjadi di Provinsi Nusa Tenggara Timur, yaitu sebesar 37,8%. Persentase stunting di NTT pada 2022 berkurang sebesar 2,5% dari tahun 2021, yaitu menjadi 35,3%. Akan tetapi, provinsi ini masih menduduki posisi pertama prevalensi stunting tertinggi di Indonesia. Sayangnya, penelitian secara keseluruhan Provinsi NTT terkait prevalensi stunting secara kuantitatif masih belum sebanyak penelitian yang dilakukan per satu kabupaten, kecamatan, atau kelurahan.

Identifikasi variabel signifikan dilakukan dengan regresi linear, namun terkadang terjadi

pelanggaran yang menghambat analisis, seperti multikolinearitas variabel independen, di mana nilai *Variance Inflation Factor* (VIF) lebih besar dari 5 (Hair, 2021:93). Berbagai pendekatan telah dikembangkan untuk mengatasi multikolinearitas, seperti regresi ridge yang menambahkan konstanta bias sehingga koefisien menyusut mendekati nol, namun hal ini berpotensi menghasilkan estimasi bias (Nugraha, 2019) dan regresi LASSO yang memilih variabel paling berpengaruh. Tetapi dapat menyebabkan bias koefisien dan hilangnya informasi penting karena penyusutan koefisien mendekati atau sama dengan nol (Santi, 2019; Santi, 2020; Santi, 2024).

*Partial Least Squares Regression* (PLSR) merupakan pendekatan yang dapat digunakan untuk mengatasi multikolinearitas tanpa menghilangkan informasi penting. PLSR melakukan reduksi dimensi dengan membentuk komponen utama yang memaksimalkan kovarians antara variabel dependen dengan setiap kemungkinan kombinasi linear dari variabel independen yang mampu menjaga informasi penting dalam model (Duan, dkk., 2020; Santi, 2011).

Selain itu ada metode *Principal Component Regression* (PCR) yang memadukan regresi dengan *principal component analysis* (PCA), di mana PCA memiliki prinsip menyederhanakan variabel independen melalui reduksi dimensi

dengan tetap mempertahankan sebagian informasi variabel asalnya. Komponen utama pada PCR merupakan kombinasi linear dari variabel independen dan hal ini menjaga informasi penting dalam model tetap ada serta multikolinearitas teratasi (Effendi dkk., 2023)

Penelitian ini bertujuan membandingkan kinerja PLSR berbasis algoritma NIPALS dan PCR dalam mengatasi permasalahan multikolinearitas pada data prevalensi stunting Provinsi Nusa Tenggara Timur tahun 2022, sekaligus mengidentifikasi faktor-faktor determinan yang berpengaruh signifikan terhadap tingkat stunting. Temuan penelitian ini diharapkan memberikan dasar empiris yang kuat bagi perumusan kebijakan dan strategi intervensi penurunan stunting di Provinsi NTT.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif, di mana menggunakan data numerik yang dianalisis secara statistik. Data sekunder yang digunakan adalah prevalensi stunting Provinsi NTT 2022 sebagai variabel dependen yang diperoleh dari publikasi hasil SSGI 2022 Kementerian Kesehatan serta 10 variabel penjelas yang diperoleh dari *website* resmi Badan Pusat Statistik NTT. Struktur data adalah *cross sectional data*. Rincian variabel penelitian disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala	Satuan
Y	Prevalensi Stunting di Provinsi NTT 2022	Rasio	%
$X_1$	Persentase rumah tangga dengan akses sanitasi layak	Rasio	%
$X_2$	Rata-rata konsumsi protein per kapita sehari	Rasio	Gram
$X_3$	Jumlah balita yang pernah mendapatkan imunisasi DPT dan HB	Rasio	Jiwa
$X_4$	Jumlah balita yang pernah mendapatkan imunisasi polio	Rasio	Jiwa
$X_5$	Jumlah balita yang pernah mendapatkan imunisasi campak	Rasio	Jiwa
$X_6$	Jumlah balita yang pernah mendapatkan imunisasi lengkap	Rasio	Jiwa
$X_7$	Indeks Pembangunan Manusia	Rasio	-
$X_8$	Persentase rumah tangga berdasarkan sumber air minum layak	Rasio	%
$X_9$	Jumlah penduduk miskin	Rasio	Ribu Jiwa
$X_{10}$	Persentase perempuan yang pernah melahirkan dalam 2 tahun terakhir dengan berat badan rendah (BBLR atau berat badan kurang dari 2500 gram)	Rasio	%

Adapun tahapan yang akan dilakukan guna mencapai tujuan dari penelitian ini adalah sebagai berikut:

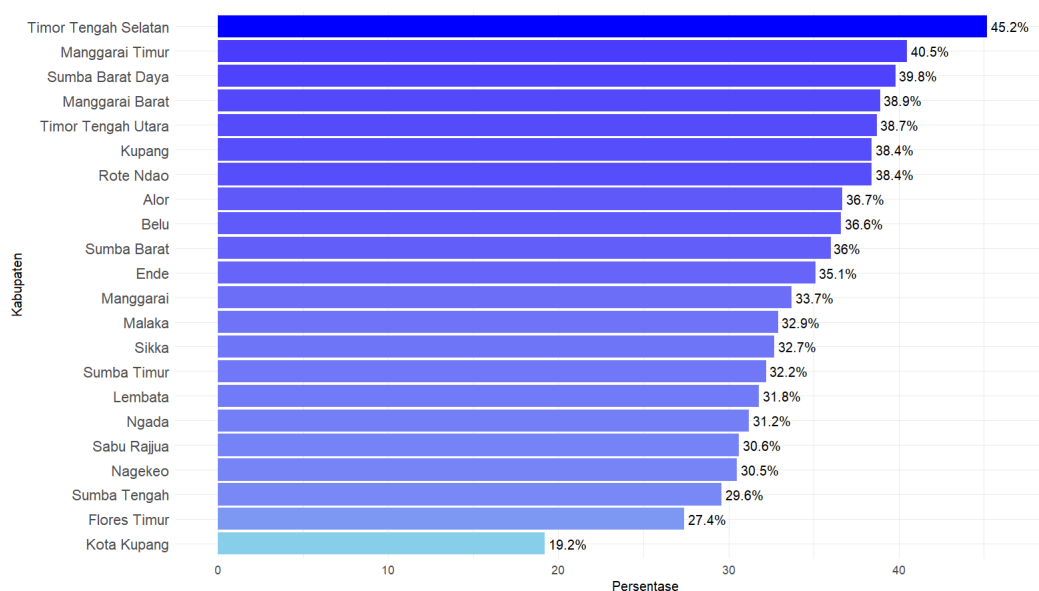
1. Melakukan input data
2. Melakukan eksplorasi data
3. Melakukan standarisasi data
4. Membangun spesifikasi regresi linear berganda
5. Melakukan estimasi parameter dengan *Ordinary Least Squares* (OLS)
6. Membangun spesifikasi model PLSR dan PCR
7. Pendugaan parameter PLSR
8. Melakukan pengujian signifikansi parameter secara parsial

9. Melakukan perhitungan perbandingan kriteria kebaikan model, yakni *Adjusted R Squared*, *Mean Absolute Error (MAE)*, dan *Root Mean Square Error (RMSE)*
10. Melakukan interpretasi hasil dari metode terbaik yang didapatkan dari hasil perbandingan kriteria kebaikan model.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Eksplorasi Data

Provinsi NTT memiliki 22 kabupaten/kota. Berdasarkan data SSGI 2022, NTT menempati posisi pertama prevalensi stunting tertinggi di Indonesia pada 2022. Gambar 1 menunjukkan Kabupaten Timor Tengah Selatan menempati posisi pertama prevalensi stunting tertinggi di NTT sebesar 45,2%. Sedangkan, kabupaten/kota dengan tingkat prevalensi terendah terjadi di Kota Kupang sebesar 19,2%. Berdasarkan Tabel 2, tingkat prevalensi stunting di NTT memiliki rata-rata 34,37% menunjukkan angka yang cukup tinggi dan memerlukan perhatian serius.



Gambar 2. Prevalensi Stunting per Kabupaten/Kota di Provinsi Nusa Tenggara Timur

Tabel 2. Statistika Deskriptif

Peubah	Mean	Median	Min.	Maks.	Simpangan Baku	Ragam
$Y$	34,37	34,4	19,2	45,2	5,465	29,862
$X_1$	73,71	78,66	45,83	88,82	14,173	200,888
$X_2$	56,03	56,58	45,16	64,73	4,859	23,606
$X_3$	4.676	4.026	1.664	9.879	2.318,443	5.375.176
$X_4$	4.577	4.076	1.646	9.750	2.260,298	5.108.949
$X_5$	4.676	4.458	1.653	10.583	2.473,29	6.117.161
$X_6$	4.492	4.256	1.599	10.327	2.275,465	5.177.740
$X_7$	64,94	64,67	57,9	80,2	4,111	16,903
$X_8$	86,51	89,73	61,18	98,86	11,149	124,299
$X_9$	51,44	41,59	18,01	120,45	26,905	723,883
$X_{10}$	15,95	14,35	7,09	30,87	6,021	36,257

Variabel jumlah balita yang mendapatkan imunisasi campak ( $X_5$ ) memiliki simpangan baku dan ragam terbesar mencerminkan ketidakmerataan cakupan imunisasi campak antar daerah. Ketimpangan serupa terlihat pada jumlah penduduk miskin ( $X_9$ ) dengan nilai minimum dan maksimum yang cukup berbeda. Ini

menunjukkan variasi jumlah kemiskinan miskin yang signifikan antar kabupaten/kota.

Variabel lain seperti akses sanitasi layak, konsumsi protein, dan imunisasi lainnya juga menunjukkan variasi yang beragam antar wilayah, sehingga diperlukan analisis lebih mendalam untuk mengidentifikasi faktor

yang paling berkontribusi terhadap prevalensi stunting di NTT.

### 3.2 Pemodelan Regresi Linear Berganda

Variabel independen yang digunakan memiliki satuan yang berbeda, sehingga sebelum dilakukan regresi linear, data terlebih dahulu distandarisasi. Parameter regresi linear berganda diestimasi menggunakan *Ordinary Least Squares* (OLS). Setelah proses standarisasi, diperoleh persamaan regresi sebagai berikut:

$$Y = 34,368 - 0,254X_1 - 1,011X_2 - 0,838X_3 - 4,242X_4 - 0,121X_5 + 4,782X_6 - 3,808X_7 + 2,081X_8 + 3,991X_9 - 0,501X_{10} \quad (1)$$

Persamaan di atas menunjukkan bahwa prevalensi stunting di Nusa Tenggara Timur (Y) bernilai 34,368 ketika semua variabel independen bernilai 0. Jika koefisien dari variabel independen ke- $i$  memiliki tanda positif, maka nilai variabel Y akan mengalami peningkatan ketika nilai variabel X mengalami peningkatan. Sebaliknya, jika tanda koefisien dari variabel independen ke- $i$  negatif, maka nilai variabel Y mengalami penurunan ketika nilai variabel X mengalami peningkatan.

### 3.3 Uji Signifikansi Parameter Simultan

Uji signifikansi parameter simultan dilakukan guna memeriksa pengaruh semua variabel independen terhadap variabel dependen. Hipotesis pada uji ini dengan  $H_0: \beta_1 = \beta_2 = \dots = \beta_j = 0$  lawan  $H_1: \text{Min } \exists \beta_j \neq 0$  untuk  $j = 1, 2, \dots, 10$  dimana tolak  $H_0$  bila  $p - \text{value} > 0,05$ .

Berdasarkan hasil uji signifikansi parameter simultan, diperoleh nilai  $p - \text{value}$  sebesar 0,008 lebih kecil dari taraf signifikansi 0,05. Hal ini menunjukkan bahwa variabel independen secara simultan memiliki pengaruh yang signifikan terhadap prevalensi stunting.

### 3.4 Uji Signifikansi Parameter Parsial

Uji signifikansi parsial dilakukan untuk mengetahui pengaruh masing-masing variabel independen terhadap variabel dependen. Hipotesis yang diuji  $H_0: \beta_j = 0$  lawan  $H_1: \exists \beta_j \neq 0$  untuk  $j = 1, 2, \dots, 10$  dimana tolak  $H_0$  bila  $p - \text{value} > 0,05$ .

Hasil pada Tabel 3 memperlihatkan bahwa hanya ada dua variabel yang signifikan, yaitu  $X_7$  dan  $X_9$  dan variabel lainnya tidak signifikan, yaitu di mana besar  $p - \text{value}$  lebih dari taraf signifikansi 5%.

Tabel 3. Uji Signifikansi Parameter Secara Parsial

Variabel	$t\text{-value}$	$p\text{-value}$	Ket.
$X_1$	-0,196	0,848	Tidak signifikan
$X_2$	-1,216	0,249	Tidak signifikan
$X_3$	-0,058	0,955	Tidak signifikan
$X_4$	-0,215	0,833	Tidak signifikan
$X_5$	-0,014	0,989	Tidak signifikan
$X_6$	0,77	0,457	Tidak signifikan
$X_7$	-3,357	0,006	Signifikan
$X_8$	1,704	0,116	Tidak signifikan
$X_9$	2,396	0,035	Signifikan
$X_{10}$	-0,576	0,576	Tidak signifikan

Hal ini menimbulkan dugaan adanya pelanggaran asumsi dalam pemodelan regresi sehingga diperlukan uji asumsi klasik untuk mendeteksi kemungkinan pelanggaran tersebut.

### 3.5 Uji Asumsi Klasik

#### Uji Normalitas

Uji normalitas bertujuan untuk memeriksa residu pada model regresi berdistribusi normal atau tidak. Pada uji ini akan digunakan uji Kolmogorov-Smirnov dengan hipotesis  $H_0$ : residu berdistribusi normal

dan  $H_1$ : residu tidak berdistribusi normal.

Berdasarkan uji yang dilakukan, hasil  $p\text{-value}$  sebesar 0,446 dimana lebih besar dari 0,05 sehingga dapat disimpulkan bahwa residu berdistribusi normal.

#### Uji Homoskedastisitas

Uji homoskedastisitas dilakukan untuk memeriksa perbedaan varians residual antar pengamatan dengan hipotesis  $H_0$ : Varians residual memenuhi asumsi homoskedastisitas dan

$H_1$ : Varians residual tidak memenuhi asumsi homoskedastisitas.

Berdasarkan uji yang dilakukan, hasil p-value diperoleh sebesar 0,545. Nilai ini lebih besar dari 0,05 sehingga dapat disimpulkan bahwa residu berdistribusi normal.

### 3.6 Pemeriksaan Multikolinearitas

Multikolinearitas dideteksi dengan memeriksa besarnya nilai *Variance Inflation Factors* (VIF), yaitu apabila nilai  $VIF > 5$ , maka variabel tersebut dinyatakan mengandung multikolinearitas.

Berdasarkan Tabel 4, terdapat lima variabel independen yang memiliki nilai VIF lebih dari 5, yaitu jumlah balita yang pernah mendapatkan imunisasi DPT dan HB ( $X_3$ ), jumlah balita yang pernah mendapatkan imunisasi polio ( $X_4$ ), jumlah balita yang pernah mendapatkan imunisasi campak ( $X_5$ ), jumlah balita yang pernah mendapatkan imunisasi lengkap ( $X_6$ ), dan jumlah penduduk miskin ( $X_9$ ).

### 3.7 Pemodelan PLSR

Pemodelan PLSR algoritma NIPALS dibangun dengan proses berikut:

1. Menentukan matriks dan vektor rata-rata terpusat.

Matriks rata-rata terpusat untuk X

$$\tilde{X} = \begin{bmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1m} - \bar{x}_m \\ x_{21} - \bar{x}_1 & \cdots & x_{2m} - \bar{x}_m \\ \vdots & \cdots & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{nm} - \bar{x}_m \end{bmatrix} \quad (2)$$

Vektor rata-rata terpusat untuk Y

$$\tilde{y} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \quad (3)$$

dengan  $\bar{y}, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$

2. Menentukan vektor bobot ( $w$ )

$$w_l = \frac{1}{\sqrt{\sum_{j=1}^m \text{cov}(x_{j(l)}, y_{(l)})^2}} \begin{bmatrix} \text{cov}(x_{1(l)}, y_{(l)}) \\ \vdots \\ \text{cov}(x_{m(l)}, y_{(l)}) \end{bmatrix} \quad (4)$$

Dengan  $x_{(l)}$  dan  $y_{(l)}$  adalah matriks residual dari perhitungan komponen sebelumnya dan  $l$  menyatakan iterasi.

3. Menentukan vektor komponen ( $t$ )

$$t_{(l)} = X_{(l)} w_{(l)} \quad (5)$$

4. Di mana  $X_{(l)}$  merupakan matriks residual variabel X yang didapatkan dari iterasi sebelumnya. Menentukan vektor *loading* ( $p$ )

$$p_{(l)} = (t_{(l)}^T t_{(l)})^{-1} (X_{(l)}^T t_{(l)}) \quad (6)$$

Tabel 4. Tabel VIF Multikolinearitas

Variabel	VIF	Keterangan
$X_1$	3,344	Tidak terjadi multikolinearitas
$X_2$	1,381	Tidak terjadi multikolinearitas
$X_3$	412,358	<b>Terjadi multikolinearitas</b>
$X_4$	775,502	<b>Terjadi multikolinearitas</b>
$X_5$	150,909	<b>Terjadi multikolinearitas</b>
$X_6$	77,069	<b>Terjadi multikolinearitas</b>
$X_7$	2,573	Tidak terjadi multikolinearitas
$X_8$	2,982	Tidak terjadi multikolinearitas
$X_9$	5,545	<b>Terjadi multikolinearitas</b>
$X_{10}$	1,515	Tidak terjadi multikolinearitas

5. Menentukan estimasi parameter  $c$

$$c_{(l)} = (t_{(l)}^T t_{(l)})^{-1} (t_{(l)}^T y_{(l)}) \quad (7)$$

Parameter  $c$  digunakan untuk membantu perhitungan matriks dan vektor residual.

6. Menentukan matriks dan vektor residual. Jika matriks residual X belum mencapai matriks nol, maka kembali lakukan langkah dari nomor 2.

Matriks residual  $X$  di iterasi ke- $l$  dapat diperoleh dengan cara berikut:

$$X_{(l)+1} = X_{(l)} - \hat{X}_{(l)}$$

Vektor residual  $y$  di iterasi ke- $l$  dapat diperoleh dengan cara berikut:

$$y_{(l)+1} = y_{(l)} - \hat{y}_{(l)}$$

Pada penelitian ini, matriks residual mencapai matriks nol pada saat iterasi ke-10, sehingga didapatkan nilai vektor bobot ( $w$ ), vektor komponen ( $t$ ), dan vektor loading ( $p$ ) iterasi ke-10 sebagai berikut:

$$W_{(10)10 \times 1} = \begin{bmatrix} 0,023 \\ -0,005 \\ 0,542 \\ -0,765 \\ 0,329 \\ -0,112 \\ 0,010 \\ -0,009 \\ 0,015 \\ -0,012 \end{bmatrix}; t_{(10)22 \times 1} = \begin{bmatrix} -0,035 \\ -0,003 \\ -0,024 \\ 0,008 \\ \vdots \\ 0,031 \\ 0,072 \\ -0,0002 \\ 0,014 \\ -0,015 \end{bmatrix}$$

$$P_{(10)10 \times 1} = \begin{bmatrix} 0,023 \\ -0,005 \\ 0,542 \\ -0,765 \\ 0,329 \\ -0,112 \\ 0,010 \\ -0,009 \\ 0,015 \\ -0,012 \end{bmatrix}$$

7. Menentukan komponen terpilih PLSR dengan nilai *Predicted Residual Error Sum of Squares* (PRESS). Nilai PRESS ditentukan dengan menggunakan:

$$PRESS_{(l)} = \sum_{i=1}^n (y_{(l)i} - \hat{y}_{(l)(-i)})^2 \quad (8)$$

Sehingga diperoleh hasil sebagai berikut:

Tabel 5. Nilai PRESS Komponen

Komponen	PRESS
1	578,871
2	603,251
3	475,246
4	325,696
<b>5</b>	<b>304,663</b>
6	311,174
7	307,147
8	350,475
9	489,794
10	684,410

Berdasarkan hasil pada tabel 7, akan dipilih nilai PRESS terkecil, yaitu PRESS dari komponen ke-5 sebesar 304,663.

8. Menentukan koefisien  $\hat{\beta}^{PLSR}$

Nilai  $\hat{\beta}^{PLSR}$  diperoleh dengan rumus berikut:

$$\hat{\beta}^{PLSR} = W(P^T W)^{-1} c \quad (9)$$

Berdasarkan rumus tersebut dan komponen yang diperoleh, didapatkan hasil sebagai berikut:

$$\hat{\beta}^{PLSR} = \begin{bmatrix} -0,092 \\ -1,055 \\ -0,829 \\ -0,545 \\ 0,282 \\ 0,683 \\ -3,804 \\ 1,958 \\ 4,111 \\ -0,342 \end{bmatrix}$$

Dari penaksiran parameter di atas terbentuklah persamaan seperti di bawah ini:

$$Y = 34,368 - 0,092X_1^* - 1,055X_2^* + 0,829X_3^* - 0,545X_4^* + 0,282X_5^* + 0,683X_6^* - 3,804X_7^* + 1,958X_8^* + 4,111X_9^* - 0,342X_{10}^* \quad (10)$$

Persamaan di atas merupakan persamaan dalam bentuk yang distandarisasi sehingga akan dilakukan transformasi kembali dalam bentuk variabel independen awal sehingga diperoleh persamaan, yaitu:

$$Y = 85,836 - 0,06X_1 - 0,217X_2 + 0,0004X_3 - 0,0002X_4 + 0,0001X_5 + 0,0003X_6 - 0,925X_7 + 0,176X_8 + 0,153X_9 - 0,057X_{10} \quad (11)$$

9. Melakukan uji signifikansi parameter secara parsial.

Selanjutnya dilakukan uji signifikansi parameter secara parsial dan didapatkan hasil  $t_{hitung}$  yang ditampilkan pada Tabel 6.

Tabel 6. Uji Signifikansi Metode PLSR

Variabel	$t_{hitung}$	$t_{tabel}$	Keterangan
$X_1$	-0,119	1,796	Tidak signifikan
$X_2$	-3,897	1,796	<b>Signifikan</b>
$X_3$	-1,897	1,796	<b>Signifikan</b>
$X_4$	-1,262	1,796	Tidak signifikan
$X_5$	0,696	1,796	Tidak signifikan
$X_6$	1,660	1,796	Tidak signifikan
$X_7$	-5,712	1,796	<b>Signifikan</b>
$X_8$	2,322	1,796	<b>Signifikan</b>
$X_9$	9,765	1,796	<b>Signifikan</b>
$X_{10}$	-0,582	1,796	Tidak signifikan

Berdasarkan Tabel 6, terlihat bahwa terdapat lima variabel independen yang memiliki nilai  $t_{hitung} > t_{tabel}$  sehingga variabel tersebut disimpulkan berpengaruh signifikan terhadap variabel dependen, yaitu  $X_2, X_3, X_7, X_8$ , dan  $X_9$ .

### 3.8 Spesifikasi PCR

*Principal Component Regression* (PCR) dilakukan dengan melakukan perhitungan nilai eigen terlebih dahulu. Nilai eigen dapat dilihat pada tabel berikut ini:

Tabel 7. Nilai Eigen

Komponen	Nilai Eigen	Proporsi Kumulatif (%)
$K_1$	<b>5,054</b>	<b>50,5</b>
$K_2$	<b>2,181</b>	<b>72,3</b>
$K_3$	<b>1,150</b>	<b>83,8</b>
$K_4$	0,747	91,3
$K_5$	0,462	95,9
$K_6$	0,236	98,3
$K_7$	0,135	99,6
$K_8$	0,027	99,91
$K_9$	0,008	99,992
$K_{10}$	0,001	100

Tabel 7 Menunjukkan terdapat 3 komponen utama dengan nilai eigen lebih dari 1, yaitu  $K_1, K_2$ , dan  $K_3$  dengan proporsi kumulatif sebesar 83,8%. Hal ini menunjukkan bahwa tiga komponen utama tersebut sudah dapat mewakili 83,8% informasi dari 10 variabel asli. Selanjutnya ditentukan vektor eigen dari tiga komponen utama tersebut guna menentukan skor komponen yang dapat dilihat pada tabel ini.

Tabel 8. Vektor Eigen

	$K_1$	$K_2$	$K_3$
$Z_1$	0,186	0,489	-0,239
$Z_2$	-0,115	0,106	0,765
$Z_3$	-0,437	0,049	-0,072
$Z_4$	-0,439	0,052	-0,085
$Z_5$	-0,438	0,022	-0,094
$Z_6$	-0,436	0,034	-0,108
$Z_7$	-0,132	0,517	-0,213
$Z_8$	0,015	0,609	0,035
$Z_9$	-0,373	-0,236	-0,096
$Z_{10}$	0,171	-0,223	-0,518

Berdasarkan tabel di atas, diperoleh persamaan berikut yang akan digunakan untuk mendapatkan skor komponen utama.

$$K_1 = 0,186Z_1 - 0,115Z_2 - 0,437Z_3 - 0,439Z_4 - 0,438Z_5 - 0,436Z_6 - 0,132Z_7 + 0,015Z_8 - 0,373Z_9 + 0,171Z_{10} \quad (12)$$

$$K_2 = 0,489Z_1 + 0,106Z_2 + 0,049Z_3 + 0,052Z_4 + 0,022Z_5 + 0,034Z_6 + 0,517Z_7 + 0,609Z_8 - 0,236Z_9 - 0,223Z_{10} \quad (13)$$

$$K_3 = -0,239Z_1 + 0,765Z_2 - 0,072Z_3 - 0,085Z_4 - 0,094Z_5 - 0,108Z_6 - 0,213Z_7 + 0,035Z_8 - 0,096Z_9 - 0,518Z_{10} \quad (14)$$

Dari persamaan tersebut, akan didapatkan skor komponen sebagai berikut:

Tabel 9. Skor Komponen Utama

$K_1$	$K_2$	$K_3$
1,023	-1,116	-0,254
-1,103	-1,031	0,151
-1,978	-0,555	-1,924
-5,400	-1,592	-0,374
-0,086	0,301	0,173
1,082	0,569	-0,207
0,481	0,546	1,087
2,466	0,940	-0,240
1,310	0,967	-0,593
-0,172	1,092	-0,760
-0,050	1,594	0,518
1,844	1,811	1,270
-2,502	0,707	1,720
2,594	-0,883	-2,836
-0,574	0,372	0,161
2,546	-2,610	-0,484
-3,552	-1,588	0,449
2,412	1,349	0,278
-1,134	-1,524	0,846
2,578	-2,221	1,564
1,386	-0,455	0,515
-3,172	3,326	-1,059

Berdasarkan tabel di atas diperoleh model regresi komponen utama sebagai berikut:

$$Y = 34,368 - 0,774K_1 - 1,862K_2 - 0,039K_3 \quad (15)$$

Transformasi regresi komponen utama dengan mensubstitusi persamaan  $K_1, K_2$ , dan  $K_3$  maka akan diperoleh persamaan sebagai berikut:

$$Y = 34,368 - 1,045Z_1 - 0,139Z_2 + 0,249Z_3 + 0,245Z_4 + 0,301Z_5 + 0,278Z_6 - 0,851Z_7 - 1,147Z_8 + 0,731Z_9 + 0,303Z_{10} \quad (16)$$

Persamaan di atas merupakan persamaan dalam bentuk baku Z sehingga akan dilakukan transformasi kembali dalam bentuk variabel independen awal sehingga diperoleh persamaan, yaitu:

$$Y = 59,434 - 0,074X_1 - 0,029X_2 + 0,00011X_3 + 0,00011X_4 + 0,00012X_5 + 0,00012X_6 - 0,207X_7 - 0,103X_8 + 0,027X_9 + 0,50X_{10} \quad (17)$$

Tahap uji signifikansi parameter secara parsial untuk menguji hipotesis  $H_0: \beta_j = 0$  lawan  $H_1: \exists \beta_j \neq 0$  untuk  $j = 1, 2, \dots, 10$  dan diperoleh hasil  $t_{hitung}$  sebagai berikut:

Tabel 10. Uji Signifikansi Metode PCR

Peubah	$t_{hitung}$	$t_{tabel}$	Keterangan
$X_1$	-1,859	1,796	<b>Signifikan</b>
$X_2$	-0,163	1,796	Tidak signifikan
$X_3$	6,490	1,796	<b>Signifikan</b>
$X_4$	4,557	1,796	<b>Signifikan</b>
$X_5$	13,842	1,796	<b>Signifikan</b>
$X_6$	16,572	1,796	<b>Signifikan</b>
$X_7$	-0,889	1,796	Tidak signifikan
$X_8$	-1,694	1,796	Tidak signifikan
$X_9$	2,296	1,796	Tidak signifikan
$X_{10}$	0,353	1,796	Tidak signifikan

Berdasarkan tabel di atas, dapat terlihat bahwa terdapat lima variabel independen yang memiliki pengaruh signifikan terhadap variabel dependen, yaitu  $X_1, X_3, X_4, X_5$ , dan  $X_6$ .

### 3.9 Pemilihan Model Terbaik

Tabel 11 menunjukkan nilai Adj.  $R^2$  PLSR lebih besar dari PCR serta nilai RMSE dan MAE dari PLSR lebih kecil daripada PCR sehingga dapat disimpulkan model PLSR merupakan

model yang lebih baik untuk digunakan menangani multikolinearitas prevalensi stunting di Nusa Tenggara Timur 2022.

Tabel 11. Kriteria Kebaikan

Metode	Adj. $R^2$	RMSE	MAE
PLSR	0,741	2,783	1,910
PCR	0,322	4,742	3,346

### Interpretasi Hasil

Metode terbaik PLSR mengidentifikasi lima variabel signifikan terhadap prevalensi stunting di NTT. Interpretasi hasil mengacu pada persamaan akhir PLSR, yaitu persamaan 10. Rata-rata konsumsi protein per kapita ( $X_2$ ) berkoefisien -1,055 menunjukkan peningkatan 1 gram konsumsi protein menurunkan 1,055% prevalensi stunting, sesuai temuan Sari (2016) tentang kekurangan protein mengakibatkan stunting. Jumlah balita yang mendapatkan imunisasi DPT-HB ( $X_3$ ) dengan koefisien -0,829 menunjukkan peningkatan 1 balita yang diimunisasi menurunkan 0,829% prevalensi stunting, selaras dengan penelitian Pardede (2022). Selanjutnya, IPM ( $X_7$ ) dengan koefisien -3,804, kenaikan 1 satuan IPM menurunkan 3,804% prevalensi stunting, sejalan dengan penelitian Sari (2024). Lalu persentase rumah tangga dengan akses sumber air minum layak ( $X_8$ ) berkoefisien 1,958, setiap kenaikan 1% akses sumber minum air layak meningkatkan 1% prevalensi stunting, bertentangan dengan Laksono (2021) yang menjelaskan bahwa peningkatan akses sumber air minum layak menurunkan prevalensi stunting. Perbedaan ini bisa terjadi karena kontaminasi bakteri E. Coli sumber air di NTT Agusanatemy (2022), serta kurangnya kesadaran warga NTT untuk menjaga kebersihan sumber mata air (Asyara, 2022). Terakhir, ada jumlah penduduk miskin ( $X_9$ ) dengan koefisien 4,111 menunjukkan setiap kenaikan 1 jiwa penduduk miskin meningkatkan 4,111% prevalensi stunting, sejalan dengan temuan dari Karyati (2022).

## 4. KESIMPULAN DAN SARAN

Metode PLSR lebih baik dibandingkan PCR dalam menangani multikolinearitas prevalensi stunting di provinsi Nusa Tenggara Timur 2022. PLSR mengidentifikasi lima variabel signifikan yang memengaruhi prevalensi stunting



rata-rata konsumsi protein per kapita sehari, jumlah balita yang pernah mendapatkan imunisasi DPT-HB, IPM, persentase rumah tangga berdasarkan sumber air minum layak, serta jumlah penduduk miskin.

Pemerintah Provinsi NTT diharapkan fokus pada kelima faktor tersebut guna menurunkan prevalensi stunting. Penelitian selanjutnya disarankan membandingkan metode PLSR algoritma NIPALS dengan algoritma SIMPLS atau metode lain dalam mengatasi multikolinearitas.

## DAFTAR PUSTAKA

- Agusanatery, & Fahira, S.(2022). Analisis Kualitas Sumber Air Minum(Air Sumur) Masyarakat di Kecamatan Kota Raja Kota Kupang Berdasarkan Jumlah Bakteri Escherichia Coli. *Jurnal Ilmiah Pendidikan Sains dan Terapan*, 2(3), p.172-181.
- Asyakra, M.L., Belia, N., Anggary, O.M.P., Rangkuti, R.A. & Syahfitri, V.N. (2022). Krisis Penyediaan Air Bersih di Provinsi Nusa Tenggara Timur. *Konferensi Nasional Ilmu Administrasi*, p.150-155.
- Duan,dkk. (2022). *Use of Partial Least Squares Regression to Identify Factors Controlling Rice Yield in Southern China*. *Agronomy Journal*, 112(3), p.1502 1516.
- Effendi, M., Ardhyatitra, R., Angelina, S.G., & Ohyver, M. (2023). *Predict Farmer Exchange Rate in the Food Crop Sector Using Principal Component Regression. Enthusiastic: International Journal of Applied Statistics and Data Science*, 3(1), p.74-84.
- Hair Jr, J.F., dkk (2021). *Artial Least Squares Structural Equation Modeling (PLS SEM) Using R: A Workbook*. Springer.
- Karyati, Y. & Julia, A. (2021). Pengaruh Jumlah Penduduk Miskin, Laju Pertumbuhan Ekonomi, dan Tingkat Pendidikan terhadap Jumlah Stunting di 10 Wilayah Tertinggi Indonesia Tahun 2010-2019. *Jurnal Riset Ilmu Ekonomi dan Bisnis*, 1(2), p.101-108.
- Laksono, A.D., & Wulandari, R.D. (2021). Dampak Stunting dalam Perekonomian di Indonesia. *Jurnal Inovasi Sektor Publik (JISP)*, 1(1), p.185-198.
- Nugraha, J., & Ali, R.G. (2019). Penerapan Metode Regresi Ridge dalam Mengatasi Masalah Multikolinearitas pada Kasus Indeks Pembangunan Manusia di Indonesia Tahun 2017. *Prosiding Sendika*, 5(2), p.239-248.
- Pardede, Tesa Trilonika, Sumargo, Bagus, & Rahayu, Widyanti (2022). Penerapan Regresi Least Absolute Shrinkage and Selection Operator (LASSO) untuk Mengidentifikasi Variabel yang Berpengaruh terhadap Kejadian Stunting di Indonesia. *Jurnal Statistika dan Aplikasinya*, 6(1), p.37-48.
- Santi, V.M., Notodiputro.K.A., Sartono, B. & Rahayu, W. (2021). *Generalized Linear Mixed Models by Penalized Lasso in Modelling the Scores of Indonesian Students*. *Journal of Physics: Conference Series*, 1869 (012140), p.1-8.
- Santi, V.M. (2011). Penerapan *Partial Least Square Regression* (PLSR) pada Pemodelan Kalibrasi Senyawa Aktif Kurkumin. *Prosiding Seminar Nasional Penelitian dan PKM Sains, Teknologi, dan Kesehatan*, 2 (1), p.535-544.
- Santi, V.M., Hayati, R., & Sumargo, B. (2024). *The Group Selection of Variables that Effected Science Scores of Indonesia's PISA using Group LASSO*. *AIP Conference Proceedings*, 29282(1), 02008-1- 02008-7. AIP Publishing.
- Santi, V.M., Notodiputro.K.A., & Sartono, B. (2019). *Variable Selection Methods Applied to the Mathematics Scores of Indonesian Students Based on Convex Penalized Likelihood*. *Journal of Physics: Conference Series*, 1402 (077096), p.1-6.
- Sari, E.M., Juffrie, M., & Nurani, N. (2016). Asupan Protein, Kalsium, dan Fosfor pada Anak Stunting dan Tidak Stunting Usia 24-59 Bulan. *Jurnal Gizi Klinik Indonesia*, 12(4), p.152-159.
- Sari, M.A., Rumbia, W.A., & Asizah, N. (2024). Analisis Pengaruh Kemiskinan dan Indeks Pembangunan Manusia terhadap Stunting di 34 Provinsi di Indonesia. *Jurnal Ekonomi, Kenangan, dan Bisnis Syariah*, 6(10), 7241-7253.
- World Health Organization* (2024). *Malnutrition*. <https://www.who.int/news-room/fact-sheets/detail/malnutrition>