

PERBANDINGAN *GULUD REGRESSION* DAN *PRINCIPAL COMPONENT REGRESSION* (PCR) TERHADAP PEMODELAN INDEKS PEMBANGUNAN MANUSIA PROVINSI JAWA TIMUR

Raihandika Ari Indhova^{1*}, Suparti², Arief Rachman Hakim³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail: dika.raihan12@gmail.com

DOI: 10.14710/j.gauss.12.2.199-208

Article Info:

Received: 2022-10-26

Accepted: 2022-12-28

Available Online: 2023-07-28

Keywords:

Human Resources; HDI; Multicollinearity; Gulud Regression; MSE; PCR

Abstract: Human resources is valuable asset in a country. Human Development Index (HDI) becomes important indicator of quality of human resources in an area. HDI value is affected by a variety of factors that are strongly related to each other so they cause multicollinearity. This observation aims to deal with multicollinearity optimally by comparing Gulud Regression to Component Regression in modeling factors that affect East Java HDI in 2020. Data that are used in this observation are East Java HDI in 2020 (Y), Life Expectancy (X_1), Infant Mortality Rate (X_2), Mean Years of Schooling (X_3), Expected Years of Schooling (X_4), Open-Unemployment Rate (X_5), Average Household Expenditure per Capita (X_6), and Labor Force Participation Rate (X_7). Based on MSE value, the Gulud Regression method is better than Principal Component Regression (PCR) method in dealing with multicollinearity problem. Based on adjusted R^2 score that is 0,954, feasibility test of the best model of Gulud Regression method is a strong model.

1. PENDAHULUAN

Sumber daya manusia (SDM) merupakan aset yang berharga dalam suatu negara. Indeks Pembangunan Manusia (IPM) menjadi indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (BPS, 2020). Menurut BPS (2014), IPM tersusun atas tiga dimensi dasar yaitu, indeks pendidikan, indeks standar hidup layak, dan indeks harapan hidup. Provinsi Jawa Timur berjumlah 40,6 juta jiwa lebih penduduk tahun 2020 (BPS, 2020). Pada tahun 2020, pencapaian pembangunan manusia di tingkat kabupaten/kota cukup bervariasi yaitu berkisar antara 62,7 (Kabupaten Sampang) hingga 82,23 (Kota Surabaya). Nilai IPM merupakan variabel terikat (Y) tidak lepas dari pengaruh berbagai macam faktor pengaruh atau variabel bebas (X). Suatu metode statistik yang dapat digunakan untuk mengetahui pengaruh variabel bebas terhadap variabel terikat yaitu metode regresi linier. Regresi linier berganda merupakan salah satu metode regresi yang sering digunakan untuk mengetahui ada pengaruh atau tidaknya dengan banyaknya variabel bebas X lebih dari 1 (Gujarati, 2004). Pada penerapannya, model regresi linier akan optimal jika tersebut terpenuhinya asumsi-asumsi klasik yaitu data residual berdistribusi normal, tidak adanya heteroskedastisitas, tidak adanya autokorelasi, dan tidak adanya multikolinieritas (Purnomo, 2017) dalam (Sihombing, 2017). Namun, tidak dipungkiri adanya hubungan yang kuat satu dengan lainnya pada faktor-faktor pengaruh atau variabel-variabel X yang diteliti, sehingga dapat menimbulkan masalah multikolinieritas. Adanya hubungan atau korelasi yang tinggi di antara variabel bebas disebut kasus multikolinieritas (Yan dan Su, 2009) dalam (Putra dan Ratnasari, 2015). Salah satu metode penanganan kasus multikolinieritas adalah *Gulud* atau *Ridge Regression* dan Regresi Komponen Utama atau *Principal Component Regression* (Montgomery dan Peck, 1991). Beberapa penelitian terdahulu yang meneliti tentang *Gulud Regression* dan *Principal Component Regression* (PCR) antara lain penelitian yang dilakukan Guntur (2017) tentang Analisis Kemiskinan dan Kerentanan Kemiskinan dengan Regresi Ridge, LASSO, dan *Elastic-Net* Di Provinsi Jawa Timur Tahun

2017. Dari hasil penelitian tersebut didapatkan hasil bahwa metode Ridge atau *Gulud* dapat berperan dalam kasus penanganan multikolinearitas. Penelitian juga dilaksanakan oleh Astuti (2014), dalam penelitiannya membandingkan metode *Partial Least Square* (PLS) dengan Regresi Komponen Utama atau *Principal Component Regression* untuk data multikolinieritas pada kasus Indeks Pembangunan Manusia di Kabupaten Gunung Kidul. Oleh karena itu peneliti mencoba membandingkan metode mana yang lebih optimal dari kedua metode tersebut dalam mengatasi kasus multikolinearitas pada pemodelan faktor-faktor yang mempengaruhi nilai IPM di Provinsi Jawa Timur tahun 2020.

2. TINJAUAN PUSTAKA

Indeks Pembangunan Manusia menjelaskan hasil pembangunan dalam bentuk pendapatan, kesehatan, pendidikan, dan aspek lain yang terakses penduduk (BPS, 2020). Menurut BPS (2015), dimensi pembentuk IPM tersebut ialah umur panjang dan hidup sehat (*a long and healthy life*), pengetahuan (*knowledge*), dan standar hidup layak (*decent standard of living*).

Regresi linear berganda yaitu metode regresi untuk menganalisis pengaruh lebih dari satu variabel bebas terhadap satu variabel terikat. Adapun model umum regresi linier berganda dengan k variabel bebas dan n pengamatan menurut Montgomery dan Peck (1991) yaitu:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i, \text{ dimana } i=1,2,\dots,n \quad (1)$$

Adapun estimator parameter terhadap β sebagai berikut :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Terdapat sifat-sifat penduga terkecil antara lain :

1. $\hat{\beta}$ Linier jika β merupakan fungsi linier dari β .

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$
2. $\hat{\beta}$ merupakan penduga tak bias dari β jika $E(\hat{\beta}) = \beta$.

$$E(\hat{\beta}) = X'X^{-1}X'X\beta = \beta$$
3. $\hat{\beta}$ mempunyai variansi minimum

$$\text{Var}(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = (X'X)^{-1}\sigma^2.$$

Adapun uji Asumsi Klasik yang harus terpenuhi sebagai berikut :

1. Uji Normalitas memiliki tujuan untuk mengetahui apakah dalam model regresi, residual data berdistribusi normal atau tidak (Siregar, 2015). Hipotesisnya yaitu :

H_0 : Residual data berdistribusi normal

H_1 : Residual data tidak berdistribusi normal

Kriteria Uji : H_0 ditolak jika $D = \sup |F_{(ei)} - F_{H(ei)}| \geq$ tabel *Kolmogorov-Smirnov*

2. Uji Non Autokorelasi menurut Sujarweni (2015) dalam (Wahab, 2020), digunakan untuk mengetahui keadaan ketika model regresi terdapat korelasi antara residual pada periode tertentu t dengan residual pada periode sebelumnya $(t - 1)$. Adapun hipotesis :

H_0 : Tidak terjadi autokorelasi

H_1 : Terjadi autokorelasi

Kriteria Uji : Rumus yang digunakan uji DW sebagai berikut :

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2}$$

Tabel 1. Kriteria Pengujian *Durbin-Watson* (DW)

d	H_0	Keputusan
$0 < d < dl$	Tidak terdapat autokorelasi positif	H_0 ditolak
$dl \leq d \leq du$	Tidak terdapat autokorelasi positif	Tidak ada keputusan
$4 - dl < d < 4$	Tidak terdapat autokorelasi negatif	H_0 ditolak
$4 - du \leq d \leq 4 - dl$	Tidak terdapat autokorelasi negatif	Tidak ada keputusan
$du < d < 4 - du$	Tidak terdapat autokorelasi negatif atau positif	Gagal tolak H_0

Jika keputusan analisis berada pada daerah ragu-ragu, untuk memastikan kembali keputusan uji non autokorelasi maka dilanjutkan dengan uji *run test* (Meiryani, 2021) dengan hipotesis sebagai berikut :

H_0 : Tidak terjadi autokorelasi

H_1 : Terjadi autokorelasi

Statistik Uji: $Z_{hit} = \frac{R - \bar{R}}{S_R}$, dengan $\bar{R} = \frac{2n_1n_2}{n_1+n_2} + 1$ dan $S_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$.

Kriteria Uji : Jika nilai $|Z_{hit}| > Z_{\frac{\alpha}{2}}$ atau nilai *Asymp.sig (2 tailed)* $< 0,05$ (α) maka H_0 ditolak dan sebaliknya.

- Uji Non Heterokedastisitas dilakukan dengan melihat ada tidaknya pola tertentu pada grafik *scatterplot* antara variabel terikat dengan residualnya. Uji *Breusch-Pagan* hipotesisnya yaitu :

H_0 : $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$ (Tidak terjadi heteroskedastisitas)

H_1 : $\sigma_i^2 \neq \sigma^2$, dimana minimal salah satu i (Terjadi heteroskedastisitas)

Kriteria Uji: H_0 di tolak jika $BP = nR^2 \sim \chi_{(n-1)}^2 > \chi_{(\frac{\alpha}{2}, n-1)}^2$.

- Uji Non Multikolinieritas menurut Umar (2011), diterapkan untuk mendeteksi adanya korelasi antar variabel bebas. Menurut Yan dan Su (2009) dalam (Putra dan Ratnasari, 2015), nilai VIF untuk parameter regresi ke- j dirumuskan pada persamaan sebagai berikut :

$$VIF_j = \frac{1}{1-R_j^2}, j = 1, 2, \dots, k.$$

Jika nilai $VIF > 5$ maka variabel bebas dalam model mengalami multikolinearitas (Neter, *et al.*, 1997).

Pengukuran ketepatan model regresi dengan uji signifikansi parameter sebagai berikut:

- Uji Signifikansi Parameter simultan diterapkan untuk menganalisis secara bersama-sama variabel bebas berpengaruh signifikan atau tidak terhadap variabel terikat (Santosa Slamet, 2014) dalam (Lestari *et al.*, 2020). Hipotesisnya sebagai berikut :

H_0 : $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$

H_1 : minimal ada satu $\beta_j \neq 0$, dengan $j = 1, 2, \dots, k$

Kriteria Uji : H_0 di tolak jika nilai $F = \frac{SS_R/(k)}{SS_E/(n-k-1)} \geq$ nilai $F(\alpha, k, n-k-1)$

- Menurut Santosa Slamet (2014) dalam (Lestari *et al.*, 2020), uji signifikansi parameter parsial atau uji statistik t digunakan untuk mengetahui apakah secara parsial atau setiap variabel bebas berpengaruh secara signifikan atau tidaknya terhadap variabel terikat. Adapun hipotesisnya:

H_0 : $\beta_j = 0$

H_1 : $\beta_j \neq 0$, dimana $j = 1, 2, \dots, k$.

Kriteria Uji : H_0 di tolak jika nilai $t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} > t_{(\frac{\alpha}{2}, n-k-1)}$.

Sebelum melakukan penerapan metode penanganan kasus multikolinearitas diperlukan tahapan standarisasi variabel. Standarisasi variabel merupakan suatu teknik yang digunakan untuk menyamakan ukuran atau satuan variabel satu dengan variabel yang lain yang sebelumnya berbeda. Berdasarkan pernyataan tersebut maka rumus standarisasi sebagai berikut:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \text{ dan } X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j}}$$

Pada penerapannya *Gulud regression* dan MKT bertujuan meminimumkan jumlah kuadrat sisaan (JKS) pada pendugaan koefisien regresi. Perbedaan kedua metode tersebut yaitu adanya penambahan penalti penyusutan dalam meminimumkan JKS dalam metode gulud. Bentuk model persamaan *Gulud Regression* yang telah terstandarisasi dapat dituliskan sebagai berikut :

$$Y_i^* = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \dots + \beta_k^* X_{ik}^* + \varepsilon_i^* \quad (2)$$

Estimator parameter *Gulud Regression* terhadap β^* sebagai berikut :

$$\hat{\beta}_R^* = (X^{*'} X^* + KI)^{-1} X^{*'} Y^*$$

Adapun sifat-sifat estimator *gulud regression* sebagai berikut :

1. Menurut Montgomery dan Peck (1991), estimator *gulud regression* $\hat{\beta}_R^*$ merupakan estimator yang bias dari β^* .

$$\hat{\beta}_R^* = (X^{*'} X^* + KI)^{-1} (X^{*'} X^*) \hat{\beta}^* = Z \hat{\beta}^*$$

2. Menurut Montgomery dan Peck (1991), matriks kovarian dari $\hat{\beta}_R^*$ sebagai berikut:

$$\text{Cov}(\hat{\beta}_R^*) = \sigma^2 (X^{*'} X^* + KI)^{-1} X^{*'} X^* (X^{*'} X^* + KI)^{-1}$$

3. Menurut Hoerl dan Kennard (1970), jumlah kuadrat error untuk ridge atau *gulud regression* sebagai berikut:

$$SSE(\hat{\beta}_R^*) = (Y^* - X^* \hat{\beta}_R^*)' (Y^* - X^* \hat{\beta}_R^*)$$

Menurut Hoerl dan Kennard (1970), memilih tetapan bias *gulud regression* (K) dengan menggunakan *ridge trace*, dengan nilai berkisar antara 0 dan 1.

Principal Component Regression (PCR) atau Regresi komponen utama adalah metode analisis regresi berasal dari metode analisis komponen utama atau *Principal Component Analysis* (PCA), dimana PCA sebagai tahap analisis antara. *Principal Component* merupakan kombinasi linear dari variabel awal (X_1, X_2, \dots, X_k) yang menghasilkan variabel baru ($X_1^*, X_2^*, \dots, X_k^*$). Persamaan model komponen utama dari variabel yang distandarkan ($X_1^*, X_2^*, \dots, X_k^*$) yaitu :

$$\begin{aligned} W_1 &= a_{11} X_1^* + a_{12} X_2^* + \dots + a_{1k} X_k^* \\ W_2 &= a_{21} X_1^* + a_{22} X_2^* + \dots + a_{2k} X_k^* \\ &\vdots \\ W_p &= a_{p1} X_1^* + a_{p2} X_2^* + \dots + a_{pk} X_k^* \end{aligned}$$

Dimana $X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j}}$, $j = 1, 2, \dots, k$ dan proporsi dari total varians ke- $k = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_k}$; $j =$

1, 2, ..., k. Menurut Ifadah (2011), PCA memiliki tahapan-tahapan sebagai berikut :

1. Uji *Kaiser-Mayer-Olkin* (KMO) dan *Barlett Test of Sphericity* digunakan untuk mengetahui apakah data layak dan dapat dianalisis dengan analisis komponen utama (Supranto, 2004). Jika nilai KMO bernilai antara 0,5 sampai dengan 1 maka PCA dapat digunakan. Jika nilai KMO kurang dari 0,5 maka PCA tidak dapat digunakan. Sedangkan *Barlett Test of Sphericity* diperlukan untuk melihat kebenaran bahwa ada atau tidaknya korelasi antar variabel bebas (X). Adapun hipotesis :

$H_0: R = I$ (tidak terdapat korelasi antar variabel bebas)

$H_1: R \neq I$ (terdapat korelasi antar variabel bebas)

Kriteria Uji : Jika nilai $\chi^2_{hitung} = -\ln|\mathbf{R}| \left[(n-1) - \frac{(2k+5)}{6} \right] \geq \chi^2_{tabel}$ atau $p\text{-value} <$ taraf signifikansi α , maka H_0 ditolak. χ^2_{tabel} yaitu $\chi^2_{(\alpha, \frac{k}{2}(k-1))}$.

2. *Anti Image Correlation* dilakukan dengan melihat angka MSA (*Measure of Sampling Adequacy*) yaitu angka korelasi yang memiliki tanda a (terletak arah diagonal dari atas kiri ke kanan bawah) yang bernilai dari 0 sampai 1, dengan kriteria sebagai berikut :
 - Jika MSA = 1, variabel bebas dapat diprediksi tanpa kesalahan oleh variabel lain.
 - Jika MSA > 0,5, variabel bebas berkemungkinan dapat diprediksi dan analisis dapat dilanjutkan.
 - Jika MSA < 0,5, variabel bebas tidak dapat diprediksi dan analisis tidak dapat dilanjutkan atau dikeluarkan dari proses analisis terlebih dahulu.
3. *Communalities* diperlukan untuk mengetahui berapa varians variabel bebas yang dapat dijelaskan oleh faktor yang terbentuk.
4. *Total Variance Explained* melibatkan nilai *eigenvalue* untuk mengetahui kemampuan setiap faktor mewakili variabel-variabel bebas yang ditunjukkan dengan besarnya varians kriteria angka *eigenvalue* dibawah 1 tidak dapat digunakan lebih lanjut pada proses PCA.
5. *Component Score Coefficient Matrics* untuk mengetahui berapa faktor yang didapatkan dalam proses PCA.

Model persamaan Principal Component Regression yang telah terstandarisasi dapat dituliskan sebagai berikut :

$$Y_i = b_0 + \sum_{m=1}^p b_m W_{im} + \varepsilon_i \quad (3)$$

Ukuran dalam menentukan model terbaik dapat dilihat dari nilai *Mean Square Error* (MSE). Semakin tinggi nilai MSE maka semakin buruk modelnya. MSE dapat dihitung dengan rumus:

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - k - 1}$$

Untuk mengukur kebaikan model menggunakan nilai R^2 yaitu jika $R^2 < 0,30$ (Performa lemah), $0,30 \leq R^2 < 0,60$ (Kinerja Moderat/sedang), dan $R^2 \geq 0,60$ (Kinerja kuat) (Sanchez, 2013).

3. METODE PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh peneliti dari website resmi Badan Pusat Statistik Provinsi Jawa Timur tahun 2020. Pada penelitian ini data yang digunakan yaitu data Indeks Pembangunan Manusia (IPM) di Jawa Timur tahun 2020 (Y), Angka Harapan Hidup (AHH) (X_1), Angka Kematian Bayi (AKB) (X_2), Rata-rata Lama Sekolah (RLS) (X_3), Harapan Lama Sekolah (HLS) (X_4), Tingkat Pengangguran Terbuka (TPT) (X_5), Rata-rata Pengeluaran Perkapita (RPPK) (X_6), dan Tingkat Partisipasi Angkatan Kerja (TPAK) (X_7). *Software* yang digunakan dalam penelitian ini adalah *software* R dan *SPSS Statistic 23*. Tahapan analisis data yang digunakan sebagai berikut:

1. Menentukan dataset penelitian sesuai faktor-faktor yang sudah ditetapkan. Kemudian melakukan eksplorasi data asli untuk mencari nilai rata-rata, standar deviasi, nilai maksimum, dan nilai minimum dari seluruh variabel penelitian.
2. Melakukan pemodelan regresi berganda beserta pengujian asumsi-asumsinya dengan meregresikan variabel (Y) dengan variabel (X) sejumlah n. Termasuk di dalamnya terdapat proses perhitungan nilai VIF pada setiap variabel bebas (X) untuk mengetahui ada tidaknya multikolinearitas.

3. Melakukan standarisasi variabel dilanjutkan penanganan multikolinearitas dengan *Gulud Regression* dan melakukan uji F dan uji t.
4. Melakukan penanganan multikolinearitas dengan *Principal Component Regression* (PCR) dan melakukan uji F dan uji t.
5. Membandingkan nilai *Mean Square Error* (MSE) untuk memperoleh metode terbaik dengan kriteria metode terbaik memiliki nilai MSE terendah.
6. Mengukur kebaikan model dengan nilai R^2 dengan kriteria semakin tinggi nilai R^2 semakin baik performanya.
7. Menganalisis hasil dan perhitungan.
8. Mengembalikan persamaan regresi ke bentuk variabel asli.
9. Membuat kesimpulan dari analisis yang telah dilakukan.

4. HASIL DAN PEMBAHASAN

Persamaan regresi linier berganda terstandarisasi yaitu:

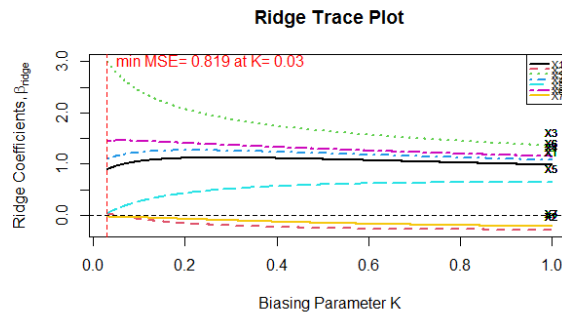
$$\hat{Y} = 19,420 + 0,31X_1 + 0,00156X_2 + 1,907X_3 + 0,877X_4 - 0,053X_5 + 0,000004X_6 - 0,006X_7 \quad (4)$$

1. Uji Normalitas, berdasarkan *output* R nilai D sebesar $0,026316 < d_{(0,05,38)}$ sebesar 0,221 dan nilai $p\text{-value} = 1 > \alpha$ maka gagal tolak H_0 sehingga dapat disimpulkan bahwa residual data berdistribusi normal
2. Uji Non Autokorelasi, berdasarkan *output* R nilai statistik DW terletak diantara dl dan du yaitu $dl = 0,9079 < d = 1,6824 < du = 2,1440$ maka gagal tolak H_0 sehingga tidak ada kesimpulan atau dapat juga dikatakan kesimpulan terdapat pada daerah ragu-ragu. Berdasarkan hasil tersebut untuk memastikan kembali keputusan uji non autokorelasi maka dilanjutkan dengan uji *run test* (Meiryani, 2021). Berdasarkan statistik uji dan *output* SPSS pada *output run test* terlihat nilai terlihat nilai $|Z_{hit}| = 0,96 < Z_{\frac{\alpha}{2}} = 1,96$ dan nilai *Asymp.sig (2 tailed)* sebesar $0,411 > 0,05$ (α), maka gagal tolak H_0 , sehingga dapat disimpulkan tidak terjadi autokorelasi.
3. Uji Non Heteroskedastisitas, berdasarkan *output* R nilai BP sebesar $6,4763 < \chi^2_{(0,05;37)}$ sebesar 52,1923 dan $p\text{-value}$ sebesar $0,4854 > \alpha$, maka gagal tolak H_0 , sehingga dapat disimpulkan bahwa tidak terjadi heteroskedastisitas.
4. Uji Non Multikolinearitas, berdasarkan *output* R, Tabel 2 terlihat bahwa variabel bebas Rata-rata Lama Sekolah (RLS) (X_3) dan Rata-rata Pengeluaran Perkapita (RPPK) (X_6) mempunyai VIF lebih besar dari 5 yaitu sebesar 9,243 dan 9,498 sehingga disimpulkan terdapat multikolinearitas.

Tabel 2. Nilai VIF Variabel bebas

Variabel Bebas	VIF
X_1	2,491
X_2	1,282
X_3	9,243
X_4	3,389
X_5	4,511
X_6	9,498
X_7	1,738

Setelah menstandarisasi variabel, maka dilakukan analisis *Gulud Regression* dan diperoleh hasil *output* R *Ridge Trace Plot* Gambar 1, tetapan bias λ sebesar 0,03 menunjukkan nilai MSE terkecil sebesar 0,819 dan menunjukkan nilai VIF pada setiap variabel bebas < 5 . Tetapan bias λ sebesar 0,03 memenuhi kriteria ($0 < \lambda < 1$), sehingga nilai tetapan bias tersebut dapat digunakan



Gambar 1. Output Ridge Trace Plot

Model *Gulud Regression* menerapkan nilai λ sebesar 0,03 diperoleh sebagai berikut :

$$\hat{Y}_R^* = 0,148X_1^* + 0,006X_2^* + 0,494X_3^* + 0,181X_4^* - 0,007X_5^* + 0,238X_6^* - 0,005X_7^* \quad (5)$$

Uji Signifikansi Parameter Simultan, berdasarkan *output* R nilai F sebesar 224,178 > $F_{(0,05,7,30)}$ sebesar 2,33 dan *p-value* sebesar 0,000 < α , maka H_0 ditolak, sehingga dapat disimpulkan bahwa secara simultan bahwa terdapat variabel bebas X^* berpengaruh signifikan terhadap variabel terikat Y^* .

Uji Signifikansi Parameter Parsial dapat dilihat pada Tabel 3.

Tabel 3. Uji Signifikansi Parameter Parsial *Gulud Regression*

Variabel	$ t_{hitung} $	$t_{(0,025;30)}$	<i>p-value</i>	Keputusan
X_1^*	4,2271	2,0422	0,0002	H_0 ditolak
X_2^*	0,2115	2,0422	0,8339	Gagal tolak H_0
X_3^*	8,7256	2,0422	0,0000	H_0 ditolak
X_4^*	4,5500	2,0422	0,0001	H_0 ditolak
X_5^*	0,1539	2,0422	0,8787	Gagal tolak H_0
X_6^*	4,2244	2,0422	0,0002	H_0 ditolak
X_7^*	-0,1523	2,0422	0,8800	Gagal tolak H_0

Berdasarkan kriteria uji dan Tabel 3 maka variabel X_1^* AHH, X_3^* RLS, X_4^* HLS, dan X_6^* RPPK secara parsial berpengaruh signifikan terhadap variabel terikat. Sedangkan variabel X_2^* AKB, X_5^* TPT, dan X_7^* TPAK secara parsial tidak berpengaruh signifikan terhadap variabel terikat (Y^*) yaitu IPM, sehingga didapatkan model akhir berikut :

$$\hat{Y}_R^* = 0,148X_1^* + 0,494X_3^* + 0,181X_4^* + 0,238X_6^* \quad (6)$$

Berdasarkan model persamaan akhir *Gulud Regression* (6), maka model persamaan variabel asal adalah :

$$\hat{Y} = -19,284 + 0,073X_1 + 0,322X_2 + 0,195X_4 + 0,000000842X_6$$

Langkah awal yang dilakukan yaitu melakukan analisis komponen utama menggunakan *software SPSS 23* sehingga diperoleh faktor komponen utama sebagai berikut :

$$W_1 = -0,045 X_1^* + 0,305 X_2^* + 0,108 X_3^* + 0,123 X_4^* + 0,352 X_5^* + 0,205 X_6^* - 0,403 X_7^*$$

$$W_2 = 0,342 X_1^* - 0,514 X_2^* + 0,245 X_3^* + 0,176 X_4^* - 0,124 X_5^* + 0,119 X_6^* - 0,304 X_7^*$$

Diperoleh model regresi dari skor komponen utama dengan variabel terikat sebagai berikut:

$$\hat{Y} = 71,871 + 3,449 W_1 + 3,492 W_2 \quad (7)$$

Uji Signifikansi Parameter Simultan, berdasarkan *output* SPSS nilai F sebesar 306,243 > $F_{(0,05,2,35)} = 3,27$ dan *p-value* sebesar 0,000 < α , maka H_0 ditolak, sehingga dapat disimpulkan bahwa secara simultan bahwa terdapat faktor komponen utama berpengaruh signifikan terhadap variabel terikat Y .

Uji Signifikansi Parameter Parsial dapat dilihat pada Tabel 4.

Tabel 4. Uji Signifikansi Parameter Parsial PCR

Faktor	Koefisien	$ t_{hitung} $	$t_{(0,025;35)}$	$p-value$	Keputusan
Constant	71,871	367,269	2,0301	0,000	H ₀ ditolak
W ₁	0,683	17,389	2,0301	0,000	H ₀ ditolak
W ₂	0,692	17,610	2,0301	0,000	H ₀ ditolak

Berdasarkan kriteria uji dan Tabel 4 maka faktor komponen utama W₁ dan W₂ secara parsial berpengaruh signifikan terhadap variabel terikat Y yaitu IPM, sehingga didapatkan model akhir berikut :

$$\hat{Y} = 71,871 + 1,039 X_1^* - 0,743 X_2^* + 1,227 X_3^* + 1,038 X_4^* + 0,781 X_5^* + 1,122 X_6^* - 0,329 X_7^* \quad (8)$$

Selanjutnya akan dirubah ke dalam variabel asal sehingga :

$$\hat{Y} = 15,270 + 0,521 X_1 - 0,012 X_2 + 0,802 X_3 + 1,133 X_4 + 0,390 X_5 + 0,00004 X_6 - 0,104 X_7 \quad (9)$$

Pemilihan model terbaik dalam kasus multikolinieritas antara metode *Gulud Regression* dan PCR didasarkan pada nilai *Mean Square Error* (MSE). Hasil perbandingan antara kedua metode tersebut disajikan pada Tabel 5.

Tabel 5. Perbandingan Metode *Gulud Regression* dan PCR

Metode	Persamaan Model	MSE
<i>Gulud Regression</i>	$\hat{Y} = -19,284 + 0,073X_1 + 0,322X_2 + 0,195X_4 + 0,000000842X_6$	0,498
PCR	$\hat{Y} = 15,270 + 0,521 X_1 - 0,012 X_2 + 0,802 X_3 + 1,133 X_4 + 0,390 X_5 + 0,00004 X_6 - 0,104 X_7$	1,455

Berdasarkan Tabel 5 menunjukkan nilai MSE yang dihasilkan oleh *Gulud Regression* lebih kecil dibandingkan dengan PCR, sehingga dapat disimpulkan bahwa metode *Gulud Regression* lebih baik dibandingkan metode PCR dalam mengatasi masalah multikolinieritas pada pemodelan Indeks Pembangunan Manusia (IPM) Jawa Timur tahun 2020.

Berdasarkan *output* R terlihat nilai *adjusted R²* sebesar 0,954 hal ini berarti 95,4% IPM Provinsi Jawa Timur tahun 2020 dipengaruhi oleh variabel AHH, AKB, HLS, dan RPPK, sedangkan sisanya 15,4 % dipengaruhi oleh faktor lainnya diluar persamaan atau variabel yang tidak diteliti. Berdasarkan kriteria koefisien determinasi maka disimpulkan model kuat. Setelah model terbaik didapatkan maka diperoleh model akhir menggunakan metode *Gulud Regression* sebagai berikut:

$$\hat{Y} = -19,284 + 0,073X_1 + 0,322X_2 + 0,195X_4 + 0,000000842X_6 \quad (10)$$

Berdasarkan persamaan (10) maka dapat diperoleh :

- Apabila variabel X₁ Angka Harapan Hidup (AHH) naik 1 satuan dan variabel lain bernilai konstan atau tidak mengalami perubahan maka variabel Y Indeks Pembangunan Manusia (IPM) akan mengalami perubahan kenaikan sebesar 0,073 satuan.
- Apabila variabel X₂ Angka Kematian Bayi (AKB) naik 1 satuan dan variabel lain bernilai konstan atau tidak mengalami perubahan maka variabel Y Indeks Pembangunan Manusia (IPM) akan mengalami perubahan kenaikan sebesar 0,322 satuan.
- Apabila variabel X₄ Harapan Lama Sekolah (HLS) naik 1 satuan dan variabel lain bernilai konstan atau tidak mengalami perubahan maka variabel Y Indeks Pembangunan Manusia (IPM) akan mengalami perubahan kenaikan sebesar 0,195 satuan.

- Apabila variabel X_6 Rata-rata Pengeluaran Perkapita (RPPK) naik 1 satuan dan variabel lain bernilai konstan atau tidak mengalami perubahan maka variabel Y Indeks Pembangunan Manusia (IPM) akan mengalami perubahan kenaikan sebesar 0,000000842 satuan.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dipaparkan didapatkan kesimpulan sebagai berikut:

1. Berdasarkan analisis uji non multikolinearitas, terdapat kasus multikolinearitas yang terjadi pada faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia Provinsi Jawa Timur tahun 2020, yaitu variabel Rata-rata Lama Sekolah (RLS) dan Rata-rata Pengeluaran Perkapita (RPPK).
2. Metode *Gulud Regression* dan *Principal Component Regression* (PCR) dapat digunakan untuk menyelesaikan masalah multikolinieritas dengan baik, dibuktikan dengan nilai VIF yang sebelumnya lebih dari 5 setelah penerapan metode menjadi tidak lebih dari 5.
3. Berdasarkan perbandingan nilai *Mean Square Error* (MSE) menunjukkan bahwa penanganan kasus multikolinieritas menggunakan metode *Gulud Regression* lebih baik dibandingkan dengan metode *Principal Komponen Regression* (PCR) terhadap pemodelan Indeks Pembangunan Manusia Provinsi Jawa Timur tahun 2020 dan didapatkan model akhir sebagai berikut:

$$\hat{Y} = -19,284 + 0,073X_1 + 0,322X_2 + 0,195X_4 + 0,000000842X_6$$

DAFTAR PUSTAKA

- BPS. 2020. *Indeks Pembangunan Manusia 2020*. Jakarta: Badan Pusat Statistik.
- BPS. 2015. *Republik Indonesia Indeks Pembangunan Manusia 2014*. <https://www.kemenpppa.go.id/lib/uploads/list/254a4-ipm-2014-bps.pdf>. Diakses : 5 Maret 2022
- Hoerl, A. E., & Kennard, R. W. 1970. *Ridge Regression: Biased Estimation For Nonorthogonal Problems*. *Journal Storage*, 12(1), 55-67.
- Ifadah, A. 2011. *Analisis Metode Principal Component Analysis (Komponen Utama) Dan Regresi Ridge Dalam Mengatasi Dampak Multikolinearitas Dalam Analisis Regresi Linear Berganda*. <http://lib.unnes.ac.id/2434/1/3416.pdf>. Diakses : 20 Juli 2022.
- Lestari, J. S., Farida, U., & Chamidah, S. 2020. *Pengaruh Kepemimpinan, Kedisiplinan, Dan Lingkungan Kerjaterhadap Prestasi Kerja Guru*. *ASSET: Jurnal Manajemen Dan Bisnis* 2 (2): 38–55. <https://doi.org/10.24269/asset.v2i2.2388>.
- Meiryani. 2021. *Memahami Uji Autokorelasi Dalam Model Regresi*. <https://accounting.binus.ac.id/2021/08/06/memahami-uji-autokorelasi-dalam-model-regresi/>. Diakses 5 Agustus 2022.
- Montgomery, D. C., & Peck, E. A. 1991. *Introduction to Linear Regression Analysis*. 2nd ed. New York: John Wiley & Sons.
- Putra, D. M., & Ratnasari, V. 2015. *Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Dengan Menggunakan Metode Regresi Logistik Ridge*. *Jurnal Sains Dan Seni ITS* 4 (2): 175–80.
- Sanchez, G. 2013. *PLS Path Modeling with R*. http://gastonsanchez.com/PLS_Path_Modeling_with_R.pdf. Diakses : 16 Juni 2022
- Siombing, R. A. 2017. *Metode Penelitian Metode Penelitian*. *Metode Penelitian Kualitatif*, no. 17: 43.

- Siregar, S. 2015. *Statistika Parametrik Untuk Penelitian Kuantitatif Dilengkapi Dengan Perhitungan Manual dan Aplikasi SPSS Versi 17*. Jakarta: Bumi Aksara.
- Umar, H. 2011. *Metode Penelitian untuk skripsi dan tesis bisnis*. Jakarta: Rajawali Pers.
- Wahab, W. 2020. *Faktor-Faktor Yang Mempengaruhi Disiplin Kerja Karyawan Pada PT. Indomas Rezeki Jaya Kabupaten Pelalawan*. *Eko Dan Bisnis (Riau Economics and Business Reviewe)* 11 (2): 127–40.
- Zou, H., & Hastie, T. 2005. *Regularization and variable selection via the elastic net*. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2), 301–320.