

PEMODELAN INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH MENGUNAKAN METODE REGRESI *RIDGE* DAN REGRESI *STEPWISE*

Erna Sulistianingsih¹, Suparti², Dwi Ispriyanti³

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

e-mail: ernasulistia26@gmail.com

DOI: 10.14710/J.GAUSS.11.3.468-477

Article Info:

Received: 2022-09-07

Accepted: 2022-12-07

Available Online: 2023-01-03

Keywords:

HDI; Multicollinearity; Ridge
Regression; Stepwise Regression.

Abstract: The Human Development Index (HDI) is an important indicator in measuring the success of national development. Central Java with a high population can be considered as an obstacle and a driver of development. To find out the factors that affect HDI, it is necessary to make a model. One of the statistical methods that can be used is multiple linear regression analysis. However, in modeling multiple linear regression there are assumptions that must be met, namely linearity, normality, homoscedasticity, non-autocorrelation, and non-multicollinearity. If the non-multicollinearity assumption is not met, then another alternative is needed to estimate the regression parameters. Several methods that can be used are ridge regression and stepwise regression methods. The best model selection is done by looking at the smallest Mean Square Error (MSE) value. In this study, ridge and stepwise regression were applied to Central Java HDI data in 2021 and the factors that influence it, namely life expectancy at birth, expected years of schooling, average length of schooling, per capita expenditure, percentage of poor people, and unemployment open. Based on the Variance Inflation Factor (VIF) value of more than 10, it can be concluded that there is a multicollinearity violation. Modeling with stepwise regression produces the best model, with the smallest MSE value. The R square model value of 0,99 indicates that the model is included in the criteria for a strong model.

1. PENDAHULUAN

Kesejahteraan masyarakat merupakan tujuan utama dalam pembangunan suatu negara. Pembangunan nasional bangsa Indonesia perlu dilakukan secara menyeluruh dan berkelanjutan guna meningkatkan kualitas hidup dan kesejahteraan masyarakat. Salah satu yang menjadi indikator penting dalam mengukur keberhasilan pembangunan nasional adalah dengan melihat Indeks Pembangunan Manusia (IPM).

Indeks Pembangunan Manusia dipengaruhi beberapa faktor. Faktor-faktor yang mempengaruhi merupakan variabel bebas (X) dan faktor yang dipengaruhi merupakan variabel terikat (Y). Suatu metode statistik yang dapat digunakan untuk mengetahui hubungan antara variabel bebas dan variabel terikat yaitu metode regresi linier berganda. Menurut Kurniawan & Yuniarto (2016), regresi linier berganda harus memenuhi asumsi linieritas, normalitas, homoskedastisitas, non autokorelasi, dan non multikolinieritas. Apabila terjadi pelanggaran pada asumsi linieritas, normalitas, homoskedastisitas, dan non autokorelasi maka dapat dilakukan penanganan dengan menggunakan transformasi seperti transformasi logaritma dan akar kuadrat. Sedangkan untuk asumsi non multikolinieritas yang tidak terpenuhi, maka dapat diatasi dengan menggunakan metode lain seperti metode regresi *ridge* dan regresi *stepwise*.

Penelitian ini bertujuan untuk memodelkan IPM Provinsi Jawa Tengah dan variabel-variabel yang diduga mempengaruhinya. Dari hasil pemodelan kedua metode kemudian dibandingkan untuk mendapatkan model terbaik.

2. TINJAUAN PUSTAKA

Indeks Pembangunan Manusia (IPM) diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 dan dipublikasikan secara berkala dalam laporan tahunan *Human Development Report* (HDR) (BPS, 2014). IPM dibangun melalui pendekatan tiga dimensi dasar yaitu umur panjang dan sehat, pengetahuan, dan kehidupan yang layak (BPS, 2015). Menurut BPS (2015), angka harapan hidup saat lahir dapat digunakan untuk mengukur dimensi kesehatan, rata-rata lama sekolah, dan harapan lama sekolah dapat digunakan untuk mengukur dimensi pengetahuan, dan pengeluaran per kapita disesuaikan digunakan untuk mengukur dimensi hidup layak. Menurut Wasilaine *et al.* (2014), IPM dapat dipengaruhi oleh persentase penduduk miskin, selain itu, pengangguran terbuka juga dapat mempengaruhi IPM (Putra, 2015).

Menurut BPS (2020), Jawa Tengah berada di urutan ketiga sebagai provinsi dengan jumlah penduduk terbanyak di Indonesia. Jumlah penduduk yang tinggi dapat dianggap sebagai penghambat pembangunan dan dapat pula dianggap sebagai pemacu pembangunan. Penduduk sebagai pemacu pembangunan karena dapat menciptakan barang dan jasa yang kemudian menggerakkan berbagai macam kegiatan ekonomi. Namun, padatnya penduduk juga dapat menimbulkan masalah yaitu banyaknya pengangguran. Menurut BPS (2021), IPM di Jawa Tengah pada tahun 2021 sebesar 72,16 lebih tinggi dibanding dengan tahun sebelumnya yakni 71,87.

Analisis regresi linier berganda yaitu hubungan secara linier antara dua atau lebih variabel bebas terhadap variabel terikat. Menurut Montgomery & Peck (1991), model umum regresi linier berganda dengan k variabel bebas dan n pengamatan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (1)$$

dengan

$i = 1, 2, 3, \dots, n$ (banyaknya pengamatan)

$j = 1, 2, 3, \dots, k$ (banyak variabel bebas)

X_{ij} : variabel bebas ke- j pada pengamatan ke- i

Y_i : variabel terikat pada pengamatan ke- i

β_0 : *intercept* dari model

β_j : koefisien regresi ke- j

ε_i : *error* pada pengamatan ke- i , diasumsikan $\varepsilon \sim NID(0, \sigma^2)$

Menurut Montgomery & Peck (1991), estimasi parameter model (1) dapat diperoleh dengan menggunakan metode kuadrat terkecil dengan meminimumkan jumlah kuadrat *error* sehingga diperoleh penaksir untuk metode kuadrat terkecil adalah:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2)$$

Pada regresi linier berganda diperlukan uji asumsi klasik. Berikut uji asumsi klasik yang digunakan:

- a. Uji linieritas

Uji linieritas adalah uji yang dilakukan untuk menentukan apakah terdapat hubungan linier antara variabel bebas dengan variabel terikat. Linieritas dapat diuji dengan menggunakan uji *Ramsey Reset Test* dengan hipotesis sebagai berikut:

H_0 : Model linier

H_1 : Model non linier

Dengan kriteria penolakan yaitu H_0 ditolak jika nilai p -value $< \alpha$.

b. Uji normalitas

Uji normalitas dapat dilakukan menggunakan uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut:

H_0 : $F(x) = F_0(x)$ (residual berdistribusi normal).

H_1 : $F(x) \neq F_0(x)$ (residual tidak berdistribusi normal).

Sedangkan statistik ujinya dapat dicari dengan rumus sebagai berikut:

$$D = \sup|F(x) - F_0(x)| \quad (3)$$

Untuk kriteria penolakannya yaitu H_0 ditolak jika $D \geq W_{(\alpha,n)}$ atau p -value $< \alpha$.

c. Uji homoskedastisitas

Uji homoskedastisitas dilakukan dengan menggunakan uji *Breusch-Pagan* dengan hipotesis sebagai berikut:

H_0 : $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$, (terjadi homoskedastisitas).

H_1 : minimal ada satu i dengan $\sigma_i^2 \neq \sigma^2$ (terjadi heteroskedastisitas).

Sedangkan statistik uji yang digunakan mengikuti distribusi *chi-square* dengan rumus sebagai berikut:

$$BP = nR^2 \sim \chi_{n-1}^2 \quad (4)$$

Untuk kriteria penolakannya yaitu H_0 di tolak jika $BP > \chi_{(\alpha,n-1)}^2$ atau p -value $< \alpha$.

d. Uji non autokorelasi

Uji non autokorelasi dilakukan dengan menggunakan uji *Durbin-Watson* dengan hipotesis sebagai berikut:

H_0 : Tidak ada autokorelasi pada *error*

H_1 : Ada autokorelasi pada *error*

Sedangkan statistik uji yang digunakan dapat dicari dengan rumus sebagai berikut:

$$d = \frac{\sum_{i=1}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2} \quad (5)$$

Untuk kriteria penolakannya yaitu H_0 ditolak jika $d < dL$ atau $d > 4-dL$ dan H_0 diterima jika $dU < d < 4-dU$

e. Uji multikolinieritas

Uji multikolinieritas dilakukan dengan menggunakan nilai *Variance Inflation Factor* (VIF) dengan rumus:

$$VIF_j = \frac{1}{(1-R_j^2)}, j = 1, 2, \dots, k \quad (6)$$

Menurut Montgomery & Peck (1991), jika nilai VIF < 10 , maka dapat dikatakan bahwa tidak terdapat multikolinieritas.

Pada regresi berganda dilakukan uji signifikansi dengan dua uji yaitu uji signifikansi secara simultan (uji F) dan uji signifikansi secara parsial (uji t). Kedua uji tersebut dijelaskan sebagai berikut:

a. Uji signifikansi secara simultan (uji F)

Uji F atau disebut juga dengan uji simultan adalah uji semua variabel bebas secara keseluruhan dan bersamaan di dalam suatu model (Kurniawan & Yuniarto, 2016). Uji F digunakan untuk melihat apakah variabel bebas secara keseluruhan berpengaruh signifikan terhadap variabel terikat atau tidak. Adapun untuk hipotesisnya dapat ditulis sebagai berikut:

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ (seluruh variabel bebas secara simultan tidak berpengaruh terhadap variabel terikat).

H_1 : minimal ada satu j dengan $\beta_j \neq 0$ (ada variabel bebas yang berpengaruh terhadap variabel terikat, untuk $j = 1, 2, \dots, k$).

Statistik uji:

$$F_{\text{hitung}} = \frac{SS_R / k}{SS_E / (n-k-1)} = \frac{MSR}{MSE} \quad (7)$$

Kriteria penolakan yaitu H_0 ditolak jika $F_{\text{hitung}} > F_{(\alpha, k, n-k-1)}$ atau $p\text{-value} < \alpha$.

b. Uji signifikansi secara parsial (uji t)

Uji t digunakan untuk melihat pengaruh dari masing-masing variabel bebas terhadap variabel terikatnya (Kurniawan & Yuniarto, 2016).

Hipotesis:

$H_0: \beta_j = 0$ (masing-masing variabel bebas tidak berpengaruh signifikan terhadap variabel terikat).

$H_1: \beta_j \neq 0$ (masing-masing variabel bebas berpengaruh signifikan terhadap variabel terikat, untuk $j = 1, 2, \dots, k$).

Statistik uji:

$$t_{\text{hitung}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (8)$$

Kriteria penolakan yaitu H_0 ditolak jika $|t_{\text{hitung}}| > t_{(\frac{\alpha}{2}, n-k-1)}$ atau $p\text{-value} < \alpha$.

Sebelum dilakukan analisis menggunakan regresi *ridge* dan *stepwise* perlu dilakukan standarisasi data karena perbedaan ukuran satuan dari variabel yang digunakan untuk analisis data. Menurut Manurung *et al.* (2005) dalam (Arrasyid *et al.*, 2021), proses standarisasi data dilakukan dengan mengurangi data pengamatan dengan nilai rata-rata variabel data, kemudian dibagi dengan standar deviasi variabel dari data tersebut. Sehingga rumus standarisasi data dapat ditulis:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \quad \text{dan} \quad X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j}} \quad (9)$$

Regresi *ridge* adalah metode estimasi koefisien regresi yang diperoleh dengan cara penambahan penalti penyusutan λ pada diagonal matriks $X'X$ (Wasilaine *et al.*, 2014). Menurut Mulyasari *et al.* (2015), Pendugaan koefisien regresi *ridge* dapat diperoleh dengan menggunakan metode *Penalized Least Square* (PLS) yaitu dengan meminimumkan persamaan:

$$PLS = \sum_{i=1}^n (Y_i^* - \beta_0^* - \sum_{j=1}^k \beta_j^* X_{ij}^*)^2 + \lambda \sum_{j=1}^k \beta_j^{*2} \quad (10)$$

Pada persamaan (7) suku pertama adalah jumlah kuadrat galat, sedangkan suku kedua adalah penalti penyusutan dalam meminimumkan jumlah kuadrat sisaan. Nilai estimasi parameter β^* diperoleh dengan menghitung nilai turunan parsial *PLS* terhadap β^* , kemudian di sama dengankan nol, sehingga diperoleh penaksir model regresi *ridge* yaitu:

$$\hat{\beta}_R^* = (X^{*'}X^* + \lambda I)^{-1}X^{*'}Y^* \quad (11)$$

Menurut Hoerl dan Kennard (1970) dalam Rashwan & El-Dereny (2011), menyarankan metode grafik yang disebut *ridge trace* untuk memilih nilai parameter *ridge* λ . *Ridge trace* adalah plot dari estimator regresi *ridge* dengan berbagai kemungkinan nilai tetapan bias λ (Putri, 2011). Pemilihan nilai λ didasarkan pada nilai MSE terkecil dari masing-masing nilai λ yang ada yaitu $0 < \lambda < 1$.

Metode *stepwise* adalah metode yang merupakan gabungan dari metode regresi *forward* dan *backward* (Kurniawan & Yuniarto, 2016). Prosedur yang dilakukan yaitu dengan menyusupkan satu demi satu variabel bebas sampai diperoleh persamaan regresi yang memuaskan (Pujilestari *et al.*, 2017). Variabel yang terpilih untuk dimasukkan ke dalam model yaitu variabel yang mempunyai korelasi tertinggi dan signifikan terhadap variabel terikat. Menurut Kurniawan & Yuniarto (2016), model regresi *stepwise* standarisasi dengan k variabel bebas dan n pengamatan adalah sebagai berikut:

$$Y_{Si}^* = \beta_1^*X_{i1}^* + \beta_2^*X_{i2}^* + \dots + \beta_k^*X_{ik}^* + \varepsilon_i^* \quad (12)$$

Setelah diperoleh model kemudian menentukan model terbaik dengan melihat nilai *Mean Square Error* (MSE). MSE adalah rata-rata kesalahan kuadrat diantara nilai aktual dan nilai peramalan. Semakin tinggi nilai MSE maka semakin buruk modelnya. MSE dapat dihitung dengan rumus:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (13)$$

Kemudian dihitung ukuran kebaikan koefisien determinasi yang menunjukkan sejauh mana kontribusi variabel bebas dalam model regresi mampu menjelaskan variasi dari variabel terikatnya. Nilai koefisien determinasi yaitu antara 0 dan 1. Jika nilai mendekati 1, artinya variabel bebas memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variabel terikat. Namun, jika nilai R^2 semakin kecil, artinya kemampuan variabel – variabel bebas dalam menjelaskan variabel terikat cukup terbatas (Ghozali, 2016). Rumus R^2 dapat ditulis dalam bentuk:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (14)$$

3. METODE PENELITIAN

Penelitian menggunakan data sekunder yang diambil dari website Badan Pusat Statistik (BPS) Provinsi Jawa Tengah Tahun 2021 yang meliputi data Indeks Pembangunan Manusia (IPM), Angka Harapan Hidup (AHH) saat lahir, Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Pengeluaran per kapita Disesuaikan, Persentase Penduduk Miskin, dan Pengangguran Terbuka.

Penelitian ini menggunakan *software* RStudio, SPSS, dan *Microsoft Excel* 2013. Berikut tahapan analisis data yang digunakan dalam penelitian ini:

1. Mencari data yang akan digunakan dalam penulisan tugas akhir meliputi data Indeks Pembangunan Manusia, angka harapan hidup saat lahir, harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita disesuaikan, persentase penduduk miskin, dan pengangguran terbuka.
2. Menentukan variabel terikat dan variabel bebas dari data yang diperoleh.
3. Melakukan standarisasi data karena data memiliki satuan yang berbeda.
4. Melakukan pemodelan regresi linier dengan metode kuadrat terkecil.

5. Melakukan uji asumsi klasik meliputi uji linieritas, uji normalitas, uji homoskedastisitas, uji non autokorelasi, dan uji non multikolinieritas.
6. Melakukan penanganan terhadap masalah multikolinieritas apabila asumsi ini tidak terpenuhi dengan menggunakan regresi *ridge* dan regresi *stepwise* yaitu:
 - A. Regresi *Ridge*
 - a. Mencari estimator parameter regresi *ridge* dengan berbagai nilai λ ($0 < \lambda < 1$).
 - b. Melakukan pemodelan regresi *ridge*.
 - c. Menghitung nilai MSE masing-masing model regresi *ridge* yang terbentuk dengan berbagai nilai λ ($0 < \lambda < 1$).
 - d. Mencari model regresi *ridge* terbaik berdasarkan nilai MSE terkecil.
 - e. Melakukan pengujian dengan menggunakan uji F dan uji t.
 - f. Menghitung nilai *Mean Square Error* (MSE).
 - B. Regresi *Stepwise*
 - a. Menghitung nilai koefisien korelasi antara semua variabel bebas dan variabel terikat.
 - b. Melakukan pemodelan regresi *stepwise*.
 - c. Menghitung nilai MSE masing-masing model regresi *stepwise* yang terbentuk.
 - d. Mencari model regresi *stepwise* terbaik berdasarkan nilai MSE terkecil.
 - e. Melakukan pengujian dengan uji F dan uji t.
 - f. Menghitung nilai MSE.
7. Pemilihan model terbaik antara regresi *ridge* dengan regresi *stepwise* dengan menggunakan MSE yang terkecil.
8. Mengembalikan persamaan regresi ke bentuk variabel asli.
9. Menghitung ukuran kebaikan *R Square*

4. HASIL DAN PEMBAHASAN

Adanya perbedaan satuan data pada variabel maka sebelum dilakukan analisis perlu di standarisasi terlebih dahulu, dengan menggunakan *Ordinary Least Square* diperoleh persamaan regresi linier berganda setelah di standarisasi yaitu:

$$\hat{Y}^* = 0,1879X_1^* + 0,1959X_2^* + 0,3623X_3^* + 0,3478X_4^* - 0,0069X_5^* - 0,0013X_6^*$$

Setelah diperoleh model persamaan regresi kemudian dilakukan uji asumsi klasik sebagai berikut:

a. Uji Linieritas

Untuk menguji linieritas dapat digunakan uji *Ramsey Reset Test*. Berdasarkan uji *Ramsey Reset Test* diperoleh nilai *p-value* sebesar 0,9778. Berdasarkan kriteria penolakan maka H_0 gagal ditolak sehingga dapat disimpulkan bahwa model linier yang berarti asumsi linieritas terpenuhi.

b. Uji Normalitas

Pada uji normalitas, diperoleh *p-value*=0,4224 > $\alpha = 0,05$, sehingga dapat disimpulkan bahwa residual berdistribusi normal.

c. Uji Homoskedastisitas

Untuk menguji homoskedastisitas, uji yang digunakan yaitu uji Breusch-Pagan. Melalui uji Breusch-Pagan diperoleh nilai BP sebesar 4,9587 dan nilai *p-value* sebesar 0,5491, sedangkan pada tabel *Chi Square* diperoleh $\chi^2_{(0,05;34)} = 48,6024$, berdasarkan kriteria penolakan maka H_0 gagal ditolak, sehingga dapat disimpulkan bahwa terjadi homoskedastisitas atau asumsi ini terpenuhi.

d. Uji Non Autokorelasi

Uji non autokorelasi dapat dilakukan dengan menggunakan uji Durbin Watson. Diperoleh nilai Durbin Watson sebesar 1,9182. Pada tabel Durbin Watson dengan $n=35$ dan

k=6, maka akan diperoleh nilai $dL=1,0974$ dan $dU=1,8835$. Sehingga nilai $4-dL=2,9026$ dan $4-dU=2,1165$. Karena nilai Durbin Watson terletak diantara dU dan $4-dU$, maka dapat disimpulkan bahwa tidak ada autokorelasi pada *error* atau asumsi ini terpenuhi.

e. Uji Multikolinieritas

Nilai VIF pada variabel rata-rata lama sekolah adalah 11,3219 lebih besar dari 10, sehingga dapat disimpulkan bahwa terjadi masalah multikolinieritas antara variabel bebas. Oleh karena itu diperlukan suatu penanganan untuk mengatasi masalah multikolinieritas menggunakan regresi *ridge* dan *stepwise*.

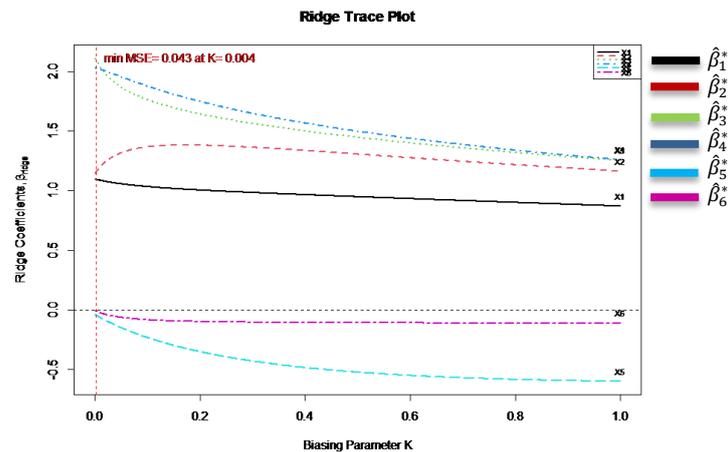
Dalam proses pengestimasiian regresi *ridge*, pemilihan nilai λ merupakan hal yang paling penting dalam penelitian ini. Penentuan nilai λ melalui pendekatan nilai VIF dan gambar *ridge trace*. Diperoleh nilai VIF dari berbagai kemungkinan nilai λ yang disajikan pada Tabel 1.

Tabel 1. Nilai VIF dengan Berbagai Nilai λ

Nilai λ	VIF X_1^*	VIF X_2^*	VIF X_3^*	VIF X_4^*	VIF X_5^*	VIF X_6^*	MSE
0,000	2,8385	8,5286	11,3219	3,4779	2,0344	1,8512	0,0451
0,001	2,8159	8,2697	10,9402	3,4357	2,0241	1,8236	0,0441
0,002	2,7937	8,0238	10,5783	3,3947	2,0138	1,7972	0,0435
0,003	2,7717	7,7901	10,2349	3,3549	2,0037	1,7720	0,0432
0,004	2,7501	7,5677	9,9088	3,3162	1,9936	1,7479	0,0431
0,005	2,7288	7,3560	9,5987	3,2785	1,9837	1,7249	0,0433
0,006	2,7078	7,1541	9,3036	3,2419	1,9739	1,7027	0,0438
0,007	2,6870	6,9616	9,0227	3,2061	1,9641	1,6815	0,0444
0,008	2,6666	6,7777	8,7549	3,1713	1,9545	1,6611	0,0452
0,009	2,6464	6,6021	8,4994	3,1373	1,9449	1,6414	0,0463
0,010	2,6265	6,4341	8,2556	3,1041	1,9355	1,6225	0,0474
0,020	2,4411	5,0900	6,3233	2,8104	1,8453	1,4661	0,0653
0,030	2,2773	4,1683	5,0236	2,5695	1,7624	1,3510	0,0893
0,040	2,1314	3,5035	4,1051	2,3664	1,6857	1,2617	0,1158
0,050	2,0006	3,0050	3,4301	2,1917	1,6144	1,1895	0,1431
0,060	1,8828	2,6192	2,9185	2,0395	1,5479	1,1292	0,1705
0,070	1,7761	2,3130	2,5205	1,9052	1,4857	1,0777	0,1975
0,080	1,6791	2,0648	2,2042	1,7859	1,4275	1,0327	0,2241
0,090	1,5905	1,8600	1,9483	1,6790	1,3728	0,9929	0,2503
0,100	1,5094	1,6885	1,7380	1,5827	1,3214	0,9572	0,2760
0,200	0,9680	0,8333	0,7618	0,9723	0,9393	0,7214	0,5108
0,300	0,6853	0,5261	0,4528	0,6730	0,7060	0,5848	0,7170
0,400	0,5166	0,3739	0,3113	0,5002	0,5523	0,4907	0,9028
0,500	0,4068	0,2853	0,2332	0,3902	0,4455	0,4205	1,0726
0,600	0,3309	0,2282	0,1847	0,3152	0,3680	0,3659	1,2298
0,700	0,2759	0,1889	0,1521	0,2615	0,3099	0,3221	1,3767
0,800	0,2347	0,1603	0,1290	0,2215	0,2652	0,2862	1,5151
0,900	0,2027	0,1387	0,1117	0,1909	0,2299	0,2563	1,6465
1,000	0,1775	0,1219	0,0985	0,1667	0,2016	0,2310	1,7720

Berdasarkan Tabel 1, nilai $\lambda=0$ sampai dengan $\lambda=1$, VIF semakin lama semakin kecil. Diperoleh nilai VIF untuk masing-masing variabel bebas X_1, X_2, X_3, X_4, X_5 , dan X_6 sebesar 2,7501; 7,5677; 9,9088; 3,3162; 1,9936; dan 1,7479 dengan menggunakan nilai MSE terkecil sebesar 0,0431. Dengan nilai $\lambda = 0,004$ diperoleh koefisien $\hat{\beta}_1^*=0,1872$; $\hat{\beta}_2^*=0,2004$; $\hat{\beta}_3^*=0,3565$; $\hat{\beta}_4^*=0,3470$; $\hat{\beta}_5^*=-0,0087$; dan $\hat{\beta}_6^*=-0,0027$.

Atas dasar koefisien estimator regresi *ridge* dapat dibuat *Ridge Trace* yaitu suatu plot antara nilai λ dengan koefisien $\hat{\beta}_R^*$ yang disajikan pada Gambar 1.



Gambar 1. Ridge Trace

Koefisien estimator *ridge* dan VIF yang diperoleh dari berbagai kemungkinan nilai λ , dipilih nilai λ dengan MSE terkecil yaitu pada $\lambda = 0,004$ dengan MSE = 0,0431. Sehingga model regresi *ridge* yang diperoleh berdasarkan *output* RStudio pada Lampiran 5 adalah sebagai berikut:

$$\hat{Y}_R^* = 0,1872X_1^* + 0,2004X_2^* + 0,3565X_3^* + 0,3470X_4^* - 0,0087X_5^* - 0,0027X_6^*$$

Dari uji secara simultan, diperoleh nilai $F_{hitung}=3756,1650$ dan $F_{tabel}=2,45$. Karena $F_{hitung}>F_{tabel}$, sehingga dapat disimpulkan bahwa semua variabel bebas secara simultan berpengaruh signifikan terhadap variabel IPM.

Pada uji parsial, diperoleh nilai t_{hitung} untuk variabel bebas X_1, X_2, X_3 , dan X_4 yaitu yaitu 20,4957; 13,1169; 19,7336; dan 30,022, serta $t_{tabel}=2,0484$. Karena $|t_{hitung}|>t_{tabel}$, maka variabel X_1, X_2, X_3 , dan X_4 memiliki pengaruh signifikan terhadap IPM. Sehingga model regresi baru untuk variabel yang signifikan sebagai berikut:

$$\hat{Y}_R^* = 0,1911X_1^* + 0,1983X_2^* + 0,3613X_3^* + 0,3478X_4^*$$

Sedangkan pada metode *stepwise* diperoleh dengan memasukan variabel bebas yang memiliki koefisien korelasi tertinggi dengan variabel terikat. Diperoleh persamaan model regresi *stepwise* sebagai berikut:

Tabel 3. Persamaan Model dan Nilai Mean Square Error (MSE)

Model	Persamaan Model	MSE
1	$\hat{Y}_S^* = 0,970X_3^*$	0,062
2	$\hat{Y}_S^* = 0,694X_3^*+0,332X_4^*$	0,027
3	$\hat{Y}_S^* = 0,195X_1^*+0,526X_3^*+0,365X_4^*$	0,009
4	$\hat{Y}_S^* = 0,191X_1^*+0,195X_2^*+0,365X_3^*+0,349X_4^*$	0,001

Berdasarkan Tabel 3 diperoleh model *stepwise* dengan nilai MSE masing-masing model. Dari ke empat model *stepwise* yang terbentuk dapat dilihat multikolinieritas yang sudah teratasi, nilai VIF dapat dilihat pada Tabel 4.

Tabel 4. Nilai VIF Variabel Bebas Model *Stepwise*

	Model	VIF
1	RLS (X3)	1,000
2	RLS (X3)	3,222
	Pengeluaran per kapita (X4)	3,222
3	RLS (X3)	4,790
	Pengeluaran per kapita (X4)	3,282
	AHH (X1)	2,118
4	RLS (X3)	8,767
	Pengeluaran per kapita (X4)	3,323
	AHH (X1)	2,121
	HLS (X2)	5,854

Pada Tabel 4 dapat diketahui bahwa model 1, model 2, model 3, dan model 4 nilai VIF lebih kecil dari 10, sehingga masalah multikolinieritas sudah teratasi. Berdasarkan Tabel 3 diperoleh model terbaik *stepwise* dengan nilai MSE terkecil yaitu pada model 4 dengan persamaan model sebagai berikut:

$$\hat{Y}_S^* = 0,191X_1^* + 0,195X_2^* + 0,365X_3^* + 0,349X_4^*$$

Dari uji secara simultan, diperoleh nilai $F_{hitung}=5750,002$ dan $F_{tabel}=2,45$. Karena $F_{hitung}>F_{tabel}$, sehingga dapat disimpulkan bahwa semua variabel bebas secara simultan berpengaruh signifikan terhadap variabel IPM.

Pada uji parsial, diperoleh nilai t_{hitung} untuk variabel bebas $X_1, X_2, X_3,$ dan X_4 yaitu 18,719; 29,022; 19,865; dan 12,240, serta $t_{tabel}=2,0484$. Karena $|t_{hitung}|$ dari variabel $X_1, X_2, X_3,$ dan $X_4 > t_{tabel}$, maka variabel $X_1, X_2, X_3,$ dan X_4 memiliki pengaruh signifikan terhadap IPM.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dipaparkan pada bab sebelumnya diperoleh kesimpulan sebagai berikut:

1. Metode regresi *ridge* mengatasi multikolinieritas dengan cara penambahan nilai λ , sedangkan penanganan multikolinieritas pada regresi *stepwise* dengan cara memasukan variabel bebas yang signifikan terhadap variabel terikat.
2. Diperoleh model persamaan regresi *ridge* dan regresi *stepwise* dengan variabel bebas yang signifikan terhadap IPM meliputi AHH, HLS, RLS, dan Pengeluaran per kapita sebagai berikut:

$$\hat{Y}_R = 5,6690 + 0,4645 X_1 + 0,9496X_2 + 1,2903X_3 + 0,0009X_4$$

$$\hat{Y}_S = 5,7533 + 0,4643 X_1 + 0,9338X_2 + 1,3036X_3 + 0,0009X_4$$

3. Model terbaik diperoleh model regresi *stepwise* dengan nilai MSE sebesar 0,001 dan *R Squared* sebesar 0,99. Sehingga model yang diperoleh dengan menggunakan metode *stepwise* termasuk model yang baik.

DAFTAR PUSTAKA

- Arrasyid, A. H., Ispriyanti, D., & Hoyyi, A. (2021). Metode Modified Jackknife Ridge Regression Dalam Penanganan Multikolinieritas (Studi Kasus Indeks Pembangunan Manusia Di Jawa Tengah). *Jurnal Gaussian*, 10(1), 104–113.
- BPS. (2014). *Indeks Pembangunan Manusia*. Badan Pusat Statistik. <https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html#subjekViewTab1>
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 23* (8th ed.). Semarang: Universitas Diponegoro.
- Kurniawan, R., & Yuniarto, B. (2016). *Analisis Regresi Dasar dan Penerapannya dengan R*. Jakarta: Kencana.
- Montgomery, D. C., & Peck, E. A. (1991). *Introduction to Linear Regression Analysis* (2nd ed.). New York: John Wiley & Sons.
- Mulyasari, R., Nugroho, S., & Rizal, J. (2015). Regresi Ridge untuk mengatasi Multikolinearitas. *E-Journal Statistika*, 1, 1–12.
- Pujilestari, S., Dwidayati, N., & Sugiman. (2017). Pemilihan Model Regresi Linear Berganda Terbaik pada Kasus Multikolinearitas Berdasarkan Metode Principal Component Analysis (PCA) dan Metode Stepwise. *Unnes Journal of Mathematics*, 6(1).
- Putri, A. P. (2011). Penggunaan Metode Ridge Trace Dan Variance Inflation Factors (Vif) Pada Regresi Ridge. E-Journal *Universitas Negeri Yogyakarta*.
- Rashwan, N. I., & El-Dereny, M. (2011). Solving Multicollinearity Problem Using Ridge Regression Models. *Int. J. Contemp. Math. Sciences*, 6(12), 585–600.
- Wasilaine, T. L., Talakua, M. W., & Lesnussa, Y. A. (2014). Model Regresi Ridge untuk Mengatasi Model Regresi Linier Berganda yang Mengandung Multikolinieritas (Studi Kasus: Data Pertumbuhan Bayi di Kelurahan Namaelo RT 001, Kota Masohi).