

ANALISIS FAKTOR KEMISKINAN DI PROVINSI SUMATERA UTARA BERDASARKAN REGRESI KOMPONEN UTAMA

Analysis of Poverty Factors in North Sumatera Based on Principal Component Regression

Najwa Khoir Aldawiyah¹, Aprillia Astuti², Rizky Dwi Kurnia³, Nadinta Kasih Amalia⁴, Dita Amelia^{5*},
M. Fariz Fadillah Mardianto⁶, Elly Ana⁷

^{1,2,3,4,5,6,7}Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Airlangga
Jl. Dr. Ir. H. Soekarno, Surabaya, 60115, Jawa Timur, Indonesia

E-mail Corresponding Author: dita.amelia@fst.unair.ac.id

Abstrak: Kemiskinan merupakan permasalahan terkait kesejahteraan masyarakat yang serius dan menjadi indikator keberhasilan ekonomi dari suatu negara. Provinsi Sumatera Utara merupakan salah satu provinsi dengan jumlah penduduk miskin terbanyak dengan total 1,2 juta jiwa di Pulau Sumatera pada tahun 2022. Tujuan penelitian ini yaitu menangani masalah multikolinearitas pada faktor-faktor yang mempengaruhi kemiskinan di Provinsi Sumatera Utara dengan menggunakan analisis komponen utama. Data yang diperoleh merupakan data yang didapatkan dari Badan Pusat Statistik Provinsi Sumatera Utara. Terdapat 8 variabel prediktor yang digunakan dan terbentuk 3 komponen utama dengan keragaman total sebesar 84,5%. Komponen utama yang terbentuk kemudian diregresikan dan diperoleh persamaan $Zscore(Y) = 3,47 \times 10^{-16} + 0,598PC_1 + 0,396PC_2 - 0,419PC_3$. Model regresi tersebut terbebas dari masalah multikolinearitas dan ketiga komponen secara signifikan berpengaruh terhadap jumlah penduduk miskin di Provinsi Sumatera Utara.

Kata Kunci: Analisis Komponen Utama, Faktor Kemiskinan, Regresi Komponen Utama, Sumatera Utara

Abstract: Poverty is a serious public welfare issue and an indicator of the economic success of a country. North Sumatera Province is one of the provinces with the highest number of poor people, with a total of 1268.9 million people on the island of Sumatera in 2022. The purpose of this research is to resolve the multicollinearity problem in the factors that affect poverty in North Sumatra Province using principal component analysis. Data was obtained from the official website of the Badan Pusat Statistik of North Sumatera Province. There were 8 predictor variables used and 3 main components were formed with a total diversity of 84.5%. The main components formed were then regressed and obtained the equation $Zscore(Y) = 3.47 \times 10^{-16} + 0.598PC_1 + 0.396PC_2 - 0.419PC_3$. The regression model is free from multicollinearity problems, and the three components significantly affect the number of poor people in North Sumatra Province.

Keywords: Principal Component Analysis, Poverty Factors, Poverty Rates, Principal Component Regression, North Sumatera

1. PENDAHULUAN

Kemiskinan dapat diartikan sebagai keadaan saat sebagian besar penduduk tidak mampu memenuhi kebutuhan dasar hidupnya, seperti sandang, pangan, dan papan. Isu kemiskinan dapat mencakup segala aspek di kehidupan manusia, meski sering kali tidak disadari keberadaannya sebagai suatu permasalahan. Provinsi Sumatera Utara merupakan salah satu provinsi dengan jumlah penduduk miskin terbanyak dengan total 1,2 juta jiwa di Pulau Sumatera [1]. Faktor-faktor sosial, ekonomi, dan lingkungan yang berdampak pada tingkat kemiskinan di Sumatera Utara sangat kompleks dan saling terkait. Hal tersebut menjadi permasalahan di Provinsi Sumatera Utara. Oleh karena itu, diperlukan analisis menyeluruh untuk memahami faktor-faktor penyebab banyaknya jumlah penduduk miskin di Sumatera Utara. Dengan mereduksi faktor-faktor yang berpengaruh, salah satu pendekatan yang dapat diterapkan yakni menggunakan metode Analisis Komponen Utama (AKU). AKU adalah metode statistika yang bertujuan untuk mengurangi dimensi data tanpa mengurangi informasi dan karakteristik data secara signifikan. Selain berfungsi untuk mengurangi faktor-faktor, Analisis Komponen Utama juga dapat diterapkan untuk mengatasi permasalahan multikolinieritas dalam Analisis Regresi Linier Berganda [2].

Penelitian yang pernah dilakukan oleh [3] dengan menggunakan regresi komponen utama pada data impor beras di Sulawesi Utara pada tahun 2006-2015. Hasil analisis diperoleh seluruh variabel mempengaruhi impor beras di Sulut sebesar 90,36%. Penelitian lainnya oleh [4] dengan menggunakan *Principal Component Regression* untuk menganalisis faktor yang mempengaruhi Indeks Pembangunan Manusia. Didapatkan hasil terdapat 4 komponen utama yang mewakili komponen lainnya, untuk regresi komponen utama dengan memperhatikan proporsi kumulatif > 80% diperoleh Indeks Pembangunan Manusia di Sulawesi Selatan dipengaruhi oleh semua variabel yang terlibat yaitu sebesar 95,7%. Dalam penelitian mengenai kemiskinan yang dilakukan [5], dengan menggunakan metode regresi linier berganda didapatkan hasil secara parsial PDRB dan IPM memiliki pengaruh negatif dan signifikan terhadap jumlah penduduk miskin di Sumatera Utara pada tahun 2002-2017 dan secara simultan variabel PDRB dan Inflasi memiliki pengaruh signifikan terhadap jumlah penduduk miskin di Sumatera Utara pada tahun 2002-2017. Penelitian lain yang dilakukan oleh [6] mengenai “analisis faktor-faktor yang mempengaruhi jumlah penduduk miskin di kabupaten dan kota Provinsi Banten”, metode yang digunakan ialah data panel dengan model FEM. Hasil penelitian menunjukkan bahwa variabel jumlah pengangguran dan PDB per kapita memiliki dampak positif signifikan, sementara variabel tingkat pendidikan memiliki dampak negatif yang signifikan pada jumlah penduduk miskin di Kabupaten Banten selama tahun 2010 hingga 2015.

Pada penelitian sebelumnya, hanya dilibatkan tiga dan empat variabel yang dianggap berpengaruh terhadap jumlah penduduk miskin, sehingga penelitian ini akan menambahkan faktor lainnya yang dianggap berpengaruh terhadap jumlah penduduk miskin di Sumatera Utara dengan menggunakan regresi komponen utama yang merupakan gabungan antara analisis komponen utama dan analisis regresi. Metode ini dipilih karena melibatkan sejumlah faktor atau variabel yang dapat menyebabkan terjadinya multikolinearitas di antara variabel. Variabel-variabel tersebut saling berkorelasi sehingga dapat mengakibatkan peningkatan *standard error* dan berpotensi memengaruhi hasil analisis.

Berdasarkan penjelasan tersebut, penelitian ini bertujuan untuk menerapkan analisis regresi komponen utama dalam konteks kemiskinan di Sumatera Utara, guna merinci faktor-faktor yang paling berkontribusi terhadap fenomena ini dan pada upaya pencapaian *Sustainable Development Goals* (SDGs) pertama, yakni pengentasan terhadap kemiskinan.

2. METODOLOGI

2.1. Data dan Sumber Data

Data yang dianalisis merupakan data sekunder merupakan data sekunder yang diperoleh dari laman resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data tersebut merupakan data *cross-section* dengan populasi dan sampel yang digunakan diduga memiliki korelasi yang signifikan dengan kemiskinan di Provinsi Sumatera Utara pada tahun 2022. Oleh karena itu, ada total 32 kabupaten dan kota di Provinsi Sumatera Utara yang menjadi observasi dalam penelitian ini. Dalam penelitian ini jumlah variabel yang digunakan yakni sebanyak 9 variabel dengan 8 variabel prediktor dan 1 variabel respon. Adapun rincian variabel dalam penelitian ini yang disajikan dalam Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan	Skala Data
Y	Jumlah Penduduk Miskin	Jiwa	Rasio
X1	Indeks Pembangunan Manusia (IPM)	%	Interval
X2	Laju pertumbuhan penduduk	%	Rasio
X3	Kepadatan penduduk	Jiwa/km ²	Rasio
X4	Laju pertumbuhan Produk Domestik Regional Bruto (PDRB) per kapita atas dasar harga konstan	%	Rasio
X5	Rata-rata lama sekolah	Tahun	Rasio
X6	Upah minimum	Rupiah	Interval
X7	PDRB atas harga konstan	Rupiah	Rasio
X8	Tingkat pengangguran terbuka	%	Rasio

2.2. Langkah Analisis Data

Langkah-langkah analisis komponen utama pada faktor penyebab kemiskinan di Sumatera Utara adalah sebagai berikut.

1. Mengumpulkan variabel dan data yang diperlukan dari berbagai sumber resmi.
2. Menggunakan statistika deskriptif untuk mendapatkan informasi umum tentang karakteristik data. Statistika deskriptif yang digunakan yakni rata-rata, standar deviasi, nilai minimum, dan nilai maksimum.
3. Melakukan standarisasi data. Standarisasi data dilakukan sebab variabel-variabel dalam penelitian ini memiliki satuan yang berbeda. Transformasi ke bentuk standar dapat dilakukan dengan menggunakan rumus berikut.

$$Z_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (1)$$

Dengan Z_i adalah bentuk standar dari data ke- i , x_i adalah data ke- i , \bar{x} adalah rata-rata data, dan σ_x merupakan simpangan baku dari data [7].

4. Uji regresi linear berganda dan uji asumsi multikolinearitas. Regresi linier berganda adalah suatu model statistika yang diaplikasikan guna mengetahui hubungan linear antara satu variabel respon dengan dua atau lebih variabel terikat [8]. Model regresi linear berganda dalam bentuk vektor dapat dituliskan sebagai berikut [9].

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

Dengan \mathbf{y} adalah vektor random yang merupakan variabel respon, \mathbf{X} adalah matrik rancangan, $\boldsymbol{\beta}$ adalah parameter regresi yang berisi $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ dengan k adalah banyaknya prediktor, dan $\boldsymbol{\varepsilon}$ merupakan vektor galat. Untuk mengestimasi nilai $\boldsymbol{\beta}$, digunakan metode OLS atau *Ordinary Least Square* yaitu dengan meminimumkan jumlah kuadrat galat sehingga didapatkan Persamaan (3) yang selanjutnya didapatkan estimasi dari $\boldsymbol{\beta}$ yakni $\hat{\boldsymbol{\beta}}$ seperti pada Persamaan (4).

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (3)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (4)$$

Maka, persamaan regresi yang terbentuk adalah:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (5)$$

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (6)$$

Sementara itu, dalam penerapannya, salah satu asumsi yang harus dipenuhi saat melakukan uji regresi linear berganda adalah tidak terdapat korelasi yang kuat antar variabel prediktor atau yang disebut sebagai multikolinearitas. Untuk mendeteksi adanya multikolinearitas antar variabel prediktor dapat melihat nilai *Variance Inflation Factor* (VIF) dan *Tolerance*. Nilai VIF dapat dihitung menggunakan rumus berikut.

$$VIF_i = \frac{1}{1 - R_i^2} \quad (7)$$

Dengan R_i^2 adalah koefisien determinasi dari variabel ke i terhadap variabel yang lain dengan $i = 1, 2, \dots, k$. Sedangkan nilai *Tolerance* didapatkan dari rumus berikut.

$$Tolerance = \frac{1}{VIF} \quad (8)$$

Suatu model regresi dikatakan terdapat multikolinearitas apabila terdapat variabel dengan nilai VIF lebih dari 10 dan atau *Tolerance* kurang dari 0,1.

5. Uji korelasi antar variabel menggunakan Bartlett's Test. Pengujian dengan uji Bartlett dilakukan untuk mengevaluasi apakah matrik korelasinya merupakan matrik identitas dengan hipotesis alternatif yakni

matrik korelasi bukanlah matrik identitas [10]. Selanjutnya menurut Indra dan Robinson, uji ini memiliki hipotesis yakni sebagai berikut:

$H_0 : \mathbf{R} = 1$ (Variabel independen tidak berkorelasi)

$H_1 : \mathbf{R} \neq 1$ (Variabel independen berkorelasi)

dengan statistik uji:

$$X^2 = \left\{ n - 1 - \frac{2p + 5}{6} \right\} \ln|\mathbf{R}| \tag{9}$$

Dimana, n adalah jumlah observasi, p adalah variabel independen, sedangkan \mathbf{R} merupakan matriks korelasi dari masing-masing variabel independen. Dengan daerah kritis akan tolak H_0 jika nilai $X^2_{hitung} > X^2_{(\frac{1}{2}p(p-1))}$ dengan kesimpulan terdapat korelasi antar variabel independen. Dalam hal ini, analisis dapat dilanjutkan apabila variabel terbukti saling berkorelasi [11].

- Uji Kaiser-Meyer-Olkin (KMO) *Measure of Sampling Adequacy* (MSA) dilakukan untuk melihat apakah data sampel yang digunakan layak untuk dianalisis dengan komponen utama. Untuk mengukur hal tersebut digunakan persamaan berikut.

$$KMO = \frac{\sum_{i \neq j}^n \sum_{j=1}^m r_{ij}^2}{\sum_{i \neq j}^n \sum_{j=1}^m r_{ij}^2 + \sum_{i \neq j}^n \sum_{j=1}^m p_{ij}^2} \tag{10}$$

Dengan r order pertama korelasi antara pasangan item dan p adalah korelasi parsial yang bersesuaian [12]. Jika didapatkan nilai KMO lebih dari 0,5 maka sampel dikatakan layak untuk analisis lebih lanjut. Sedangkan untuk MSA a masing-masing item didapatkan dari persamaan berikut.

$$MSA_j = \frac{\sum_{k \neq j}^m r_{jk}^2}{\sum_{k \neq j}^m r_{jk}^2 + \sum_{k \neq j}^m p_{jk}^2} \tag{11}$$

Nilai MSA yang didapatkan adalah paling kecil 0,5 agar analisis komponen dapat dilanjutkan [10], [13].

- Perbaiki model dengan eliminasi variabel yang memiliki nilai MSA kurang dari 0,5 (jika diperlukan).
- Analisis komponen utama (AKU). AKU adalah sebuah metode statistik multivariat yang digunakan untuk mengurangi dimensi dari dataset dengan tujuan menghasilkan variabel-variabel baru yang bersifat ortogonal (tidak saling berkorelasi) dan mempertahankan sebanyak mungkin informasi dari variabilitas total data asli. Tujuan utama dari AKU adalah mengurangi jumlah variabel dalam dataset sambil tetap mempertahankan sebagian besar informasi yang terkandung dalam data asli. Perhitungan komponen utama diuraikan sebagai berikut. Misalkan terdapat vektor random \mathbf{y} sebanyak p buah sedemikian sehingga \mathbf{S} adalah matrik kovarian dari $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p$ dengan nilai akar ciri $\lambda_1 > \lambda_2 > \dots > \lambda_p$ [14]. Diketahui persamaan matrik

$$\mathbf{ASA}' = \mathbf{C}'\mathbf{S}\mathbf{C} = \begin{pmatrix} s_{z1}^2 & 0 & \dots & 0 \\ 0 & s_{z2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_{zp}^2 \end{pmatrix} \tag{12}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_p \end{pmatrix} \tag{13}$$

Dengan \mathbf{C} matrik orthogonal yang kolom-kolomnya merupakan vektor akar ciri dari \mathbf{S} yang telah dinormalisasi. Komponen utama yang pertama, yang dinyatakan dengan PC_1 merupakan kombinasi linear dari \mathbf{y}_i sedemikian sehingga

$$PC_1 = \mathbf{a}'_1\mathbf{y} = a_{11}\mathbf{y}_1 + a_{12}\mathbf{y}_2 + \dots + a_{1p}\mathbf{y}_p \tag{14}$$

$$PC_p = \mathbf{a}'_p \mathbf{y} = a_{p1}y_1 + a_{p2}y_2 + \dots + a_{pp}y_p$$

Dengan \mathbf{a}'_i dipilih untuk memaksimalkan rasio dari varian PC_p dengan variasi total, dengan batasan :

$$\sum_{i=1}^p a_{1i}^2 = 1 \tag{15}$$

Varians dari komponen utama yang terbentuk adalah diagonal utama dari matrik \mathbf{ASA}' yang juga merupakan nilai akar ciri dari matrik \mathbf{S} , sehingga

$$Var(PC_i) = s_{zi}^2 = \lambda_i \tag{16}$$

Berdasarkan Persamaan (16) , maka total variansi dari keseluruhan komponen dinyatakan dalam

$$\sum_{i=1}^p Var(PC_i) = \sum_{i=1}^p \lambda_i = tr(\mathbf{S}) \tag{17}$$

Dan proporsi keragaman yang dapat dijelaskan oleh k komponen pertama dinyatakan sebagai berikut

$$Proportion\ of\ Variance = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \tag{18}$$

9. Analisis regresi linear berganda menggunakan komponen utama yang terbentuk.
10. Interpretasi dan simpulan hasil analisis

3. HASIL DAN PEMBAHASAN

3.1. Statistik Deskriptif

Hasil uji statistika deskriptif pada variabel-variabel yang digunakan dalam penelitian ini, disajikan dalam Tabel 2 berikut.

Tabel 2. Statistika Deskriptif Variabel Penelitian

Variabel	Mean	Simpangan Baku	Minimum		Maksimum	
			Nilai	Kab/Kota	Nilai	Kab/Kota
Y	38,43	34,56	4,52	Pakpak Bharat	187,74	Medan
X1	71,79	4,41	62,93	Nias Barat	81,76	Medan
X2	1,27	0,41	0,50	Sibolga	2,44	Pakpak Bharat
X3	1044,08	1933,70	39,70	Pakpak Bharat	8865,91	Medan
X4	2,87	0,67	1,01	Nias Selatan	4,14	Langkat
X5	9,18	1,39	5,64	Nias	11,48	Medan
X6	2.748.923,39	228.589,26	2.522.609	Dairi, Nias Selatan, Pakpak Bharat, Samosir, Nias Uatra, Nias Barat	3.370.64 5	Medan
X7	17.637,35	30.123,98	909,62	Pakpak Bharat	165.120, 01	Medan
X8	4,65	2,72	0,26	Pakpak Bharat	9,36	Pematangsiantar

3.2. Regresi Linier Berganda dan Deteksi Multikolinearitas

Sebelum melakukan regresi linier berganda, seluruh variabel yang digunakan dalam penelitian ini terlebih dahulu ditransformasi menjadi bentuk baku. Dengan bantuan *software*, persamaan regresi yang terbentuk adalah sebagai berikut.

$$\begin{aligned} Zscore(Y) = & -7,73 * 10^{-17} + 0,068 * Zscore(X1) + 0,026 * Zscore(X2) \\ & -0,111 * Zscore(X3) + 0,082 * Zscore(X4) - 0,282 * Zscore(X5) \\ & -0,188 * Zscore(X6) + 1,056(X7) + 0,184 * Zscore(X8) \end{aligned} \tag{19}$$

Uji asumsi multikolinearitas umumnya dilakukan pada regresi linier berganda untuk mengetahui apakah di antara variabel prediktor terdapat korelasi yang kuat. Suatu data dikatakan memiliki gejala multikolinearitas ketika nilai VIF lebih dari 10 atau dengan nilai *Tolerance* kurang dari 0,1. Tabel 3 berikut menyajikan signifikansi dari masing-masing variabel, nilai VIF, dan *Tolerance* dari masing-masing variabel yang digunakan dalam penelitian ini.

Tabel 3. Hasil Uji Regresi Linier Berganda

Variabel	T-value	Sig.	Tolerance	VIF
Zscore(X1)	0,334	0,741	0,085	11,759
Zscore(X2)	0,273	0,787	0,401	2,491
Zscore(X3)	-0,969	0,342	0,267	3,748
Zscore(X4)	0,632	0,533	0,211	4,730
Zscore(X5)	-1,357	0,187	0,082	12,247
Zscore(X6)	-1,950	0,063	0,380	2,632
Zscore(X7)	9,709	0,000	0,298	3,351
Zscore(X8)	2,402	0,024	0,604	1,656

Berdasarkan Tabel 3, dengan menggunakan *alpha* 0,05 dari 8 variabel yang diteliti hanya 2 variabel yang memiliki *p-value* < 0,05 yakni variabel PDRB atas dasar harga konstan (X7) dan tingkat pengangguran terbuka (X8). Variabel tersebut secara statistik berpengaruh terhadap jumlah penduduk miskin. Selain itu, terdapat variabel IPM (X1) dan rata-rata lama sekolah (X5) yang memiliki nilai VIF > 10 dan nilai *Tolerance* < 0,1 sehingga dapat disimpulkan bahwa di antara variabel-variabel yang diduga mempengaruhi kemiskinan di Sumatera Utara terjadi multikolinearitas. Untuk menangani multikolinearitas tersebut, digunakan analisis komponen utama.

3.3. Hasil Pengujian Bartlett & KMO

Uji Bartlett dan KMO dilakukan untuk mengetahui kelayakan setiap variabel yang digunakan untuk dilakukan analisis lebih lanjut. Hipotesis nol yang diuji yaitu variabel-variabel yang digunakan belum mencukupi untuk dianalisis lebih lanjut. Hasil uji Bartlett's dan KMO disajikan pada Tabel 4 berikut.

Tabel 4. Hasil Pengujian Bartlett's & KMO

KMO MSA	Bartlett's Test of Sphericity. Sig.
0,584	0,000

Pada Tabel 4 hasil pengujian menggunakan *software*, nilai signifikansi uji Bartlett's sebesar 0,000 dan memenuhi daerah kritis *alpha* 0,05. Keputusan yang diambil yaitu menolak hipotesis nol sehingga variabel-variabel yang mempengaruhi kemiskinan di Sumatera Utara pada tahun 2022 sudah mencukupi persyaratan untuk dianalisis lebih lanjut. Selanjutnya, nilai KMO MSA yang diperoleh yaitu sebesar 0,584 dan lebih dari nilai minimum MSA yang dipersyaratkan sebesar 0,5. Oleh karena itu, kumpulan variabel yang digunakan selanjutnya dapat dilakukan analisis MSA untuk mereduksi variabel mana saja yang belum layak untuk analisis komponen utama.

Tabel 5. Nilai MSA Variabel

Variabel	MSA
Zscore(X1)	0,686
Zscore(X2)	0,288
Zscore(X3)	0,523
Zscore(X4)	0,487
Zscore(X5)	0,516
Zscore(X6)	0,836
Zscore(X7)	0,593
Zscore(X8)	0,814

Tabel 5 menyajikan nilai MSA untuk masing-masing variabel penelitian. Berdasarkan Tabel 5 terlihat bahwa terdapat 2 variabel dengan nilai MSA kurang dari 0,5. Nilai MSA < 0,5 menunjukkan bahwa korelasi antar variabel tersebut tidak mencukupi untuk analisis lebih lanjut. Variabel dengan nilai MSA kurang dari 0,5 yakni Laju Perumbuhan Penduduk (X2) dan Laju Pertumbuhan PDRB Per Kapita atas Dasar Harga Konstan (X4).

3.4. Perbaikan Model

Perbaikan model dilakukan dengan mengeliminasi variabel-variabel yang memiliki nilai MSA < 0,5 secara bertahap. Variabel yang dieliminasi pada tahapan pertama adalah variabel dengan nilai MSA terendah, yakni variabel laju pertumbuhan penduduk. Setelah proses eliminasi yang pertama, berdasarkan Tabel 6 diperoleh nilai MSA untuk variabel lain lebih dari 0,5 sehingga dari 8 variabel yang digunakan hanya 7 yang dapat melanjutkan ke proses analisis selanjutnya.

Tabel 6. Nilai MSA Variabel

Variabel	MSA
Zscore(X1)	0,689
Zscore(X3)	0,706
Zscore(X4)	0,662
Zscore(X5)	0,611
Zscore(X6)	0,767
Zscore(X7)	0,715
Zscore(X8)	0,854

Tabel 7. Hasil Pengujian Bartlett's & KMO

KMO MSA	Bartlett's Test of Sphericity. Sig.
0,703	0,000

Berdasarkan Tabel 7 diperoleh nilai KMO sebesar 0,703. Sementara itu, hasil uji Bartlett memperoleh signifikansi sebesar 0,000 sehingga dengan menggunakan *alpha* 5% diperoleh kesimpulan bahwa 7 variabel tersebut telah memenuhi persyaratan untuk analisis lebih lanjut.

3.5. Analisis Komponen Utama

Analisis komponen utama merupakan metode statistik yang digunakan untuk mereduksi dan menyederhanakan variabel yang diamati dan membentuk variabel baru yang selanjutnya disebut komponen utama [15]. Analisis komponen utama ini bertujuan untuk mengatasi adanya multikolinearitas antar variabel bebas. Terdapat 3 kriteria dalam memilih komponen utama yaitu nilai kaar ciri lebih dari 1, melihat bentuk kurva *scree* plot, dan variansi kumulatif komponen yang mencapai 70% sampai 80% [16]. Komponen utama yang terbentuk diharapkan dapat menyederhanakan variabel tetapi mampu menerangkan keragaman total yang maksimal [17]. Parameter yang dilihat pada analisis komponen utama yaitu nilai komunalitas variabel, total variansi, dan nilai *factor loading* yang dirotasi [18].

Tabel 8. Nilai Komunalitas Variabel

	Initial	Nilai Komunalitas
Zscore(X1)	1,000	0,953
Zscore(X3)	1,000	0,888
Zscore(X4)	1,000	0,917
Zscore(X5)	1,000	0,968
Zscore(X6)	1,000	0,851
Zscore(X7)	1,000	0,798
Zscore(X8)	1,000	0,547

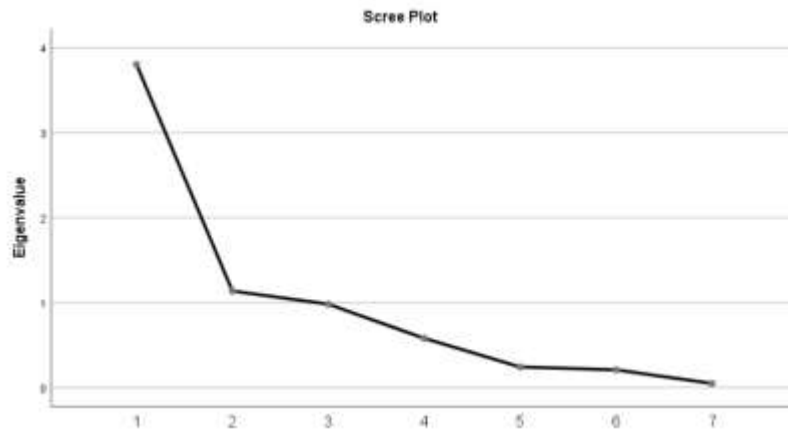
Tabel 8 menyajikan nilai komunalitas dari masing-masing variabel. Komunalitas menyatakan jumlah nilai keragaman yang bisa dijelaskan oleh komponen yang terbentuk [19]. Pada tabel di atas, keseluruhan variabel faktor kemiskinan yang digunakan dalam penelitian ini memiliki nilai komunal di atas 0,5. Hal ini berarti komponen yang terbentuk mampu menjelaskan variabel-variabel dengan keragaman di atas 50%. Nilai komunal tertinggi mencapai 96,8% oleh variabel Zscore(X5) yaitu rata-rata lama sekolah. Sementara nilai komunal terendah sebesar 0,547 pada variabel Zscore(X8) yang merupakan tingkat pengangguran terbuka di setiap kabupaten/kota di Sumatera Utara.

Selanjutnya, untuk menentukan banyaknya komponen utama maka perlu melihat total keragaman yang dijelaskan oleh masing-masing komponen yang memiliki akar ciri lebih dari 1 dan melihat bentuk kurva pada

grafik *scree plot* [19]. Tabel 9 menunjukkan nilai akar ciri untuk masing-masing komponen dan Gambar 1 menunjukkan grafik *scree plot* pembentukan komponen.

Tabel 9. Hasil Ekstraksi Komponen Baru

Komponen	Total	Initial Eigenvalues	
		% Varians	Kumulatif %
1	3,804	54,336	54,336
2	1,137	16,238	70,574
3	0,981	14,016	84,590
4	0,580	8,279	92,869
5	0,242	3,464	96,333
6	0,208	2,978	99,311
7	0,048	0,689	100,000



Gambar 1. Grafik Scree Plot

Berdasarkan Tabel 9 dan Gambar 1 terdapat 2 komponen yang memiliki nilai akar ciri lebih dari 1 dengan keragaman kumulatif sebesar 70,5%. Namun, komponen ketiga memiliki nilai akar ciri yang tidak jauh dari 1 dan keragaman yang dapat dijelaskan juga cukup besar sehingga keragaman kumulatif dari ketiga komponen mencapai 84,5%. Untuk tujuan analisis dan interpretasi lebih lanjut, maka banyaknya komponen yang akan digunakan dalam penelitian ini adalah sejumlah 3 komponen yang dapat menjelaskan keragaman total sebesar 84,5%.

Untuk mengidentifikasi *factor loadings* pada masing-masing komponen baru yang terbentuk, digunakan metode rotasi VARIMAX. Rotasi VARIMAX merupakan metode rotasi orthogonal yang digunakan untuk menyederhanakan kolom-kolom pada matriks komponen utama yang terbentuk [20].

Tabel 10. Matrik Komponen yang Dirotasi

	Komponen		
	1	2	3
Zscore(X1)	0,328	0,888	0,239
Zscore(X3)	0,749	0,506	-0,267
Zscore(X4)	0,116	0,287	0,906
Zscore(X5)	0,155	0,956	0,177
Zscore(X6)	0,665	0,039	0,638
Zscore(X7)	0,841	0,126	0,275
Zscore(X8)	0,666	0,272	0,171

Berdasarkan Tabel 6, terlihat bahwa masing-masing variabel memiliki nilai *factor loadings* di atas 0,5 terhadap ketiga komponen utama. Persamaan komponen utama yang terbentuk adalah sebagai berikut.

$$\begin{aligned}
 PC1 = & 0,328 * Zscore(X1) + 0,749 * Zscore(X3) + 0,116 * Zscore(X4) \\
 & + 0,155 * Zscore(X5) + 0,665 * Zscore(X6) + 0,841 * Zscore(X7) \\
 & + 0,666 * Zscore(X8)
 \end{aligned}
 \tag{20}$$

$$PC2 = 0,888 * Zscore(X1) + 0,506 * Zscore(X3) + 0,287 * Zscore(X4) + 0,956 * Zscore(X5) + 0,039 * Zscore(X6) + 0,126 * Zscore(X7) + 0,272 * Zscore(X8) \quad (21)$$

$$PC3 = 0,239 * Zscore(X1) - 0,267 * Zscore(X3) + 0,906 * Zscore(X4) + 0,177 * Zscore(X5) + 0,638 * Zscore(X6) + 0,275 * Zscore(X7) + 0,171 * Zscore(X8) \quad (22)$$

3.6. Interpretasi Lanjutan

Berdasarkan persamaan komponen utama yang diperoleh yaitu pada Persamaan (20), (21), dan (22), selanjutnya dilakukan analisis regresi menggunakan bantuan *software*. Persamaan regresi yang diperoleh yakni sebagai berikut.

$$Zscore(Y) = 3,47 * 10^{-16} + 0,598 * PC1 + 0,396 * PC2 - 0,419 * PC3 \quad (23)$$

Tabel 11 menunjukkan nilai VIF dan *Tolerance* dari tiap-tiap komponen utama saat diregresikan dengan variabel respon Y. Diperoleh nilai VIF sama dengan 1 untuk semua komponen. Selain itu, secara statistik, setiap komponen memiliki pengaruh yang signifikan terhadap jumlah penduduk miskin dengan nilai *p-value* < 0,05 sehingga dapat disimpulkan bahwa komponen utama yang terbentuk telah menangani masalah multikolinearitas pada model regresi faktor kemiskinan di Provinsi Sumatera Utara. Berikut disajikan nilai signifikansi uji parsial, VIF, dan *Tolerance* pada model regresi komponen utama.

Tabel 11. Hasil Uji Regresi Komponen Utama

	T-Value	Sig.	Tolerance	VIF
PC1	5,78	0,000	1,000	1,000
PC2	3,83	0,001	1,000	1,000
PC3	-4,05	0,000	1,000	1,000

Nilai *R-Squared* yang diperoleh model adalah sebesar 69%. Artinya, sebesar 69% jumlah penduduk miskin dapat dijelaskan oleh komponen utama 1, komponen utama 2, dan komponen utama 3. Sisanya, dijelaskan oleh faktor lainnya. Berikutnya persamaan regresi yang diperoleh diuji untuk mengetahui apakah model tersebut secara statistik berarti atau tidak dengan hipotesis sebagai berikut [21].

H_0 : $\delta_k = 0$ (Secara keseluruhan, model regresi tidak berarti)

H_0 : $\delta_k \neq 0$ (Secara keseluruhan, model regresi berarti)

Tabel 12. Hasil Uji Signifikansi Model Regresi

Sumber	db	Adj SS	Adj MS	F	P-Value
Regresi	3	22,091	7,364	21,552	0,000
Galat	29	9,909	0,342		
Total	32	32,00			

Berdasarkan Tabel 12 diperoleh *P-Value* regresi sebesar 0,000. Dengan menggunakan *alpha* 5% maka keputusan yang didapatkan adalah H_0 ditolak sehingga secara keseluruhan model regresi dengan PC1, PC2, dan PC3 sebagai variabel prediktor dan jumlah penduduk miskin sebagai variabel respon secara statistik signifikan atau berarti. Nilai MSE yang diperoleh juga relatif kecil yakni sebesar 0,342.

Langkah berikutnya adalah mengubah Persamaan (23) ke dalam bentuk variabel penelitian yang telah distandarisasi dan didapatkan persamaan regresi sebagai berikut.

$$Zscore(Y) = 3,47 * 10^{-16} + 0,447 * Zscore(X1) - 0,760 * Zscore(X3) - 0,196 * Zscore(X4) + 0,397 * Zscore(X5) - 0,145 * Zscore(X6) + 0,437 * Zscore(X7) + 0,434 * Zscore(X8) \quad (24)$$

Berdasarkan Persamaan (24) berikut rincian interpretasi dari persamaan regresi tersebut:

- Angka kemiskinan akan bernilai $3,47 * 10^{-16}$ apabila semua variabel bernilai 0.
- Angka kemiskinan akan bertambah sebesar 0,447 untuk setiap pertambahan satu satuan Indeks Pembangunan Manusia atau IPM (X1) apabila variabel bebas lainnya bernilai tetap.

- Angka kemiskinan akan berkurang sebesar 0,760 untuk setiap penambahan satu satuan kepadatan penduduk (X3) apabila variabel bebas lainnya bernilai tetap.
- Angka kemiskinan akan berkurang sebesar 0,196 untuk setiap penambahan satu satuan laju pertumbuhan PDRB per kapita atas dasar harga konstan (X4) apabila variabel bebas lainnya bernilai tetap.
- Angka kemiskinan akan bertambah sebesar 0,397 untuk setiap penambahan satu satuan rata-rata lama sekolah (X5) apabila variabel bebas lainnya bernilai tetap.
- Angka kemiskinan akan berkurang sebesar 0,145 untuk setiap penambahan satu satuan upah minimum (X6) apabila variabel bebas lainnya bernilai tetap.
- Angka kemiskinan akan bertambah sebesar 0,437 untuk setiap penambahan satu satuan PDRB harga konstan (X7) apabila variabel bebas lainnya bernilai tetap.
- Angka kemiskinan akan bertambah sebesar 0,434 untuk setiap penambahan satu satuan tingkat pengangguran terbuka (X8) apabila variabel bebas lainnya bernilai tetap.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan mengenai penggunaan metodologi Analisis Komponen Utama untuk menganalisis faktor kemiskinan di Provinsi Sumatera Utara dapat disimpulkan bahwa:

1. Dari 8 variabel yang digunakan dalam penelitian ini dapat dibentuk 3 komponen utama yang saat komponen-komponen tersebut diregresikan, model tersebut terbebas dari masalah multikolinearitas. Hal ini dibuktikan dari nilai VIF yang kurang dari 10 dan ketiga komponen tersebut secara statistik signifikan mempengaruhi variabel respon jumlah penduduk miskin di Provinsi Sumatera Utara.

2. Model taksiran regresi komponen utama untuk faktor penyebab kemiskinan tersebut adalah :

$$Y = 3,47 * 10^{-16} + 0,598 * PC1 + 0,396 * PC2 - 0,419 * PC3$$

3. Model taksiran regresi komponen utama yang telah diubah menjadi model regresi variabel bebas X terhadap variabel tak bebas dari angka kemiskinan tersebut adalah :

$$\begin{aligned} Zscore(Y) = & 3,47 * 10^{-16} + 0,447 * Zscore(X1) - 0,760 * Zscore(X3) \\ & - 0,196 * Zscore(X4) + 0,397 * Zscore(X5) - 0,145 * Zscore(X6) \\ & + 0,437 * Zscore(X7) + 0,434 * Zscore(X8) \end{aligned}$$

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Persentase Penduduk Miskin Menurut Kab/Kota (Persen), 2020-2022," Sumatera Utara, 2022.
- [2] M. S. Noya, V. Delsen, A. Z. Wattimena, and S. D. Saputri, "PENGUNAAN METODE ANALISIS KOMPONEN UTAMA UNTUK MEREDUKSI FAKTOR-FAKTOR INFLASI DI KOTA AMBON," *Jurnal Ilmu Matematika dan Terapan*, vol. 11, no. 2, pp. 109–118, Dec. 2017.
- [3] M. Sriningsih, D. Hatidja, and J. D. Prang, "PENANGANAN MULTIKOLINEARITAS DENGAN MENGGUNAKAN ANALISIS REGRESI KOMPONEN UTAMA PADA KASUS IMPOR BERAS DI PROVINSI SULUT," *Jurnal Ilmiah Sains*, vol. 18, no. 1, pp. 18–24, 2018.
- [4] S. Susilawati and D. Didiharyono, "Application of Principal Component Regression in Analyzing Factors Affecting Human Development Index," *Jurnal Varian*, vol. 6, no. 2, pp. 199–208, May 2023, doi: 10.30812/varian.v6i2.2366.
- [5] P. S. M. J. Silaban, P. S. Br Sembiring, V. A. Br Sitepu, and J. P. Br Sembiring, "the Pengaruh IPM dan PDRB terhadap Jumlah Penduduk Miskin di Sumatera Utara Tahun 2002-2017," *Jesya (Jurnal Ekonomi & Ekonomi Syariah)*, vol. 4, no. 1, pp. 311–321, Dec. 2020, doi: 10.36778/jesya.v4i1.288.

- [6] R. Annisa and H. Sutjipto, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI JUMLAH PENDUDUK MISKIN DI KABUPATEN DAN KOTA PROVINSI BANTEN," *Tirtayasa EKONOMIKA*, vol. 12, no. 2, pp. 301–316, 2017.
- [7] R. S. Susanti, M. Syafri, and A. Hasin, "Penentuan Indikator Kesehatan Keluarga dengan Menggunakan Metode Analisis Komponen Utama (PCA)," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 11, no. 1, pp. 46–51, 2023.
- [8] F. Daniel, "MENGATASI PENCILAN PADA PEMODELAN REGRESI LINEAR BERGANDA DENGAN METODE REGRESI ROBUST PENAKSIR LMS," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 13, no. 3, pp. 145–156, Oct. 2019, doi: 10.30598/barekengvol13iss3pp145-156ar884.
- [9] A. C. Rencher and G. B. Schaalje, *LINEAR MODELS IN STATISTICS*, 2nd ed. Canada: John Wiley & Sons, Inc.
- [10] G. Haumahu and N. Lewaherilla, "PENERAPAN ANALISIS KOMPONEN UTAMA DALAM MEREDUKSI FAKTOR-FAKTOR PENYEBAB DIARE DI PROVINSI MALUKU THE APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO REDUCE DIARRHEA FACTORS IN MALUKU PROVINCE," *Mathematics & Applications (MAP Journal)*, vol. 2, no. 1, pp. 41–46, 2020.
- [11] Drajat Indra Purnama and Pardomuan Robinson Sihombing, "PERBANDINGAN ANALISIS KOMPONEN UTAMA DAN ROBUST PCA (ROBPCA) (STUDI KASUS: PADA ANALISIS DATA RATA-RATA PENGELUARAN PER KAPITA SEBULAN UNTUK KOMODITAS MAKANAN DI PROVINSI SULAWESI SELATAN)," *lppm bina bangsa*, vol. 1, no. 1, pp. 67–76, Mar. 2021.
- [12] U. Lorenzo-Seva and P. J. Ferrando, "MSA: The Forgotten Index for Identifying Inappropriate Items Before Computing Exploratory Item Factor Analysis," *Methodology*, vol. 17, no. 4, pp. 296–306, 2021, doi: 10.5964/meth.7185.
- [13] M. Faisal, S. Ali, and D. Ratna, "ANALISIS KOMPONEN UTAMA PADA DINAS KETENAGAKERJAAN BAGIAN PENEMPATAN DAN PERLUASAN KERJA MENCARI PEKERJAAN MENURUT GOLONGAN PEKERJAAN," *Journal of Innovation Research and Knowledge (JIRK)*, vol. 2, no. 12, pp. 4561–4568, 2023.
- [14] A. C. Rencher, "Methods of Multivariate Analysis Second Edition."
- [15] M. Sriningsih, D. Hatidja, and J. D. Prang, "Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras di Provinsi Sulawesi Utara," *Jurnal Ilmiah Sains*, vol. 18, no. 1, 2018.
- [16] N. Sari, H. Yasin, and A. Prahutama, "GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT ANALYSIS (GWRPCA) PADA PEMODELAN PENDAPATAN ASLI DAERAH DI JAWA TENGAH," *JURNAL GAUSSIAN*, vol. 5, no. 4, pp. 717–726, 2016, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [17] S. Yulianto and U. Putriana, "ANALISIS KOMPONEN UTAMA (AKU) UNTUK PENGELOMPOKAN AREA PELAYANAN DAN JARINGAN (APJ) DAERAHJAWA TENGAHDAN D.I. YOGYAKARTA," *JASDM*, vol. 1, no. 1, pp. 7–11, 2019.
- [18] A. Ilmaniati and B. E. Putro, "ANALISIS KOMPONEN UTAMA FAKTOR-FAKTOR PENDAHULUAN (ANTECEDENTS) BERBAGI PENGETAHUAN PADA USAHA MIKRO, KECIL, DAN MENENGAH (UMKM) DI INDONESIA," *Januari*, vol. 11, no. 1, 2019, doi: 10.24853/jurtek.11.1.67-78.
- [19] Z. W. Baskara, L. Harsyiah, D. N. A. Paramartha, and Q. D. Utama, "Analisis Faktor Untuk Pemetaan Karakteristik pada Percobaan Dekafeinasi Kopi Robusta," *EIGEN MATHEMATICS JOURNAL*, pp. 15–20, Jun. 2022, doi: 10.29303/emj.v5i1.139.
- [20] A. N. Nitasari *et al.*, "Reduksi Faktor-Faktor yang Mempengaruhi Kualitas Air Hujan dengan Metode Analisis Komponen Utama," *Zeta - Math Journal*, vol. 8, no. 1, pp. 7–15, Feb. 2023, doi: 10.31102/zeta.2023.8.1.7-15.

- [21] H. T. Antono and M. Lutfi, “Variabel-Variabel yang Berpengaruh Terhadap Swabakar Batubara Menggunakan Regresi Komponen Utama,” *Statistika*, vol. 14, no. 1, pp. 25–30, 2014.