Arus Jurnal Sains dan Teknologi (AJST)
Website: http://jurnal.ardenjaya.com/index.php/ajst
Email: jurnal.ajst@ardenjaya.com

Perbandingan Metode *Principal Component Analysis* (PCA) dan *Partial Least Square* (PLS) dalam Penanganan Multikolinearitas pada Kasus Kemiskinan di Provinsi Sulawesi Tenggara Tahun 2023

INFO PENULIS

Chafifah Apriliani Asrat
FMIPA Universitas Halu Oleo
chafifahaprilianiasrat@gmail.com

Makkulau
FMIPA Universitas Halu Oleo

Irma Yahya
FMIPA Universitas Halu Oleo

INFO ARTIKEL

ISSN: 3026-3603
Vol.3, No. 1 April 2025
http://jurnal.ardenjaya.com/index.php/ajst

Saran Penulisan Referensi:

Asrat, C. A., Makkulau, & Yahya, I. (2025). Perbandingan Metode *Principal Component Analysis* (PCA) dan *Partial Least Square* (PLS) dalam Penanganan Multikolinearitas pada Kasus Kemiskinan di Provinsi Sulawesi Tenggara Tahun 2023. *Arus Jurnal Sains dan Teknologi, 3*(1), 68-82

Abstrak

Multikolinearitas merupakan salah satu permasalahan dalam analisis regresi linear berganda yang dapat mempengaruhi kestabilan dan keakuratan estimasi parameter. Oleh karena itu, diperlukan metode yang dapat mengatasinya agar hasil analisis menjadi lebih valid. Penelitian ini bertujuan untuk membandingkan penanganan multikolinearitas pada kasus kemiskinan di Provinsi Sulawesi Tenggara tahun 2023 dengan menggunakan metode *Principal Component Analysis* (PCA) dan *Partial Least Square* (PLS). Analisis dilakukan dengan membentuk komponen baru dari variabel independen menggunakan kedua metode tersebut. Model terbaik ditentukan berdasarkan nilai *Adjusted* R^2, *Root Mean Square Error* (RMSE), dan *Akaike Information Criterion* (AIC). Hasil penelitian menunjukkan bahwa metode PLS menghasilkan nilai *Adjusted* R^2 sebesar 93,26%, RMSE sebesar 0,2596, dan AIC sebesar 7,0880, sedangkan PCA menghasilkan nilai *Adjusted* R^2 sebesar 92,92%, RMSE sebesar 2,5524, dan AIC sebesar 86,1822. Dengan demikian, metode PLS lebih direkomendasikan dalam menangani multikolinearitas karena mampu menghasilkan model regresi yang lebih akurat, efisien, dan informatif.

Kata kunci: PCA, PLS, Multikolinearitas, dan Kemiskinan.

Abstract

Multicollinearity is one of the problems in multiple linear regression analysis that can affect the stability and accuracy of parameter estimation. Therefore, a method is needed that can overcome it so that the analysis results are more valid. This study aims to compare the handling of multicollinearity in poverty cases in Southeast Sulawesi Province in 2023 using the Principal Component Analysis (PCA) and Partial Least Square (PLS) methods. The analysis was carried out by forming new components from independent variables using both methods. The best model was determined based on the Adjusted R^2, Root Mean Square Error (RMSE), and Akaike Information Criterion (AIC) values. The results showed that the PLS method produced an Adjusted R^2 value of 93.26%, RMSE of 0.2596, and AIC of 7.0880, while PCA produced an Adjusted R^2 value of 92.92%, RMSE of 2.5524, and AIC of 86.1822. Thus, the PLS method is more recommended in handling multicollinearity because it is able to produce a more accurate, efficient, and informative regression model.

Keywords: PCA, PLS, Multicollinearity, and Poverty.

A. Pendahuluan

Analisis regresi linear berganda merupakan metode statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen dengan beberapa variabel independen. Salah satu asumsi penting dalam regresi adalah tidak adanya multikolinearitas antar variabel independen (Candra, 2024).

Multikolinearitas terjadi ketika dua atau lebih variabel independen memiliki korelasi tinggi satu sama lain. Efek multikolinearitas dapat menjadikan nilai model tidak dapat menjelaskan hubungan antara variabel dependen dan variabel independen secara baik. Keberadaan multikolinearitas akan menyebabkan varians parameter yang diestimasikan akan menjadi lebih besar dari yang seharusnya dengan demikian tingkat akurasi dari estimasi akan menurun (Supriyadi, dkk, 2017). Oleh karena itu, diperlukan pendekatan alternatif untuk mengatasi permasalahan tersebut tanpa harus menghapus variabel penting dari model (Candra, dkk, 2024).

Beberapa metode telah dikembangkan untuk menangani multikolinearitas, di antaranya *Principal Component Analysis* (PCA) dan *Partial Least Square* (PLS). PCA merupakan metode transformasi yang membentuk komponen baru yang tidak saling berkorelasi (Silitonga, dkk, 2021), sedangkan PLS mengurangi ukuran atau volume dari variabel independen dengan membuat variabel baru yang terbentuk dari gabungan linear antara variabel yang menjelaskan dengan dimensi yang lebih kecil (Gunawan, dkk, 2024). Kedua metode ini mampu mereduksi dimensi data dan mempertahankan informasi penting dalam model regresi.

Penelitian ini menggunakan studi kasus tentang kemiskinan di Provinsi Sulawesi Tenggara Tahun 2023. Kemiskinan merupakan salah satu isu strategis di Indonesia, termasuk di Provinsi Sulawesi Tenggara (Huda, dkk, 2023). Data BPS menunjukkan adanya peningkatan jumlah penduduk miskin di wilayah tersebut pada tahun 2023 (Badan Pusat Statistik Provinsi Sulawesi Tenggara, 2023). Tinggi rendahnya tingkat kemiskinan dipengaruhi oleh berbagai faktor, salah satunya adalah pertumbuhan jumlah penduduk di suatu daerah. Peningkatan populasi yang tidak seimbang dengan ketersediaan lapangan kerja dapat memicu berbagai masalah sosial dan ekonomi, seperti pengangguran, yang pada akhirnya berdampak pada bertambahnya jumlah penduduk miskin.

A1. Tinjauan Pustaka

A.1.1 Regresi Linear Berganda

Regresi berganda adalah tentang memprediksi variabel dependen berdasarkan dua atau lebih variabel independen. Tujuan menggunakan beberapa variabel independen adalah untuk mendapatkan prediksi yang lebih akurat atau untuk menjelaskan lebih banyak variasi dalam variabel dependen (Sudrajat, 2020). Adapun persamaan regresi linear berganda untuk p variabel independen adalah sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{kn} + \varepsilon_i,\ i = 1,2,\ldots,n,\ p = 1,2,\ldots,k, \quad (1)$$

dimana X adalah variabel independen, Y adalah variabel dependen, β_0 adalah intersep/titik potong antara sumbu tegak y dan garis fungsi linear, $\beta_1, \beta_2, \ldots, \beta_k$ adalah koefisien-koefisien regresi atau koefisien kemiringan, ε_i adalah faktor sisaan, dan i adalah pengamatan ke-i.

A.1.2 Uji Multikolinearitas

Uji Multikolinearitas bertujuan untuk mendeteksi apakah variabel independen pada model regresi saling berkorelasi. Untuk memenuhi kriteria *Best Linear Unbiased Estimator* (BLUE), tidak boleh mendapat korelasi antara setiap variabel independen pada model regresi (Muthahharah dan Fatwa, 2022).

A.1.2.1 Pengertian Multikolinearitas

Kata multikolinearitas sendiri pada awalnya diperoleh oleh seseorang yaitu bernama Ragnar Frisch, multikolinearitas memiliki arti bahwa adanya hubungan linier ataupun hubungan yang saling memiliki keterkaitan atau keterkaitan satu sama lain yang dimana hubungan ini bersifat sempurna atau pasti di antara variabel-variabel yang independen atau variabel dependen di dalam suatu model regresi (Gunawan, dkk, 2024).

A.1.2.2 Penyebab Terjadinya Multikolinearitas

Gunawan, dkk (2024) menyatakan bahwa multikolinearitas tidaklah datang begitu saja, tetapi disebabkan oleh beberapa faktor-faktor yang ada, yakni:

a. Pemilihan Sistem penghimpunan data atau informasi yang dipakai,
b. Batasan yang terdapat dalam model atau populasi yang dipilih dari sampel,
c. Spesifikasi model, dan
d. Model yang "*overdetermined*" artinya adanya persamaan yang lebih banyak daripada yang tidak diketahui.

A.1.2.3 Akibat Apabila Terjadi Multikolinearitas

Terjadinya multikolinearitas dapat menyebabkan pemakaian metode regresi menjadi kurang tepat karena taksiran regresinya tidak stabil dan variabel koefisien regresinya sangat besar (Azizah, dkk, 2021). Multikolinearitas menyebabkan estimator mempunyai varian yang besar akibatnya interval estimasi cenderung lebih besar, sehingga membuat variabel independen secara statistik tidak signifikan padahal nilai koefisien determinasi (R^2) tinggi, sehingga sulit mendapatkan estimasi yang tepat (Supriyadi, dkk, 2017).

A.1.2.4 Cara Mendeteksi

Supriyadi, dkk (2017) mengatakan, ada beberapa cara untuk mengetahui keberadaan multikolinearitas dalam suatu model regresi, yaitu:

a. Menganalisis matriks korelasi Jika antara dua atau lebih variabel independen memiliki korelasi yang cukup tinggi, biasanya di atas 0,9, maka hal tersebut mengindikasikan terjadinya multikolinearitas.
b. *Variation Inflation Factor* (VIF) adalah salah satu cara dalam mendeteksi adanya multikolinearitas, dalam penulisan ini menggunakan nilai VIF. Multikolinearitas dalam sebuah regresi dapat diketahui apabila nilai VIF ≥ 10. Gunawan, dkk (2024) mengatakan nilai VIF ini dapat diketahui dengan cara:

$$VIF_p = \frac{1}{1 - R^2{}_p}, \tag{2}$$

dengan $R^2{}_p$ adalah koefisien determinasi antara X_p dengan variabel independen lainnya pada suatu persamaan.
c. TOL (*Tolerance*) Jika nilai *Tolerance* kurang dari 0,1 atau nilai VIF melebihi 10, maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar variabel independen.

A.1.2.5 Penanganan Multikolinearitas

Beberapa cara yang dapat digunakan dalam menyelesaikan masalah multikolinearitas yaitu dengan menghilangkan variabel yang berkorelasi tinggi, menambah jumlah amatan, melakukan transformasikan data ke dalam bentuk lain, atau menggunakan metode regresi lain yang lebih advance. Jika terdapat pelanggaran asumsi multikolinearitas, terdapat beberapa prosedur yang dapat digunakan untuk mengatasinya, seperti menambahkan data, menghilangkan satu atau beberapa variabel independen yang memiliki korelasi tinggi dari model regresi dan menggunakan metode analisis yang lain seperti regresi *ridge* (Candra, dkk, 2024).

A.1.3 Metode *Principal Component Analysis* (PCA)

Nasution, dkk (2019) mengatakan bahwa PCA merupakan kombinasi linear dari variabel awal yang secara geometris kombinasi linear ini merupakan sistem koordinat baru yang diperoleh dari rotasi sistem semula. Metode PCA sangat berguna digunakan jika data yang ada memiliki jumlah variabel yang besar dan memiliki korelasi antar variabelnya. Metode PCA mempunyai langkah penting yang perlu dilakukan sebelum melakukan analisis data menggunakan metode PCA ini, yaitu standarisasi data. Prosedur pengerjaan PCA bertujuan untuk menyederhanakan dan menghilangkan faktor atau indikator yang kurang dominan dan kurang relevan tanpa mengurangi informasi utama dari data asli.

A.1.4 Metode *Partial Least Square* (PLS)

Menurut Bastien, dkk (2005) untuk meregresikan variabel dependen Y dengan variabel independen $X_1, X_2, \ldots, X_k$, metode PLS akan mencari komponen-komponen baru yang berperan sebagai variabel independen untuk mengestimasi parameter regresi. Model regresi PLS dapat diperoleh dengan memanfaatkan uji statistik dari regresi linier sederhana. Hasil dari uji statistik ini memungkinkan untuk memilih variabel independen yang signifikan untuk membangun komponen PLS. Tujuan PLS adalah membentuk komponen yang dapat menangkap informasi dari variabel independen untuk memprediksi variabel dependen. Dalam pembentukan komponen PLS, digunakan variabel dependen dan variabel-variabel independen yang distandarisasi.

A.1.5 Pengujian Kelayakan Model Regresi Terbaik

Untuk melakukan deteksi kelayakan model regresi terbaik, secara umum digunakan uji serentak/simultan (Uji F) dan uji parsial (Uji t).

A.1.5.1 Uji Simultan (Uji F)

Uji F digunakan untuk menguji apakah ada hubungan regresi antara variabel dependen (Y) dengan variabel-variabel independen (X_1, X_2, ..., X_k) secara bersama-sama (simultan).

Hipotesis:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ (variabel yang masuk dalam model tidak signifikan).

H_a: minimal terdapat satu $\beta_p \neq 0$ (paling tidak terdapat satu variabel yang masuk dalam model yang signifikan).

Statistik uji:

$$F_{hit} = \frac{JKR/(k)}{JKG/(n-k-1)} = \frac{KTR}{KTG}, \tag{3}$$

dengan k adalah jumlah variabel independen dalam model, n adalah jumlah total observasi (sampel) yang digunakan dalam analisis, $KTR = \frac{JKR}{k}$, dan $KTG = \frac{JKG}{n-k-1}$.

Daerah Penolakan:

Tolak H_0 jika $F_{hit} \geq F_{(1-\alpha;k,n-k-1)}$ dan

Terima H_0 jika $F_{hit} < F_{(1-\alpha;k,n-k-1)}$ (Wohon, dkk, 2017).

A.1.5.2 Uji Parsial (Uji t)

Tujuan dari uji parsial (uji t) adalah untuk mengetahui seberapa jauh pengaruh dari variabel independen terhadap variabel dependen secara parsial. Pengujian hipotesis akan dilakukan dengan menggunakan tingkat signifikan (α) sebesar 0,05.

Hipotesis:

$H_0 : \beta_p = 0$ (tidak ada pengaruh signifikan)

Untuk hipotesis dua arah:

$H_a : \beta_p \neq 0$ (ada pengaruh signifikan, bisa positif atau negatif).

Untuk hipotesis satu arah:

$H_1 : \beta_p > 0$ (pengaruhnya signifikan positif). Tolak H_0, jika $t_{hit} > t_{(1-\alpha;n-k-1)}$.

$H_1 : \beta_p < 0$ (pengaruhnya signifikan negatif). Tolak H_0, jika $t_{hit} < t_{(\alpha;n-k-1)}$.

Statistik uji:

$$t_{hit} = \frac{b_k}{s\{b_k\}}. \tag{4}$$

Daerah Penolakan:

Untuk uji dua arah:

Tolak H_0 jika $|t_{hit}| > t_{(1-\frac{\alpha}{2};n-k-1)}$ dan

Terima H_0 jika $|t_{hit}| < t_{(1-\frac{\alpha}{2};n-k-1)}$ (Wohon, dkk, 2017)

A.1.6 Ukuran Kesesuaian Model

Model yang baik tentunya memiliki tingkat akurasi yang lebih baik dan dapat mengurangi sisaan dalam memprediksi. Untuk mendapatkan model yang baik diperlukan sebuah metode dalam mengukur kebaikan model. Penelitian ini akan menggunakan *Adjusted* R^2, RMSE, dan AIC untuk membandingkan dan memilih model terbaik.

A.1.6.1 Koefisien Determinasi Ganda Terkoreksi (*Adjusted R²*)

Koefisien determinasi ganda terkoreksi (*adjusted coefficient multiple determination*), dilambangkan oleh R^2_{adj} mengoreksi R^2 dengan cara membagi setiap jumlah kuadrat dalam rumus bagi R^2 dengan derajat independennya masing-masing (Wohon, dkk, 2017). Nilai ini dapat dihitung dengan rumus berikut:

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-k-1}, \tag{5}$$

dengan n adalah jumlah amatan dan $R^2 = 1 - \frac{JKG/(n-k-1)}{JKT/(n-1)}$. Jika konstan tidak termasuk model bagian, maka (n - 1) diganti dengan n.

A.1.6.2 *Root Mean Squared Error* (RMSE)

Milniadi dan Adiwijaya (2023) mengatakan, RMSE adalah aturan penilaian kuadrat yang juga mengukur ukuran rata-rata sisaan. RMSE adalah akar kuadrat dari perbedaan kuadrat rata-rata antara prediksi data dan pengamatan aktual. Rumus dari RMSE adalah:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \tag{6}$$

dimana y_i adalah nilai hasil observasi dan $\hat{y}_i$ adalah nilai hasil prediksi.

A.1.6.3 *Akaike Information Criterion* (AIC)

AIC merupakan kriteria untuk memilih model dalam ekonometrika. Menurut Azizah, dkk (2021), AIC menjadi sarana perbandingan antara model statistik, yang mana dianggap sebagai dasar kriteria untuk mengevaluasi kebaikan model. Nilai yang lebih rendah dari indeks menunjukkan model yang disukai atau dapat dikatakan bahwa model terbaik adalah model yang memberikan nilai AIC terkecil. AIC didefinisikan sebagai berikut:

$$AIC = -2\log(L(\hat{\theta}\,|\,y)) + 2k, \tag{7}$$

dimana:

$L(\hat{\theta}\,|\,y)$: fungsi *likelihood* parameter yang diestimasi

k : jumlah parameter yang diestimasi.

A.1.7 Kemiskinan

Kemiskinan didefinisikan sebagai kondisi dimana seseorang atau sekelompok orang tidak mampu memenuhi kebutuhan dasar untuk mencapai standar hidup yang layak. Kebutuhan dasar ini mencakup pangan, pakaian, tempat tinggal, pendidikan, dan layanan kesehatan. Bank Dunia mendefinisikan kemiskinan ekstrim sebagai hidup dengan kurang dari $1,90 per hari, tetapi standar ini bisa bervariasi tergantung pada konteks lokal. Kemiskinan merupakan masalah global yang kompleks dan berkelanjutan, yang mempengaruhi jutaan orang di seluruh dunia. (Nurrahmah, dkk, 2024). Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan Rp 582.932 per kapita per bulan (Badan Pusat Statistik Indonesia, 2024). Terdapat berbagai faktor yang menyebabkan dan memperburuk jumlah penduduk kemiskinan, diantaranya:

A.1.7.1 Jumlah Penduduk

Jumlah penduduk adalah jumlah masyarakat yang tinggal pada suatu wilayah dan menetap untuk mendiami suatu daerah pada suatu waktu tertentu (Yenny dan Anwar, 2020).

A.1.7.2 Jumlah Pengangguran

Pengangguran terjadi karena ketidakseimbangan antara jumlah angkatan kerja dengan kesempatan kerja yang tersedia. Pengangguran sangat erat kaitannya terhadap masalah kemiskinan, hal ini ditimbulkan karena banyaknya orang yang kehilangan pekerjaan, sehingga mengalami penurunan standar kehidupan (Irawati dan Pakereng, 2023).

A.1.7.3 Kepadatan Penduduk

Kepadatan penduduk adalah perbandingan antara jumlah penduduk dengan luas wilayah yang dihuni. Kepadatan penduduk diakibatkan oleh tiga komponen yaitu: kelahiran, kematian, dan migrasi (Irham dan Putri, 2023).

A.1.7.4 Laju Pertumbuhan Penduduk

Menurut Badan Pusat Statistik Indonesia (2010), laju pertumbuhan penduduk adalah rata-rata pertumbuhan penduduk tahunan antar dua sensus.

B. Metodologi

Data yang digunakan adalah data sekunder yaitu data jumlah penduduk miskin, jumlah penduduk, jumlah pengangguran, kepadatan penduduk, dan laju pertumbuhan penduduk di Provinsi Sulawesi Tenggara pada tahun 2023 dan berasal dari situs Badan Pusat Statistik Provinsi Sulawesi Tenggara (BPS Sultra, 2023). Adapun variabel yang digunakan dalam penelitian ini adalah variabel dependen (Y) yaitu jumlah penduduk miskin dan variabel independen (X) yaitu jumlah penduduk (X_1), jumlah pengangguran (X_2), kepadatan penduduk (X_3), dan laju pertumbuhan penduduk (X_4).

B.1 Tahapan Penelitian

Penelitian ini dilakukan dengan prosedur:

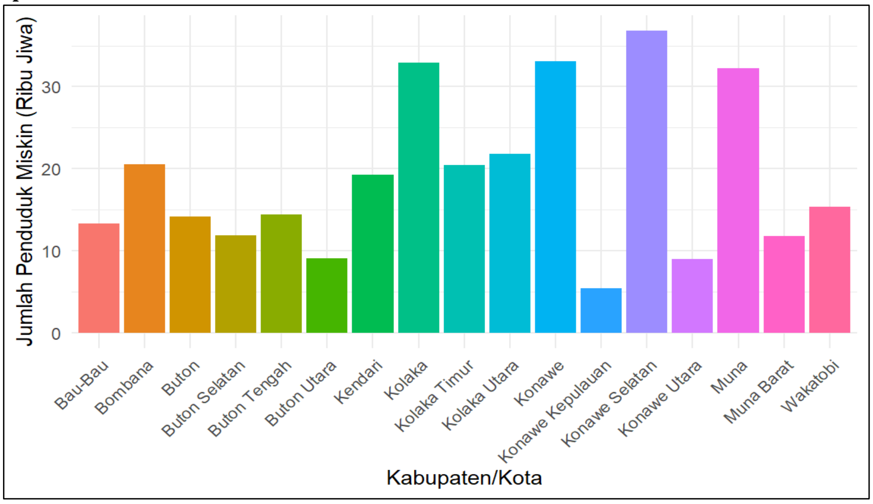
1. Menyiapkan data kemiskinan di Provinsi Sulawesi Tenggara Tahun 2023.
2. Melakukan analisis statistika deskriptif pada data.
3. Melakukan uji multikolinearitas dengan melihat nilai VIF.
4. Mengatasi multikolinearitas dengan melakukan analisis regresi dengan metode PCA sebagai berikut:
   a. Melakukan standarisasi pada variabel independen.
   b. Menghitung matriks varian kovarian, lalu mencari nilai eigen dan vektor eigen dari matriks varian kovarian yang telah diperoleh.
   c. Menentukan nilai proporsi PC (%).
   d. Menghitung matriks korelasi variabel dengan PC berdasarkan vektor eigen.
   e. Menghitung skor PC dari vektor eigen.
   f. Memilih PC berdasarkan proporsi varians dan matriks korelasi variabel.
   g. Meregresikan variabel dependen Y terhadap skor PC SK.
   h. Melakukan analisis hasil model berdasarkan multikolinearitas dan kelayakan model regresi PCA.
   i. Menghitung nilai *Adjusted* R^2, RMSE, dan AIC dari model regresi PCA.
   j. Melakukan transformasi balik variabel standar Z menjadi variabel independen X.
5. Mengatasi multikolinearitas dengan melakukan analisis regresi dengan metode PLS sebagai berikut:
   a. Melakukan standarisasi pada variabel independen dan variabel dependen.
   b. Membentuk komponen PLS pertama ($\mathbf{t}_1$)
      - Melakukan analisis regresi variabel dependen terstandarisasi terhadap masing-masing variabel $\mathbf{X}^*_p$ untuk mengetahui variabel independen yang signifikan untuk membangun komponen PLS pertama ($\mathbf{t}_1$)
      - Melakukan uji signifikansi pada masing-masing variabel $\mathbf{X}^*_p$
      - Menghitung nilai $\text{cov}(\mathbf{Y}, \mathbf{X}_p)$
      - Membentuk komponen PLS pertama
   c. Membentuk komponen PLS kedua ($\mathbf{t}_2$)
      - Melakukan analisis regresi variabel dependen terstandarisasi terhadap $\mathbf{t}_1$ dan masing-masing variabel $\mathbf{X}^*_p$ untuk melihat apakah masih ada variabel independen yang signifikan, jika masih, maka perhitungan komponen PLS kedua akan dilanjutkan
      - Melakukan uji signifikansi pada masing-masing variabel $\mathbf{X}^*_p$
      - Menghitung nilai $\text{cov}(\mathbf{Y}_1, \mathbf{X}_{1p})$
      - Membentuk komponen PLS kedua
   d. Perhitungan komponen PLS berhenti jika semua kovariansi parsial atau variabel independen tidak signifikan.
   e. Meregresikan variabel dependen Y terhadap komponen PLS yang telah dihitung sebelumnya.

   f. Melakukan analisis hasil model berdasarkan multikolinearitas dan kelayakan model regresi PLS.
   g. Menghitung nilai *Adjusted* R^2, RMSE, dan AIC dari model regresi PLS.
   h. Melakukan transformasi balik komponen PLS menjadi variabel independen X.
6. Membandingkan nilai *Adjusted* R^2, RMSE, dan AIC dari perhitungan model regresi PCA dan model regresi PLS.
7. Melakukan penarikan kesimpulan.

C. Hasil dan Pembahasan

C.1 Deskripsi Data

Badan Pusat Statistik Provinsi Sulawesi Tenggara pada 17 Juli 2023 merilis berita tentang persentase penduduk miskin, dikatakan bahwa jumlah penduduk miskin pada Maret 2023 sebesar 321,53 ribu orang, naik 6,79 ribu orang terhadap September 2022, dan naik 11,74 ribu orang terhadap Maret 2022.



Gambar 1. Diagram batang jumlah penduduk miskin kabupaten/kota di Provinsi Sulawesi Tenggara Tahun 2023

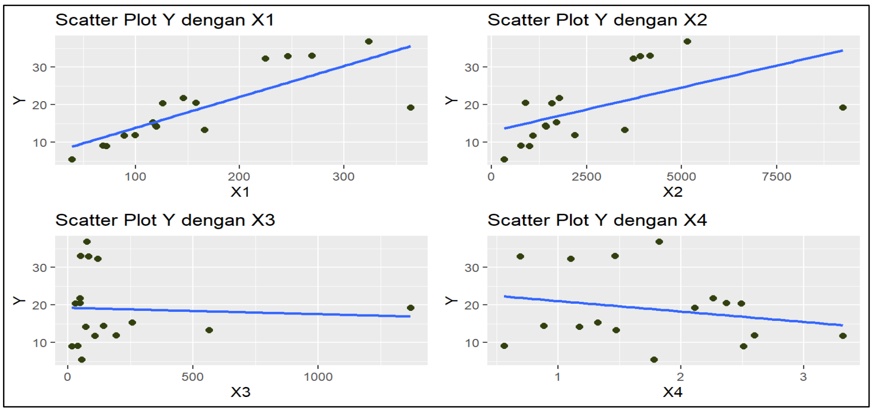
Dari Gambar 1 menunjukan bahwa jumlah penduduk miskin berdasarkan kabupaten/kota di Provinsi Sulawesi Tenggara tahun 2023 memiliki jumlah yang bervariasi, dapat dilihat jumlah penduduk miskin terendah yang berada pada Kabupaten Konawe Kepulauan sebanyak 5 ribu jiwa dan jumlah penduduk miskin tertinggi yang berada pada Kabupaten Konawe Selatan mencapai lebih dari 35 ribu jiwa terlihat memiliki selisih yang besar.

Deskripsi terkait kemiskinan di Provinsi Sulawesi Tenggara Tahun 2023 dapat diketahui dari ukuran pemusatan dan penyebaran data. Ukuran pemusatan menggunakan nilai rata-rata (means), median, minimum, dan maksimum. Sedangkan ukuran penyebaran menggunakan nilai standar deviasi.

Tabel 1. Statistik Deskriptif Data

Variabel	Minimum	Maksimum	Mean	Median	Variansi	Standar Deviasi
Y	5,44	36,84	18,9159	15,35	92,0829	9,596
X_1	39	364,22	161,7065	125,77	8616,9462	92,8275
X_2	334	9248	2582	1701	4920429,875	2218,2042
X_3	17,12	1370,33	193,0301	76,41	109080,9023	330,274
X_4	0,56	3,32	1,7594	1,78	0,5934	0,7703

Tabel 1 menunjukkan mean jumlah penduduk miskin di Provinsi Sulawesi Tenggara tahun 2023 adalah sebesar 18,92, dengan variansi sebesar 92,0829. Variansi tertinggi terletak pada variabel jumlah pengangguran (X2), kepadatan penduduk (X3), dan jumlah penduduk (X1) secara berturut-turut. Variansi yang tinggi pada variabel tersebut menunjukkan besarnya keragaman dari masing-masing wilayah.



Gambar 1. Diagram pencar variabel *Y* terhadap masing-masing variabel *X*

Dari Gambar 2 terlihat bahwa variabel yang memiliki hubungan linear positif terhadap jumlah penduduk miskin (Y) adalah jumlah penduduk (X1) dan jumlah pengangguran (X2). Sedangkan kepadatan penduduk (X3) dan laju pertumbuhan penduduk (X4) memiliki hubungan linear yang negatif terhadap jumlah penduduk miskin yang berarti bahwa peningkatan jumlah penduduk miskin akan menyebabkan nilai kepadatan penduduk dan laju pertumbuhan penduduk menjadi menurun begitupun sebaliknya.

C.2 Pendeteksian Multikolinearitas

Penelitian ini menggunakan metode PCA dan PLS untuk mengatasi multikolinearitas pada model regresi. Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinearitas antara variabel independen adalah dengan melihat nilai VIF. Berikut salah satu hasil pengujian asumsi multikolinearitas dari keempat variabel independen yang digunakan dalam penelitian ini.

Tabel 2. Hasil Uji Multikolinearitas

Variabel	**VIF**
X_1	16,6703
X_2	30,2594
X_3	6,3331
X_4	1,0468

Dari Tabel 2 terbukti bahwa terdapat variabel yang memiliki gejala multikolinearitas. Dapat dilihat terdapat dua variabel independen yang nilai VIF-nya lebih dari 10, sehingga harus dilakukan penanganan untuk mengatasi multikolinearitas pada data dengan menggunakan metode PCA dan PLS.

C.3 Penerapan Metode PCA pada Model Regresi

Metode PCA sangat berguna digunakan jika data yang ada memiliki jumlah variabel yang besar dan memiliki korelasi antar variabelnya. Perhitungan dari PCA didasarkan pada perhitungan nilai eigen dan vektor eigen yang menyatakan penyebaran data dari suatu dataset. Berikut adalah nilai eigen, vektor eigen, dan nilai proporsi PC berdasarkan vektor eigen yang sudah diperoleh.

Tabel 3. Vektor Eigen, Nilai Eigen, dan Proporsi *PC*

Variabel	**Komponen Utama (*PC*)**			
	PC_1	PC_2	PC_3	PC_4
Z_1	-0,5743	-0,1106	0,5893	-0,5574
Z_2	-0,6265	0,0126	0,1200	0,7700
Z_3	-0,5224	0,2336	-0,7590	-0,3106
Z_4	0,0688	0,9659	0,2494	0,0012
Nilai Eigen	2,5008	1,0290	0,4505	0,0197
Proporsi *PC*	0,6252	0,2573	0,1126	0,0049
Kumulatif Proporsi *PC*	0,6252	0,8825	0,9951	1

Setiap *PC* mewakili variabel-variabel yang dianalisis. Kemampuan setiap *PC* mewakili variabel-variabel yang dianalisis ditunjukkan oleh besarnya varians yang dijelaskan, yang disebut dengan nilai eigen. Nilai eigen menunjukkan kepentingan relatif masing-masing komponen dalam menghitung varians keempat variabel yang dianalisis. Dari Tabel 3 dapat dilihat nilai eigen yang lebih besar dari satu (λ_p > 1) dan *PC* yang dapat menyumbangkan data lebih besar dari 75% yaitu PC_1 dan PC_2. Namun, meskipun metode umum seperti *kaiser criterion* merekomendasikan pemilihan *PC* dengan nilai eigen lebih dari 1, dalam penelitian ini seluruh *PC* digunakan tanpa membatasi hanya pada yang memiliki λ_p > 1. Pendekatan ini dilakukan untuk memastikan bahwa tidak ada informasi yang terbuang dan seluruh varians yang tersedia tetap dipertimbangkan dalam pembentukan model regresi.

Langkah selanjutnya adalah dengan mencari nilai skor komponen yang terbentuk untuk membentuk persamaan regresi komponen utama. Skor komponen ini nantinya akan menentukan banyaknya variabel yang akan terbentuk untuk menggantikan variabel independen dengan variabel baru. Berikut adalah skor komponen yang terbentuk.

Tabel 4. Skor Komponen ke-1, ke-2, ke-3, dan ke-4

Observasi ke-	PC_1	PC_2	PC_3	PC_4
1	0,7190	-0,7818	-0,2381	-0,0354
2	-0,6593	-0,9464	0,4143	0,0878
3	-0,9146	-0,5958	1,0009	0,0351
4	-0,8183	-1,5115	0,5129	0,0576
5	-1,5397	-0,1850	1,4558	0,0297
6	0,7838	0,6578	0,4167	-0,428
7	0,3861	-0,4558	-0,6277	-0,0964
8	0,5979	0,5409	0,3461	-0,047
9	1,2232	-1,5126	-0,7222	0,0689
10	1,3469	0,9144	-0,0065	0,1528
11	0,8235	0,8391	0,3259	0,0241
12	1,6139	0,0618	-0,5773	0,086
13	1,1432	1,9753	0,1575	-0,0038
14	0,5921	-1,0946	-0,5005	-0,1055
15	0,5729	1,1254	-0,1422	0,2368
16	-4,9666	1,0689	-0,9457	-0,0085
17	-0,9039	-0,0999	-0,8698	-0,0543

Nilai *PC* yang telah diperoleh ini akan menjadi variabel independen dan kemudian diregresikan terhadap variabel dependen *Y*, sehingga diperoleh model regresi PCA seperti berikut.

$$\hat{Y} = 18,9159 - 2,9957PC_1 - 2,8898PC_2 + 11,1223PC_3 - 4,6908PC_4$$

Persamaan regresi di atas memiliki nilai *adjusted* R^2 0,9292. Nilai tersebut dapat menjelaskan 92,92% variabilitas data dependen berdasarkan $\mathbf{PC}_p$ yang digunakan sebagai variabel independen, sisanya yaitu 7,08% menjelaskan sisaan atau variabel lain. Selanjutnya akan dilakukan pendeteksian multikolinearitas untuk melihat apakah model regresi PCA tersebut masih memiliki gejala multikolinearitas.

Tabel 5. Hasil Uji Multikolinearitas Model Regresi PCA

Variabel	**VIF**
PC_1	1
PC_2	1
PC_3	1
PC_4	1

Dari Tabel 5 terbukti bahwa tidak terdapat gejala multikolinearitas pada model regresi PCA. Dapat dilihat dua komponen PCA tersebut memiliki nilai VIF yang kurang dari 10, sehingga dapat dikatakan metode PCA ini berhasil mengatasi masalah multikolinearitas pada data.

Berdasarkan hasil uji kelayakan model regresi PCA secara simultan, diperoleh nilai statistik uji sebesar 53,54 dengan jumlah komponen PCA (*k*) = 4, derajat kebebasan 1 (db_1) = 4, derajat kebebasan 2 (db_2) = 12, dan $\alpha = 5\% = 0,05$, maka diperoleh nilai F_{tabel} = 3,2592. Dari hasil uji diketahui nilai F_{hit} = 53,54 atau lebih dari F_{tabel}. Berdasarkan kaidah keputusan disimpulkan bahwa H_0 ditolak, artinya paling tidak terdapat satu *PC* yang masuk dalam model yang signifikan. Selanjutnya akan dilakukan uji t yaitu uji kelayakan model secara parsial, untuk melihat *PC* mana saja yang signifikan. Berikut adalah hasil uji kelayakan model regresi PCA dengan menggunakan uji t.

Tabel 6. Hasil Uji Kelayakan Model Regresi PCA Secara Parsial

Parameter	Koefisien	SE Koefisien	Nilai t	*P-value*
(*Intercept*)	18,9159	0,6191	30,5561	$9{,}47\times10^{-13}$
PC_1	-2,9957	0,4035	-7,4240	$8{,}01\times10^{-06}$
PC_2	-2,8898	0,6290	-4,5940	$6{,}17\times10^{-04}$
PC_3	11,1223	0,9507	11,6988	$6{,}42\times10^{-08}$
PC_4	-4,6908	4,5444	-1,0322	0,3223

Diketahui nilai $t_{tabel(0{,}975;12)}$ = 2,179, berdasarkan kaidah keputusan jika |nilai t| > nilai t_{tabel}, maka H_0 ditolak. Dari tabel di atas dapat dilihat variabel yang memenuhi kriteria tersebut adalah variabel PC_1, PC_2, dan PC_3, sehingga didapatkan kesimpulan bahwa variabel PC_1, PC_2, dan PC_3 berpengaruh secara signifikan terhadap variabel Y. Sedangkan variabel PC_4 tidak berpengaruh secara signifikan terhadap variabel Y, ini dikarenakan |nilai t| ≤ nilai t_{tabel}, maka gagal menolak H_0.

Sekarang, akan dilakukan pengukuran model terbaik dengan menggunakan *Adjusted* R^2, RMSE, dan AIC untuk membandingkan dan memilih model terbaik. Berikut adalah hasil perhitungan ukuran kebaikan model regresi PCA.

Tabel 7. Hasil Ukuran Kebaikan Model Regresi PCA

Metode	Nilai
Adjusted R^2	0,9292
RMSE	2,5524
AIC	86,1822

Tabel 7 memperlihatkan hasil dari ukuran kebaikan model regresi PCA, dari tabel tersebut menunjukkan nilai *Adjusted* R^2 sebesar 0,9292 artinya model regresi PCA yang diperoleh dapat menjelaskan 92,92% variabilitas data dependen dengan PC terpilih yang digunakan sebagai variabel independen, sisanya yaitu 7,08% menjelaskan sisaan atau variabel lain. Kemudian nilai RMSE-nya adalah sebesar 2,5524 artinya prediksi model regresi PCA meleset jauh dari nilai asli dan terakhir nilai AIC yaitu sebesar 86,1822 nilai ini sendiri cukup besar, artinya model kurang efisien dan cenderung kurang baik.

C.4 Penerapan Metode PLS pada Model Regresi

Metode PLS digunakan karena mampu membentuk komponen baru yang tidak hanya menangkap informasi dari variabel independen, tetapi juga mempertimbangkan hubungannya dengan variabel dependen. Selanjutnya dilakukan perhitungan komponen PLS seperti berikut.

C.4.1 Perhitungan Komponen PLS Pertama ($\mathbf{t}_1$)

Pertama dilakukan analisis regresi variabel dependen terstandarisasi terhadap masing-masing variabel $\mathbf{X}^*_p$ untuk mengetahui variabel independen yang signifikan untuk membangun komponen PLS pertama ($\mathbf{t}_1$). Berikut adalah hasil analisis regresi variabel dependen terstandarisasi terhadap masing-masing variabel $\mathbf{X}^*_p$.

Tabel 8. Hasil Analisis Regresi terhadap Masing-masing Variabel $\mathbf{X}^*_p$

Variabel	Koefisien	SE Koefisien	Nilai t	*P-value*
X^*_1	0,7957	0,1564	5,088	0,0001
X^*_2	0,5405	0,2172	2,4879	0,0251
X^*_3	-0,0579	0,2578	-0,2245	0,8254
X^*_4	-0,2228	0,2517	-0,8851	0,3901

Uji signifikansi pada taraf nyata 5% menunjukkan bahwa variabel X^*_1 dan X^*_2 signifikan untuk membangun komponen PLS pertama, karena *p-value* dari variabel tersebut memiliki nilai yang kurang dari 0,05. Dengan menggunakan persamaan di bawah, perhitungan komponen PLS pertama adalah sebagai berikut.

$$\mathbf{t}_1 = \frac{1}{\sqrt{\sum_{p=1}^{k} \operatorname{cov}(\mathbf{Y},\mathbf{X}_p)^2}} \sum_{p=1}^{k} \operatorname{cov}(\mathbf{Y},\mathbf{X}_p)\mathbf{X}^*_p,$$

$$\mathbf{t}_1 = \frac{708,7899X^*_1 + 11504,25X^*_2 + 0X^*_3 + 0X^*_4}{\sqrt{708,7899^2 + 11504,25^2 + 0^2 + 0^2}}$$

$$= 0,0615X^*_1 + 0,9981X^{*\,*}_2 + 0X^*_3 + 0X^*_4$$

Dari persamaan di atas dapat dilakukan substitusi variabel independen terstandarisasi (X^*_1 dan X^*_2) yang sebelumnya sudah dihitung ke dalam persamaan untuk mendapatkan nilai komponen PLS pertama. Berikut adalah nilai komponen PLS pertama ($\mathbf{t}_1$)

Tabel 9. Nilai Komponen PLS Pertama ($\mathbf{t}_1$)

Observasi ke-	Nilai t_1
1	-0,5427
2	0,5592
3	0,7828
4	0,6542
5	1,2651
6	-0,7633
7	-0,4264
8	-0,3724
9	-0,879
10	-0,7734
11	-0,4706
12	-1,0928
13	-0,722
14	-0,5531
15	-0,2202
16	3,1336
17	0,421

C.4.2 Perhitungan Komponen PLS Kedua ($\mathbf{t}_2$)

Seperti pada proses sebelumya, langkah pertama dalam pembentukan komponen PLS kedua adalah dengan memeriksa apakah komponen kedua ini masih diperlukan atau tidak. Pemeriksaan ini dilakukan dengan melakukan analisis regresi variabel dependen terstandarisasi terhadap $\mathbf{t}_1$ dan masing-masing variabel $\mathbf{X}^*_p$ untuk melihat apakah masih ada variabel independen yang signifikan dalam membangun komponen PLS kedua ($\mathbf{t}_2$). Berikut adalah hasil analisis regresi variabel dependen terstandarisasi terhadap $\mathbf{t}_1$ dan masing-masing variabel $\mathbf{X}^*_p$.

Tabel 10. Hasil Analisis Regresi terhadap t_1 dan Masing-masing Variabel $\mathbf{X}^*_p$

Variabel	Koefisien	SE Koefisien	Nilai t	*P-value*
X^*_1	2,0588	0,2513	8,1918	$1{,}04\times10^{-6}$
X^*_2	-33,4405	4,0788	-8,1986	$1{,}04\times10^{-6}$
X^*_3	-1,1654	0,1464	-7,959	$1{,}45\times10^{-6}$
X^*_4	-0,1764	0,2176	-0,8104	0,4313

Uji signifikansi pada taraf nyata 5% menunjukkan bahwa masih ada variabel yang signifikan membangun komponen PLS kedua yaitu variabel X^*_1, X^*_2, dan X^*_3. Hal ini dikarenakan *p-value* dari variabel tersebut memiliki nilai yang kurang dari 0,05. Dengan menggunakan persamaan di bawah, perhitungan komponen PLS kedua adalah sebagai berikut.

$$\mathbf{t}_2 = \frac{1}{\sqrt{\sum_{p=1}^{k} \operatorname{cov}(\mathbf{Y}_1, \mathbf{X}_{1p})^2}} \sum_{p=1}^{k} \operatorname{cov}(\mathbf{Y}_1, \mathbf{X}_{1p}) \mathbf{X}^*_{1p},$$

$$\mathbf{t}_2 = \frac{0,2769X^*_{11} - 0,0171X^*_{12} - 0,4842X^*_{13} + 0X^*_{14}}{\sqrt{0,2769^2 + (-0,0171)^2 + (-0,4842)^2 + 0^2}}$$

$$= 0,4962X^*_{11} - 0,0306X^*_{12} - 0,8677X^*_{13} + 0X^*_{14}$$

Dari persamaan $\mathbf{t}_2$ dapat dilakukan substitusi variabel $\mathbf{X}_{1p}$ yang merupakan nilai sisaan dari regresi $\mathbf{X}_p$ terhadap $\mathbf{t}_1$ yang sebelumnya telah distandarisasi ke dalam persamaan untuk mendapatkan nilai komponen PLS kedua. Berikut adalah nilai komponen PLS kedua ($\mathbf{t}_2$).

Tabel 11. Nilai Komponen PLS Kedua ($\mathbf{t}_2$)

Observasi ke-	Nilai t_2
1	0,0091
2	1,1058

3	2,0219
4	1,5608
5	2,6151
6	0,7577
7	-0,8431
8	0,4442
9	-0,5527
10	-0,4422
11	0,2393
12	-1,0191
13	-0,5681
14	-0,2877
15	-0,8945
16	-2,5771
17	-1,5693

C.4.3 Perhitungan Komponen PLS Ketiga ($\mathbf{t}_3$)

Perhitungan komponen PLS ketiga, memiliki langkah yang mirip dengan langkah sebelumnya yaitu pertama dilakukan pemeriksaan apakah komponen ketiga ini masih diperlukan atau tidak. Pemeriksaan ini dilakukan dengan melakukan analisis regresi variabel dependen terstandarisasi terhadap $\mathbf{t}_1$, $\mathbf{t}_2$ dan masing-masing variabel $\mathbf{X}^*_p$ untuk melihat apakah masih ada variabel independen yang signifikan dalam membangun komponen PLS ketiga ($\mathbf{t}_3$). Berikut adalah hasil analisis regresi variabel dependen terstandarisasi terhadap $\mathbf{t}_1$, $\mathbf{t}_2$, dan masing-masing variabel $\mathbf{X}^*_p$.

Tabel 12. Hasil Analisis Regresi terhadap $\mathbf{t}_1$, $\mathbf{t}_2$, dan masing-masing variabel $\mathbf{X}^*_p$

Variabel	**Koefisien**	**SE Koefisien**	**Nilai t**	***P-value***
X^*_1	0,4911	0,431	1,1394	0,2751
X^*_2	-8,0031	6,9979	-1,1436	0,2734
X^*_3	0,4604	0,4039	1,1397	0,275
X^*_4	-0,0229	0,0686	-0,3338	0,7439

Uji signifikansi pada taraf nyata 5% menunjukkan bahwa sudah tidak ada variabel yang signifikan membangun komponen PLS ketiga. Hal ini dikarenakan *p-value* dari variabel memiliki nilai yang lebih besar dari 0,05, sehingga perhitungan berhenti pada komponen PLS kedua.

Komponen PLS baru yang telah diperoleh kemudian diregresikan dengan variabel dependen terstandarisasi. Dari hasil analisis regresi yang dilakukan diperoleh model persamaan regresi linear berganda antara komponen PLS baru dengan variabel dependen seperti berikut.

$$\hat{Y} = -1,025\times 10^{-16} + 0,5286t_1 + 0,6003t_2$$

Persamaan regresi di atas memiliki nilai *adjusted* R^2 sebesar 0, 9326. Nilai tersebut dapat menjelaskan 93,26% variabilitas data dependen dengan komponen PLS baru yang digunakan sebagai variabel independen, sisanya yaitu 6,74% menjelaskan sisaan atau variabel lain. Selanjutnya akan dilakukan pendeteksian multikolinearitas untuk melihat apakah model regresi PLS tersebut masih memiliki gejala multikolinearitas.

Tabel 13. Hasil Uji Multikolinearitas Model Regresi PLS

Variabel	**VIF**
t_1	1
t_2	1

Dari hasil Tabel 13 terbukti bahwa tidak terdapat gejala multikolinearitas pada model regresi PLS. Dapat dilihat dua komponen PLS tersebut memiliki nilai VIF yang kurang dari 10, sehingga dapat dikatakan metode PLS ini berhasil mengatasi masalah multikolinearitas pada data.

Berdasarkan hasil uji kelayakan model regresi PLS secara simultan, diperoleh nilai statistik uji sebesar 111,7 dengan jumlah komponen PLS (k) = 2, derajat kebebasan 1 (db_1) = 2, derajat kebebasan 2 (db_2) = 14, dan $\alpha = 5\% = 0,05$, maka diperoleh nilai F_{tabel} = 3,7389. Dari hasil uji diketahui nilai F_{hit} = 111,7 atau lebih besar dari F_{tabel}. Berdasarkan kaidah keputusan disimpulkan bahwa H_0 ditolak, artinya paling tidak terdapat satu komponen PLS yang masuk dalam model yang signifikan. Selanjutnya akan dilakukan uji t yaitu uji kelayakan model secara parsial, untuk melihat komponen PLS mana saja yang signifikan. Berikut adalah hasil uji kelayakan model regresi PLS dengan menggunakan uji t.

Tabel 14. Hasil Uji Kelayakan Model Regresi PLS Secara Parsial

Parameter	Koefisien	SE Koefisien	Nilai t	*P-value*
(*Intercept*)	$-1{,}025\times10^{-16}$	0,0630	0,000	1,00
t_1	0,5286	0,0616	8,593	$5{,}90\times10^{-7}$
t_2	0,6003	0,0491	12,231	$7{,}32\times10^{-9}$

Diketahui nilai $t_{tabel(0,975;14)}$ = 2,145, berdasarkan kaidah keputusan jika |nilai t| > nilai t_{tabel}, maka H_0 ditolak. Dari Tabel 14 dapat dilihat bahwa semua komponen PLS memenuhi kriteria tersebut, sehingga diperoleh kesimpulan bahwa semua komponen PLS berpengaruh secara signifikan terhadap variabel Y terstandarisasi.

Sekarang, akan dilakukan pengukuran model terbaik dengan menggunakan *Adjusted* R^2, RMSE, dan AIC untuk membandingkan dan memilih model terbaik. Berikut adalah hasil perhitungan ukuran kebaikan model regresi PLS.

Tabel 15. Hasil Ukuran Kebaikan Model Regresi PLS

Metode	Nilai
Adjusted R^2	0,9326
RMSE	0,2596
AIC	7,0880

Tabel 15 memperlihatkan hasil dari ukuran kebaikan model regresi PLS, dari tabel tersebut menunjukkan nilai *Adjusted* R^2 sebesar 0,9326 artinya model regresi PLS yang diperoleh dapat menjelaskan 93,26% variabilitas data dependen dengan komponen PLS baru yang digunakan sebagai variabel independen, sisanya yaitu 6,74% menjelaskan sisaan atau variabel lain. Kemudian nilai RMSE-nya adalah sebesar 0,2596 nilai ini cenderung lebih kecil dari nilai RMSE Model regresi PCA artinya prediksi model regresi PLS sangat dekat dari data aslinya dan terakhir nilai AIC yaitu sebesar 7,0880 nilai ini sendiri lebih kecil dibandingkan nilai AIC Model regresi PCA, artinya model efisien, tidak kompleks berlebihan, dan modelnya bagus.

D. Kesimpulan

Berdasarkan hasil analisis dan pembahasan diperoleh kesimpulan sebagai berikut.

1. Masalah multikolinearitas yang diatasi dengan metode PCA menghasilkan nilai *Adjusted* R^2 sebesar 92,92%%, RMSE sebesar 2,5524, dan AIC sebesar 86,1822 dengan persamaan model regresi sebagai berikut:
$$\hat{Y} = 18{,}9159 - 2{,}9957SK_1 - 2{,}8898SK_2 + 11{,}1223SK_3 - 4{,}6908SK_4,$$
dan setelah ditranformasi ke variabel asalnya, diperoleh persamaan model regresi sebagai berikut:
$$\hat{Y} = 3{,}9836 + 0{,}1208X_1 - 0{,}0002X_2 - 0{,}0185X_3 - 0{,}2974X_4.$$
2. Masalah multikolinearitas yang diatasi dengan metode PLS menghasilkan nilai *Adjusted* R^2 sebesar 93,26%, RMSE sebesar 0,2596, dan AIC sebesar 7,0880 dengan dengan persamaan model regresi sebagai berikut:
$$\hat{Y} = -1{,}025\times10^{-16} + 0{,}5286t_1 + 0{,}6003t_2,$$
dan setelah ditranformasi ke variabel asalnya, diperoleh persamaan model regresi sebagai berikut:
$$\hat{Y} = -34{,}2059 + 0{,}0266X_1 + 0{,}0118X_2 - 0{,}0024X_3.$$
3. Metode PLS menghasilkan model regresi dengan *Adjusted* R^2 lebih tinggi, serta nilai RMSE dan AIC yang lebih kecil. Artinya secara keseluruhan, metode PLS lebih baik dalam menghasilkan model yang akurat, efisien, dan informatif. Dengan demikian, dalam menangani multikolinearitas pada kasus kemiskinan di Provinsi Sulawesi Tenggara tahun 2023, metode PLS lebih direkomendasikan dibandingkan metode PCA.

Saran

Berdasarkan pembahasan, diketahui bahwa metode PLS lebih unggul dibandingkan PCA dalam menangani multikolinearitas pada kasus kemiskinan di provinsi sulawesi tenggara tahun 2023. Namun, jika variabel independen yang digunakan sangat banyak sampai mencapai ratusan, metode PLS ini tidak direkomendasikan untuk dikerjakan secara manual. Atau bisa juga menggunakan metode regresi lain seperti regresi *ridge* dan LASSO.

Untuk penelitian selanjutnya, disarankan dilakukan perbandingan dengan metode lain seperti regresi *ridge* dan LASSO, atau untuk memperoleh hasil yang lebih menyeluruh, disarankan untuk menguji kedua metode ini pada kasus lain atau menggunakan data bangkitan agar dapat mengidentifikasi metode yang paling optimal dalam menangani multikolinearitas pada berbagai jenis data.

E. Referensi

Azizah, I.N., Arum, P.R., dan Wasono, R., (2021). Model Terbaik Uji Multikolinearitas untuk Analisis Faktor-faktor yang Mempengaruhi Produksi Padi di Kabupaten Blora Tahun 2020, *Prosiding Seminar Nasional UNIMUS*., 4(1), 61-69.

Badan Pusat Statistik Indonesia. (2010). Pertumbuhan dan Persebaran Penduduk Indonesia Hasil Sensus Penduduk 2010. *https://www.bps.go.id/id/publication/2011/05/23/35c86fd4076ef 657ef89125d/pertumbuhan-dan-persebaran-penduduk-indonesia-hasil-sensus-penduduk-2010.html*. Diakses pada 18 Februari 2025, pukul 15.30 WITA.

Badan Pusat Statistik Indonesia. (2024). Persentase Penduduk Miskin Maret 2024 turun menjadi 9,03 persen. *https://www.bps.go.id/id/pressrelease/2024/07/01/2370/persentase-penduduk-misk in-maret-2024-turun-menjadi-9-03-persen-.html.* Diakses pada 2 November 2024, pukul 16.54 WITA.

Badan Pusat Statistik Provinsi Sulawesi Tenggara. (2023). Persentase Penduduk Miskin Maret 2023 11,43 Persen*. https://sultra.bps.go.id/en/pressrelease/2023/07/17/1057/persentase-penduduk-miskin-maret-2023-11-43-persen.ht ml*. Diakses pada 16 Oktober 2024, pukul 10.24 WITA.

Badan Pusat Statistik Provinsi Sulawesi Tenggara. (2024). Jumlah Penduduk Umur 15 Tahun Keatas yang Menganggur Menurut Kabupaten/Kota (Jiwa), 2022-2024. *https://sultra.bps.go.id/id/statistics-table/2/Mzk4IzI=/jumlah-penduduk-umur-15-tahun-keatas-yang-menganggur-menurut-kabupaten-kota--jiwa-.html*. Diakses pada 18 Februari 2025, pukul 15.25 WITA.

Bastien, P., Vinzi, V.E., dan Tenenhaus, M., (2005). PLS generalised linear regression, *Computational Statistics & Data Analysis*., 48(1), 17-46.

Candra, W.P.A., Sukarsa, I.K.G., dan Gandhiadi, G.K., (2024). Perbandingan antara Latent Root Regression dan Ridge Regression dalam Mengatasi Multikolinearitas, *INNOVATIVE: Journal Of Social Science Research*., 4(1), 10300-10312.

Gunawan, R.A., Zulkarmain, D.P., dan Arianto, S.T., (2024). Perbandingan Metode Ordinary Least Square (OLS) dan Metode Partial Least Square (PLS) untuk Mengatasi Multikolinearitas, *Socius: Jurnal Penelitian Ilmu-ilmu Sosial*., 1(6), 97-103.

Huda, A.C., Az-Zahra, A., Yasmin, F.P., Ningrum, I.W.K., Putra, W.S., dan Budiasih, (2023). Analisis Regresi Spasial Persentase Kemiskinan di Kawasan Timur Indonesia Tahun 2022, *Seminar Nasional Official Statistics 2023*. 2023(1)., 747-756.

Irawati, M. dan Pakereng, M.A.I., (2023). Analisis Pengaruh Jumlah Pengangguran terhadap Jumlah Kemiskinan menggunakan Metode Regresi Linier (Studi Kasus: Kota Salatiga), *Jurnal Ekonomi dan Manajemen Teknologi*., 7(2), 401-408.

Irham, A.R. dan Putri, R.M., (2023). Kepadatan Penduduk terhadap Indeks Pembangunan Manusia di Provinsi Lampung, *Jurnal Media Komunikasi Geografi*., 24(1), 91-100.

Milniadi, A.D. dan Adiwijaya, N.O., (2023). Analisis Perbandingan Model Arima dan LSTM dalam Peramalan Harga Penutupan Saham (Studi Kasus: 6 Kriteria Kategori Saham Menurut Peter Lynch), *SIBATIK JOURNAL*., 2(6), 1683-1692.

Muthahharah, I. dan Fatwa, I., (2022). Analisis Regresi Linear Berganda untuk Media Pembelajaran Daring terhadap Prestasi Belajar Mahasiswa di STKIP Pembangunan, *Jurnal Matematika dan Statistika serta Aplikasinya*., 10(1), 53-60.

Nasution, M.Z., Nababan, A.A., Syaliman, K.U., Novelan, M.S., dan Jannah, M., (2019). Penerapan Principal Component Analysis (PCA) dalam Penentuan Faktor Dominan yang Mempengaruhi Pengidap Kanker Serviks, *Jurnal Mantik Penusa*., 3(1), 204-120.

Nurrahmah, A., Ningsih, R.S., Dafa A, N., Esabela P, D., dan Puspita, A.M.I., (2024). Menelusuri Akar Kemiskinan di Indonesia: Strategi dan Harapan untuk Masa Depan, *Jurnal Penelitian Pendidikan Indonesia*., 1(4), 31-34.

Silitonga, Y.C., Kamid., dan Multahadah, C., (2021). Perbandingan Metode Stepwise dan Principal Component Analysis (PCA) pada Kasus Faktor-faktor yang Mempengaruhi Pendapatan Asli Daerah (PAD) Di Provinsi Jambi, *Gamma Pi: Jurnal Matematika dan Terapan*., 3(2), 12-20.

Sudrajat, D., (2020). *Pengantar Statistika Pendidikan Disertai Aplikasi Program SPSS*, Surakarta: Pusat Kajian Bahasa dan Budaya.

Supriyadi, E., Mariani, S., dan Sugiman., (2017). Perbandingan Metode Partial Least Square (PLS) dan Principal Component Regression (PCR) untuk Mengatasi Multikolinearitas pada Model Regresi Linear Berganda, *UNNES Journal of Mathematics*., 6(2), 117-128.

Wohon, S.C., Hatidja, D., dan Nainggolan, N., (2017). Penentuan Model Regresi Terbaik dengan menggunakan Metode Stepwise (Studi Kasus: Impor Beras di Sulawesi Utara), *Jurnal Ilmiah Sains*., 17(2), 80-88.

Yenny, N.F. dan Anwar, K., (2020). Pengaruh Jumlah Penduduk terhadap Pertumbuhan Ekonomi di Kota Lhokseumawe, *Jurnal Ekonomika Unimal*., 1(2), 26-31.