

## Optimasi Konstanta Bias Regresi *Ridge* untuk Pemodelan Kemiskinan Jawa Barat

Mela Permata Sari \*, Suliadi Suliadi

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

melapermata236@gmail.com, suliadi@unisba.ac.id

**Abstract.** The issue of multicollinearity in multiple linear regression can disrupt parameter estimation. Ridge regression offers a solution by introducing a bias constant ( $k$ ) to the parameter estimation and typically applied uniformly across all variables. This study implements the Particle Swarm Optimization (PSO) method to determine distinct bias constants for each variable in ridge regression to model poverty levels in 27 regencies/cities in West Java Province using independent variables such as the Human Development Index (HDI), Open Unemployment Rate, Gross Regional Domestic Product (GRDP), and Average Years of Schooling. The results show that only HDI required a penalty ( $k = 0.3833$ ), while the other variables did not. The resulting ridge regression model achieved an  $R^2$  of 65.41% and a Mean Square Error (MSE) of 0.3905. GRDP was identified as the only statistically significant variable.

**Keywords:** : Ridge Regression, Bias Constant, PSO.

**Abstrak.** Masalah multikolinearitas dalam regresi linier berganda dapat mengganggu estimasi parameter. Regresi ridge merupakan solusi dengan menambahkan konstanta bias ( $k$ ), tetapi biasanya konstanta ini bersifat tunggal untuk semua variabel. Studi ini mengimplementasikan metode *Particle Swarm Optimization* (PSO) untuk menentukan konstanta bias yang berbeda pada setiap variabel dalam regresi ridge. Penelitian ini memodelkan tingkat kemiskinan di 27 kabupaten/kota Provinsi Jawa Barat dengan variabel bebas Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka, Produk Domestik Regional Bruto (PDRB), dan Rata-rata Lama Sekolah (RLS). Hasil menunjukkan bahwa hanya IPM yang memerlukan penalti ( $k = 0.3833$ ), sementara variabel lainnya tidak. Model ridge yang dihasilkan memiliki  $R^2$  sebesar 65,41% dan Mean Square Error (MSE) sebesar 0,3905. PDRB ditemukan sebagai satu-satunya variabel yang signifikan secara statistik.

**Kata Kunci:** Regresi Ridge, Konstanta Bias, PSO.

## A. Pendahuluan

Dalam regresi linier berganda dengan pendekatan Metode Kuadrat Terkecil (MKT), salah satu asumsi penting adalah tidak adanya multikolinearitas, yaitu kondisi di mana terdapat hubungan yang sangat kuat antar variabel bebas. Multikolinearitas dapat menyebabkan nilai standard error menjadi tinggi dan menghasilkan estimasi koefisien regresi yang tidak stabil. Untuk mengatasi masalah tersebut, digunakan metode regresi ridge yang diperkenalkan oleh A.E. Hoerl pada tahun 1962. Regresi ridge menambahkan suatu konstanta bias ( $k$ ) ke dalam proses estimasi untuk mengurangi kesalahan kuadrat rata-rata (MSE) dan memperbaiki stabilitas model. Namun, umumnya nilai konstanta bias ini diterapkan secara tunggal untuk seluruh variabel bebas, tanpa memperhatikan tingkat kekolinearitas yang berbeda-beda pada masing-masing variabel. Beberapa studi sebelumnya, seperti yang dilakukan oleh Hoerl dan Kennard (1970), Muniz dan Kibria (2009), serta Wong et al. (2015), telah mengusulkan berbagai metode untuk menentukan nilai  $k$ , namun sebagian besar masih menggunakan pendekatan satu nilai tunggal. Sebagai alternatif, Bas et al. (2017) mengusulkan penggunaan algoritma *Particle Swarm Optimization* (PSO) untuk menentukan nilai  $k$  yang berbeda-beda untuk setiap variabel. Pendekatan ini bertujuan untuk menurunkan nilai *VIF* (*Variance Inflation Factor*) dan sekaligus meminimalkan jumlah kuadrat residu, sehingga menghasilkan model yang lebih akurat.

PSO adalah algoritma optimasi berbasis populasi yang terinspirasi dari perilaku sosial hewan seperti ikan dan burung. Setiap individu atau "partikel" dalam kawanan merepresentasikan kandidat solusi dan akan bergerak berdasarkan pengalaman terbaiknya sendiri serta partikel lain dalam populasi. PSO memiliki keunggulan dalam hal efisiensi komputasi karena tidak memerlukan turunan dari fungsi objektif dan hanya membutuhkan sedikit parameter.

Kemiskinan masih menjadi masalah utama di Indonesia, termasuk di Provinsi Jawa Barat yang merupakan provinsi dengan jumlah penduduk terbanyak. Menurut Badan Pusat Statistik (BPS), persentase penduduk miskin di Indonesia pada September 2022 mencapai 9,57%, sementara di Jawa Barat tercatat sebesar 8,06% atau sekitar 4 juta jiwa. Analisis awal pada data menunjukkan adanya multikolinearitas ( $VIF > 10$ ) antar variabel bebas, sehingga regresi ridge menjadi pendekatan yang tepat untuk digunakan.

Tujuan dari penelitian ini adalah memodelkan faktor-faktor yang memengaruhi tingkat kemiskinan di Provinsi Jawa Barat dengan menggunakan metode ridge regression dengan konstanta bias yang ditentukan secara spesifik untuk setiap variabel menggunakan PSO (Bas et al., 2017). Variabel terikat yang digunakan adalah tingkat kemiskinan, sementara variabel bebasnya meliputi Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka, Produk Domestik Regional Bruto (PDRB), dan Rata-Rata Lama Sekolah (RLS).

## B. Metode

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik Provinsi Jawa Barat. Data yang digunakan terkait dengan data Kemiskinan dari 27 kabupaten/kota di Provinsi Jawa Barat pada Tahun 2022 dengan variabel sebanyak 5 yang dapat dilihat pada Tabel 1.

**Tabel 1.** Variabel Penelitian

| Variabel | Keterangan                                | Satuan        |
|----------|-------------------------------------------|---------------|
| Y        | Kemiskinan                                | Persen        |
| X1       | Indeks pembangunan Manusia (IPM)          | Persen        |
| X2       | Pengangguran                              | Persen        |
| X3       | Pendapatan Domestik Regional Bruto (PDRB) | Miliar Rupiah |
| X4       | Rata-rata Lama Sekolah (RLS)              | Tahun         |

## Regresi Ridge

Regresi ridge merupakan salah satu teknik regularisasi dalam regresi linier yang dirancang untuk mengatasi permasalahan multikolinearitas, yaitu kondisi ketika terdapat hubungan linier yang tinggi antara dua atau lebih variabel independen. Keberadaan multikolinearitas dapat menyebabkan varian koefisien regresi menjadi sangat besar, sehingga menghasilkan estimasi parameter yang tidak stabil dan tidak dapat diandalkan. Regresi ridge memperkenalkan suatu nilai penalti  $k$ , yang juga disebut sebagai parameter bias, ke dalam proses estimasi. Teknik ini memiliki struktur yang mirip dengan metode Ordinary Least Squares (OLS), dengan perbedaan utama pada adanya penambahan penalti terhadap besar koefisien. Bentuk estimasi OLS dan regresi ridge dirumuskan sebagai berikut (Uslu et al., 2017):

$$\text{Estimasi regresi OLS: } \hat{\beta} = (X'X)^{-1}X'Y.$$

$$\text{Estimasi regresi ridge: } \hat{\beta}_R = (X'X + kI)^{-1}X'Y$$

Koefisien regresi ridge bersifat bias, sebagaimana ditunjukkan melalui transformasi berikut:  $\hat{\beta}_R = (X'X + kI)^{-1}X'Y = Z\hat{\beta}$  dengan  $Z = (X'X + kI)^{-1}(X'X)$ ,  $E(\hat{\beta}_R) = E(Z\hat{\beta}) = Z\beta$

## Penentuan Nilai Parameter Bias K

Pemilihan nilai  $k$  merupakan hal yang sangat penting dalam regresi ridge, karena nilai tersebut mengontrol keseimbangan antara bias dan varian. Beberapa pendekatan yang umum digunakan untuk menentukan nilai  $k$  optimal.

## Standarisasi Data

Menurut Kutner et al. (2005), sebelum menerapkan regresi ridge, variabel-variabel sebaiknya distandarisi, yaitu dipusatkan dan diskalakan agar memiliki nilai tengah nol dan simpangan baku satu. Standarisasi dirumuskan sebagai berikut :  $y_i^* = \frac{Y_i - \bar{Y}}{s_y}$ ,  $i = 1, 2, \dots, n$ ,  $x_{ij}^* = \frac{X_{ij} - \bar{X}_j}{s_{xj}}$ ,  $i = 1, 2, \dots, n, j = 1, 2, \dots, p$  dengan  $s_y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$ ,  $s_{xj} = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1}}$ , dimana:  $s_y$  = Standar deviasi variabel terikat,  $s_{xj}$  = Standar deviasi variabel bebas,  $\bar{Y}$  = Rata-rata dari variabel terikat,  $\bar{X}_j$  = Rata-rata dari variabel bebas ke- $j$ . Model regresi setelah transformasi adalah:  $y_i^* = \beta_1^* x_{1i}^* + \beta_2^* x_{2i}^* + \dots + \beta_p^* x_{pi}^* + e_i$ . Koefisien hasil standarisasi memiliki hubungan dengan koefisien asli:  $\beta_j^* = \left(\frac{s_y}{s_{xj}}\right) \beta_j$ ,  $j = 1, 2, \dots, p$ ,  $\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \dots - \beta_p \bar{X}_p = \bar{Y} - \sum_{j=1}^p \beta_j \bar{X}_j$ .

## Particle Swarm Optimization

PSO pertama kali diperkenalkan oleh Eberhart dan Kennedy (1995) sebagai suatu metode komputasi evolusioner yang terinspirasi dari perilaku sosial kawanan organisme, seperti burung atau ikan. PSO bertujuan untuk mencari solusi optimal (baik maksimum maupun minimum) dari suatu permasalahan melalui interaksi kolektif antar agen pencari yang disebut partikel. Dalam PSO, setiap partikel memiliki tiga komponen utama: posisi, kecepatan (*velocity*), dan nilai kecocokan (*fitness*). Posisi mewakili solusi kandidat, *velocity* menunjukkan arah serta laju perpindahan solusi dalam ruang pencarian, sedangkan nilai *fitness* mencerminkan kualitas solusi berdasarkan fungsi objektif yang didefinisikan. Setiap partikel bergerak dalam ruang pencarian dengan memperbarui posisi dan *velocity*-nya berdasarkan dua informasi penting: pengalaman terbaik individu (*personal best* atau *pbest*) dan pengalaman terbaik kelompok (*global best* atau *gbest*). Proses pembaruan ini mengintegrasikan dua mekanisme eksplorasi (menjelajahi ruang solusi secara luas) dan eksploitasi (memfokuskan pencarian pada area potensial) secara simultan untuk mencapai solusi optimal (Suliadi, 2024).

Misalkan ruang solusi berdimensi  $D$ , maka posisi terbaik individu ke- $i$  dan posisi terbaik seluruh kawanan dinotasikan sebagai:  $Pbest_i = p_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ ,  $i = 1, 2, \dots, N$ ,  $gbest = p_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ . Misalkan ingin dicari vektor  $k = (k_1, k_2, \dots, k_p)$  yang meminimalkan fungsi objektif  $f(k)$ . Misal  $k_i(j)$  menyatakan posisi partikel ke- $i$  pada iterasi ke- $t$ , maka:

$$pbest_i = \{k_i(t') \mid f(k_i(t')) = \min\{f(k_i(j))\}_{j=1}^t\}$$

$$gbest = p_g = \{k_i(t') \mid f(k_i(t')) = \min\{f(k_i(j))\}_{i=1}^N\}_{j=1}^t$$

Untuk masalah maksimisasi, operator *min* digantikan oleh *max* pada kedua persamaan tersebut. Pada iterasi ke-(t+1), kecepatan dan posisi partikel ke-ii diupdate menggunakan rumus berikut:

$$v_{id}(t+1) = wv_{id}(t) + c_1r_1(p_{id} - k_{id}(t)) + c_2r_2(p_{gd} - k_{id}(t)), d = 1, 2, \dots, D$$

$$k_i(t+1) = k_i(t) + v_i(t+1)$$

$$v_i(.) = (v_{i1}(.), v_{i2}(.), \dots, v_{iD}(.))$$

di mana  $r_1, r_2 \sim U(0, 1)$  diambil secara acak dari distribusi uniform, w: faktor inersia,  $c_1, c_2$ : koefisien percepatan individu dan sosial. Setelah pembaruan posisi dan velocity dilakukan, nilai *pbest* dan *gbest* akan diperbarui berdasarkan nilai fitness  $f(.)$  terbaru. Proses iterasi dilanjutkan hingga tercapai kondisi konvergensi.

### Algoritma PSO

Berikut langkah-langkah algoritma PSO:

#### Algoritma 1: Particle Swarm Optimization

1. Input Parameter:
  - a. N: jumlah partikel
  - b. D: dimensi permasalahan
  - c. LB, UB: batas bawah dan atas ruang pencarian
2. Inisialisasi:
  - a. Untuk setiap partikel  $i=1$  hingga  $NN$ :
    - i. Untuk setiap dimensi  $d=1$  hingga  $DD$ :
      1.  $vid(0) \leftarrow \text{random}[LB, UB]$  (inisialisasi *velocity*)
      2.  $xid(0) \leftarrow \text{random}[LB, UB]$  (inisialisasi posisi)
      3.  $pbest_i \leftarrow x(0)$  (set pBest awal)
3. Evaluasi Awal:
  - a. Hitung nilai fitness dari semua partikel
  - b. Tentukan  $gbest$  dari semua  $pbest_i$
4. Iterasi (selama belum konvergen):
  - a. Untuk setiap partikel dan dimensi:
    - i. Update sesuai Persamaan (2.40)
    - ii. Update posisi  $k_i(t+1) = k_i(t) + V_i(t+1)$
    - iii. Evaluasi fitness
    - iv. Update  $pbest$   $f(k_i(t)) < f(pbest_i)$  jika lebih baik
    - v. Update  $gbest$   $f(k_i(t)) < f(gbest_i)$  jika ditemukan solusi lebih baik
  - b. Ulangi hingga kondisi konvergensi terpenuhi
5. Output:
  - a. Solusi terbaik diberikan oleh  $gbest$

### Kriteria Konvergensi

Konvergensi dalam PSO dicapai ketika: Stabilitas posisi: Perubahan posisi partikel sangat kecil selama beberapa iterasi berturut-turut. Stabilitas nilai fitness: Nilai fitness terbaik tidak berubah secara signifikan selama sejumlah iterasi.

## Regresi Ridge menggunakan PSO

Bas et al. (2017) mengusulkan suatu pendekatan untuk menentukan nilai konstanta bias  $k$  pada regresi ridge dengan memanfaatkan algoritma PSO. Pendekatan ini dirancang untuk secara simultan meminimalkan kesalahan prediksi dan mengurangi multikolinearitas antar variabel penjelas. Fungsi objektif yang digunakan mencakup dua komponen utama, yaitu MAPE dan VIF, sehingga optimisasi dilakukan dengan tujuan memperoleh nilai  $k$  yang dapat menghasilkan  $VIF < 10$  sekaligus meminimalkan jumlah kuadrat galat (SSE). Permasalahan optimisasi ini dirumuskan dalam bentuk fungsi tujuan sebagai berikut:  $f(k_1, k_2, \dots, k_p) = MAPE + \phi(k_1, k_2, \dots, k_p)$  dengan kendala:  $k_1, k_2, \dots, k_p \geq 0$ . Komponen MAPE didefinisikan sebagai:  $MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$ ,  $\phi(k_1, k_2, \dots, k_p) = \begin{cases} 0 & , VIF_j < 10, j = 1, 2, \dots, p \\ \sum_{j=1}^p VIF_j & , \text{selainnya} \end{cases}$ , an  $VIF_j$  adalah elemen diagonal ke- $j$  dari matrik  $(X'X + K)^{-1}X'X(X'X + K)^{-1}$  atau  $VIF_j = [((X'X + K)^{-1}X'X(X'X + K)^{-1})_{jj}]$ . Namun, berdasarkan temuan dalam studi ini, penggunaan MAPE sebagai komponen utama dalam fungsi objektif memiliki keterbatasan karena nilainya dapat melebihi 100 dan memiliki rentang yang tidak seragam dengan VIF. Untuk mengatasi hal tersebut, digunakan alternatif fungsi kesalahan berupa kebalikan koefisien determinasi  $Inv. R^2$  yang didefinisikan sebagai berikut:

$$Inv. R^2 = 1 - R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \times 100.$$

Fungsi ini memberikan rentang nilai yang pasti, yaitu antara 0 hingga 100, sehingga lebih sebanding dengan nilai VIF dan mencegah dominasi salah satu komponen dalam proses optimisasi. Hal ini menjadikan fungsi objektif yang digunakan menjadi lebih seimbang dan stabil dalam pencarian solusi optimal menggunakan algoritma PSO.

## Algoritma 2 Penentuan konstanta bias $k$

Input:

N: Ukuran kawanan

D: Dimensi permasalahan =  $p$  (banyaknya variabel bebas)

LB: Batas bawah ruang pencarian = 0 (nilai minimum konstanta bias)

UP: Batas atas ruang pencarian = 1 (nilai maximum konstanta bias)

Fungsi objektif:

$$f(k) = Inv. R^2(k) + VIF(k) = Inv. R^2(k_1, k_2, \dots, k_p) + VIF(k_1, k_2, \dots, k_p)$$

Output:

Gbest: posisi (solusi) =  $k = (k_1, k_2, \dots, k_p)$  [konstanta bias regresi ridge]

1. Start
2. Inisialisasi nilai awal posisi ( $k$ ) dan *velocity* ( $v$ ) untuk seluruh kawanan
3. For  $i=1$  to  $N$  do
4.     For  $d=1$  to  $D$  do
5.          $v_{id}(0) \leftarrow \text{random}[LB, UB]$
6.          $k_{id}(0) \leftarrow \text{random}[LB, UB]$
7.     End for
8.      $v_i(0) = (v_{i1}(0), v_{i2}(0), \dots, v_{id}(0))$  #vektor *velocity*
9.      $k_i(0) = (k_{i1}(0), k_{i2}(0), \dots, k_{id}(0))$  #vektor *posisi*
9. End for
10. Matrik posisi:  $P^T = [k_1(0), k_2(0), \dots, k_N(0)]$
11. Hitung skor fungsi tujuan setiap partikel  $f_i(k), i = 1, 2, \dots, N \rightarrow \text{Algoritma 3}$
12.  $Pbest_i = p_i \leftarrow k_i(0), i = 1, 2, \dots, N$
13. Tentukan  $gbest = p_g$  menggunakan persamaan:
$$gbest = p_g = \{k_i, (t') | f(k_i, (t')) = \min \left\{ \{f(k_i(j))\}_{i=1}^N \right\}_{j=1}^t \}$$
14.  $t \leftarrow 1$  #iterasi

15. While [belum konvergensi]
  16.     For i=1 to N
  17.         For d=1 to D
  18.             Update velocity menggunakan persamaan
$$v_{id}(t) = wv_{id}(t-1) + c_1r_1(p_{id} - k_{id}(t-1)) + c_2r_2(p_{gd} - k_{id}(t-1)), d = 1, 2, \dots, D$$

$$V_i(t) = (v_{1d}(t), v_{2d}, \dots, v_{pd}(t))$$
  19.             End for
  20.             Update posisi menggunakan persamaan
$$k_i(t) = k_i(t-1) + V_i(t)$$
  21. Hitung skor fungsi tujuan partikel ke-I iterasi ke-t,  $f(k_i(t))$  menggunakan Algoritma 3
  22.     Jika  $f(k_i(t)) < f(p_i)$  [update pbest
  23.          $p_i \leftarrow k_i(t)$
  24.     End if
  25.     Jika  $(k_i(t)) < f(p_g)$  [update gbest
  26.          $p_g \leftarrow k_i(t)$
  27.     End if
  28. End for
  29.  $t \leftarrow t + 1$
  30. Cek konvergensi : konvergen/belum
  31. End while
  32.  $solusi = gbest = p_g$
  33. finish
- 

### Algoritma 3 : Perhitungan fungsi tujuan untuk satu partikel

---

Input:

$X^*$ : matrik data variabel bebas yang sudah distandarisasi  
 $Y^*$ : vector data variabel terikat yang sudah distandarisasi  
 $k = (k_1, k_2, \dots, k_p)$  posisi suatu partikel

Objective function:

$$f(k) = Inv.R^2(k) + VIF(k) = Inv.R^2(k_1, k_2, \dots, k_p) + VIF(k_1, k_2, \dots, k_p)$$


---

Output:

$K = Diag(k)$  #Matrik diagonal dengan diagonal utama = k

---

1. start
  2.  $\widehat{\beta}^* = (X^{*'}X^* + K)^{-1}X^{*'}Y^* = (b_1, b_2, \dots, b_p)^t$
  3.  $\widehat{y}_i = b_1x_{1i}^* + b_2x_{2i}^* + \dots + b_px_{pi}^*, i = 1, 2, \dots, n$
  4.  $Inv.R^2 = \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \times 100$
  5.  $VIF_j = [((X'X + K)^{-1}X'X(X'X + K)^{-1})_{jj}]$
  6.  $\phi(k_1, k_2, \dots, k_p) = \begin{cases} 0 & , VIF_j < 10, j = 1, 2, \dots, p \\ \sum_{j=1}^p VIF_j & , \text{selainnya} \end{cases}$
  7.  $f(k) = MAPE(k) + \phi(VIF(k))$
  8. Return  $\leftarrow f(k)$
  9. Finish
-

### Uji Signifikansi dalam Regresi Ridge

Signifikansi koefisien regresi ridge dapat diuji menggunakan uji statistik t, dengan Hoptesis  $H_0: \beta_j^* = 0$  (variabel bebas ke-j tidak berpengaruh terhadap respon).  $H_1: \beta_j^* \neq 0$  (variabel bebas ke-j berpengaruh terhadap respon). Statistik uji:  $t = \frac{\hat{\beta}_j^*}{se(\hat{\beta}_j^*)}$ . Dimana  $\hat{\beta}_j^*$  adalah estimasi koefisien regresi ke-j dibawah model regresi ridge, dan  $se(\hat{\beta}_j^*)$  adalah estimasi dari galat baku. Estimasi *standar error* dari elemen ke-j dari  $\hat{\beta}^*$  diperoleh sebagai akar kuadrat dari elemen ke-j dari diagonal matriks kovarians.  $Var(\hat{\beta}^*) = \hat{\sigma}^2 (X^{*'}X^* + K)^{-1}X^{*'}X^*(X^{*'}X^* + K)^{-1}$ . Dimana  $\hat{\sigma}^2 = \frac{\sum e^2}{db.mse} = \frac{n - tr(2H - HH')}{n}$ .  $H = X^*(X'X + K)^{-1}X^{*'}$ .

### C. Hasil Penelitian dan Pembahasan

#### Regresi Linear Berganda

Model awal dibangun menggunakan metode kuadrat terkecil (OLS). Hasil estimasi koefisien regresi dan signifikansinya ditunjukkan pada Tabel 2.

**Tabel 2.** Hasil Regresi Linear Berganda

| Variabel  | Koefisien | Std. Error | t-hitung | p-value | Signifikan |
|-----------|-----------|------------|----------|---------|------------|
| Intercept | 26.860    | 15.010     | 1.789    | 0.0873  | Tidak      |
| X1        | -0.1304   | 0.2997     | -0.435   | 0.6677  | Tidak      |
| X2        | 0.1865    | 0.1826     | 1.021    | 0.3182  | Tidak      |
| X3        | -0.000016 | 0.0000066  | -2.377   | 0.0266  | Ya         |
| X\$       | -0.5157   | 0.4397     | -1.172   | 0.2539  | Tidak      |

**Tabel 3.** Hasil nilai VIF

| Variabel | VIF       |
|----------|-----------|
| $X_1$    | 13.407736 |
| $X_2$    | 1.448201  |
| $X_3$    | 1.537359  |
| $X_4$    | 13.164621 |

Nilai koefisien determinasi ( $R^2$ ) sebesar 65.41% dan MSE sebesar 3.2352 menunjukkan bahwa model mampu menjelaskan sekitar dua pertiga variasi tingkat kemiskinan. Namun, masalah multikolinearitas terdeteksi dengan nilai VIF  $X_1$  dan  $X_4 > 10$  (masing-masing 13.41 dan 13.16), sehingga model OLS tidak dapat diandalkan sepenuhnya.

#### Regresi Ridge Menggunakan PSO

Untuk mengatasi multikolinearitas, dilakukan regresi ridge dengan konstanta bias yang dioptimasi menggunakan *Particle Swarm Optimization (PSO)*. Hasil dari proses PSO menunjukkan bahwa hanya variabel IPM ( $X_1$ ) yang membutuhkan penalti, yaitu  $k = 0.3833$ , sedangkan variabel lainnya tidak ( $k = 0$ ).

**Tabel 4.** Hasil VIF Regresi Ridge

| Variabel | Nilai VIF |
|----------|-----------|
| X1       | 9.7948    |
| X2       | 1.5001    |
| X3       | 1.5289    |
| X4       | 9.9889    |

Seluruh nilai VIF berada di bawah ambang batas 10, yang menandakan bahwa model ridge berhasil mengatasi masalah multikolinearitas.

### Regresi Ridge

Berikut adalah hasil estimasi koefisien setelah dilakukan penalti menggunakan nilai k dari PSO.

**Tabel 5.** Nilai Koefisien Regresi Ridge

| Variabel | Koefisien | Std. Error | t-hitung | p-value | Signifikan |
|----------|-----------|------------|----------|---------|------------|
| X1       | -0.1528   | 0.3764     | -0.406   | 0.6885  | Tidak      |
| X2       | 0.1552    | 0.1473     | 1.053    | 0.3030  | Tidak      |
| X3       | -0.3750   | 0.1487     | -2.522   | 0.0190  | Ya         |
| X4       | -0.5768   | 0.3801     | -1.517   | 0.1427  | Tidak      |

Nilai  $R^2 = 65.41\%$  dan  $MSE = 0.3905$  menunjukkan peningkatan akurasi model dibandingkan regresi OLS ( $MSE$  awal = 3.2352). Hanya variabel PDRB yang berpengaruh signifikan secara statistik terhadap tingkat kemiskinan.

### Pengembalian Koefisien Regresi

Setelah dilakukan estimasi parameter menggunakan metode Ordinary Least Squares (OLS) pada data asli tanpa proses standarisasi, diperoleh model regresi linier berganda yang merepresentasikan hubungan antara tingkat kemiskinan (variabel dependen) dengan variabel-variabel independen yaitu Indeks Pembangunan Manusia (IPM), tingkat pengangguran, Produk Domestik Regional Bruto (PDRB), dan rata-rata lama sekolah. Model ini memberikan gambaran awal mengenai arah dan besarnya pengaruh masing-masing variabel bebas terhadap tingkat kemiskinan.

Adapun bentuk persamaan regresi linier berganda yang diperoleh adalah sebagai berikut:

$$\hat{Y} = 25.18204 - 0.09833X_1 + 0.1879X_2 - 0.0000159X_3 - 0.55811X_4$$

1. **Intercept ( $b_0 = 25.18204$ ):** Jika semua variabel bebas bernilai nol, maka tingkat kemiskinan diperkirakan sebesar 25.18 persen. Namun, secara praktis, interpretasi intercept biasanya tidak bermakna secara langsung jika nilai nol tidak relevan secara konteks.
2.  **$X_1$  (IPM) =  $-0.09833$ :** Setiap peningkatan 1 unit dalam Indeks Pembangunan Manusia (IPM) diperkirakan akan menurunkan tingkat kemiskinan sebesar 0.098 persen poin, dengan asumsi variabel lain tetap. Hubungan ini bersifat negatif, sesuai dengan teori bahwa pembangunan manusia yang lebih baik dapat mengurangi kemiskinan.
3.  **$X_2$  (Pengangguran) =  $0.18797$ :** Setiap kenaikan 1 persen poin tingkat pengangguran diperkirakan meningkatkan tingkat kemiskinan sebesar 0.188 persen poin. Ini menunjukkan hubungan positif antara pengangguran dan kemiskinan.
4.  **$X_3$  (PDRB) =  $-0.0000159$ :** Setiap peningkatan 1 milyar PDRB diperkirakan menurunkan tingkat kemiskinan sebesar 0.0000159 persen poin. Koefisien ini sangat kecil karena PDRB memiliki skala yang besar, sehingga pengaruh per unitnya tampak kecil..
5.  **$X_4$  (Rata-rata Lama Sekolah) =  $-0.55811$ :** Setiap peningkatan 1 tahun rata-rata lama sekolah diperkirakan menurunkan tingkat kemiskinan sebesar 0.558 persen poin. Ini menunjukkan bahwa pendidikan memiliki pengaruh negatif yang cukup kuat terhadap kemiskinan, sesuai teori bahwa pendidikan membantu meningkatkan kesejahteraan ekonomi.

Untuk mengevaluasi performa model, digunakan dua indikator utama, yaitu Mean Squared Error (MSE) dan koefisien determinasi ( $R^2$ ). Nilai MSE sebesar 3,2352 menunjukkan rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, yang mencerminkan tingkat kesalahan prediksi model. Semakin kecil nilai MSE, semakin baik akurasi model. Sementara itu, nilai  $R^2$  sebesar 65,41% menunjukkan bahwa sekitar 65% variasi dalam tingkat kemiskinan dapat dijelaskan oleh variabel-variabel independen dalam model. Sisanya, sebesar 34,59%, dijelaskan oleh faktor-faktor lain yang tidak dimasukkan ke dalam model ini.

## D. Kesimpulan

### Kesimpulan

Berdasarkan hasil analisis, ditemukan adanya permasalahan multikolinearitas di antara variabel-variabel bebas yang digunakan dalam model. Untuk mengatasi permasalahan tersebut, diterapkan metode *Ridge Regression* dengan pendekatan optimasi menggunakan *Particle Swarm Optimization* (PSO) untuk menentukan nilai konstanta bias yang optimal. Hasil optimasi menunjukkan bahwa nilai konstanta bias ( $k$ ) yang diperoleh untuk masing-masing variabel adalah sebagai berikut: IPM ( $X_1$ ) sebesar 0,3833, sedangkan untuk Pengangguran ( $X_2$ ), PDRB ( $X_3$ ), dan Rata-rata Lama Sekolah ( $X_4$ ) diperoleh nilai bias sebesar 0. Hal ini mengindikasikan bahwa hanya variabel  $X_1$  yang memerlukan regularisasi tambahan untuk menurunkan tingkat multikolinearitas dalam model. Model regresi ridge yang diperoleh berdasarkan data yang telah distandarisasi dirumuskan sebagai berikut:  $Y^* = -0.153X_1^* + 0.155X_2^* - 0.375X_3^* - 0.577X_4^*$ . Setelah model dikembalikan ke bentuk semula berdasarkan data asli (tanpa standarisasi), diperoleh model regresi ridge sebagai berikut:  $\hat{Y} = 25.18204 - 0.09833X_1 + 0.1879X_2 - 0.0000159X_3 - 0.55811X_4$ . Model tersebut menghasilkan nilai koefisien determinasi sebesar  $R^2 = 65,41\%$ , yang menunjukkan bahwa sekitar 65% variasi dalam tingkat kemiskinan dapat dijelaskan oleh keempat variabel bebas. Dari hasil pengujian signifikansi, diketahui bahwa hanya variabel PDRB ( $X_3$ ) yang signifikan secara statistik dalam model ini, mengindikasikan peran penting sektor ekonomi terhadap kondisi kemiskinan di Provinsi Jawa Barat.

### Saran

Berdasarkan hasil penelitian yang telah dilakukan, berikut beberapa saran yang dapat diajukan:

1. Implikasi Kebijakan: Mengingat peran signifikan PDRB dalam menurunkan tingkat kemiskinan, pemerintah Provinsi Jawa Barat disarankan untuk meningkatkan produktivitas ekonomi daerah, khususnya di sektor-sektor yang memiliki kontribusi besar terhadap PDRB, sebagai strategi dalam menekan angka kemiskinan.
2. Implikasi Akademik dan Metodologis: Pendekatan PSO dalam regresi ridge terbukti efektif dalam mengatasi multikolinearitas, terutama ketika nilai bias ditentukan secara optimal melalui proses optimasi. Oleh karena itu, metode ini layak untuk diterapkan dalam berbagai penelitian kuantitatif lainnya, khususnya di bidang ekonomi, kesehatan, dan pendidikan yang seringkali melibatkan variabel-variabel dengan korelasi tinggi.

### Daftar Pustaka

- Bas, E., Egrioglu, E., & Uslu, V. R. (2017). Shrinkage Parameters for Each Explanatory Variable Found via Particle Swarm Optimization in Ridge Regression. *Peertechz*, 12-20.
- BPS. (2021). Produk Domestik Regional Bruto. Surakarta: BPS Kota Surakarta.
- BPS. (2024, Maret 15). Rata-rata Lama Sekolah tahun 2023. Retrieved from Badan Pusat Statistik Kabupaten Karanganyar: <https://karanganyarkab.bps.go.id/id/news/2024/03/15/144/rata-rata-lama-sekolah-tahun-2023.html>
- firinguetti, L. (1999). A generalized ridge regression estimator and its finite. *Communications in Statistics-Theory and Methods*, 28: 1217-1229.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 55-67.
- Hoerl, A. E., Kennard, R. W., & Baldwin, K. F. (1975). ridge regression: some simulation. *open jurnal of statistic*, 105-123.

- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. Newyork: MCGraw-Hill Irwin.
- Muniz, & Kiberia. (2009). On same Ridge Regression Estimators an Empirical Comparisons. *Communication in Statistics-Simulation and Computation*, 38(3):621-630.
- Suliadi. (2024). *Particle Swarm Optimization: An Introduction*. Dept. of Statistics Bandung Islamic University.
- Wong, Yiu, K., & Sung Nok, C. (2015). An iterative approach to minimize the mean squared error in ridge. *Computational Statistics*.
- Dyar Al Falah Hilman, & Aceng Komarudin Mutaqin. (2023). Penerapan Regresi Double Poisson untuk Memprediksi Pertandingan dan Klasemen Liga 1 Indonesia. *Jurnal Riset Statistika*, 97–106. <https://doi.org/10.29313/jrs.v3i2.2784>
- Muhammad Bangkit Riksa Utama, & Hajarisman, N. (2021). Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg. *Jurnal Riset Statistika*, 1(1), 35–42. <https://doi.org/10.29313/jrs.v1i1.24>
- Dewi, E. F., & Hajarisman, N. (n.d.). *Penanganan Data Hilang pada Pemodelan Persamaan Terstruktur melalui Metode Full Information Maximum Likelihood (FIML)*. <https://journals.sangkuriangpublisher.com/datamath>