 ISSN: 2502-6526

PENERAPAN METODE REGRESI GULUD DAN REGRESI KOMPONEN UTAMA DALAM MENGATASI PENYIMPANGAN MULTIKOLINEARITAS PADA ANALISIS REGRESI LINEAR BERGANDA

Sri Siska Wirdaniyati[1], Edy Widodo[2]
[1]Mahasiswa Prodi Statistika Universitas Islam Indonesia [2]Dosen Prodi Statistika Universitas Islam Indonesia
srisiskaw08@gmail.com, edykafifa@gmail.com

Abstrak

Dalam mengkaji hubungan atau pengaruh dua atau lebih variabel bebas terhadap variabel terikat, maka model regresi yang digunakan adalah model regresi linear berganda. Pada analisis regresi linear berganda perlu dilakukan pengujian asumsi klasik agar mendapatkan penaksir yang tidak bias, linear, dan terbaik *(Best Linear Unbiased Estimator/BLUE).* Salah satu asumsi klasik yaitu uji multikolinearitas yang merupakan pegujian untuk mengetahui apakah terdapat korelasi antara variabel bebas pada model regresi. Tujuan dari penelitian ini adalah untuk mengetahui solusi/metode terbaik dan paling efektif dalam mengatasi penyimpangan multikolinearitas sehingga dapat memberikan permodelan regresi linear berganda terbaik. Metode yang digunakan adalah regresi gulud *(ridge regression)* dan regresi komponen utama *(principal component regression)* dengan menggunakan empat data simulasi yang berbeda. Berdasarkan hasil penelitian diketahui bahwa keempat data tersebut memiliki nilai *RMSE (Root Mean Square Error)* lebih kecil dan nilai *Adjusted R Square* (R^2_{Adj}) lebih besar pada regresi gulud daripada regresi komponen utama, sehingga metode regresi gulud merupakan solusi terbaik dan paling efektif dalam mengatasi penyimpangan multikolinearitas pada analisis regresi linear berganda.

Kata Kunci: *Multikolinearitas, Regresi Gulud, Regresi Komponen Utama.*

1. PENDAHULUAN

Dalam kondisi sehari-hari sering ditemukan adanya hubungan antara variabel satu dengan variabel lainnya. Sebagai contoh: dalam bidang pemasaran diketahui adanya hubungan antara volume penjualan dengan biaya advertensi, bidang perternakan diketahui adanya hubungan antara jumlah pakan yang diberikan dengan berat badan ternak, dan sebagainya. Untuk mengetahui bentuk hubungan ini dapat digunakan analisis regresi dengan metode *Ordinary Least Square (OLS)* dan persyaratan statistik yang harus terpenuhi yang disebut uji asumsi klasik (Supramono dan Sugiarto, 1993). Apabila asumsi klasik ini tidak terpenuhi, khususnya terdapat penyimpangan multikolinearitas, maka akan berdampak pada interval estimasi yang cenderung lebih lebar dan nilai hitung statistik uji t akan kecil, sehingga membuat variabel bebas secara statistik tidak signifikan mempengaruhi variabel terikat meskipun koefisien determinasi masih relatif besar.

Regresi gulud diajukan sebagai suatu cara untuk mengatasi penyimpangan kekolinearanganda (multikolinearitas). Metode regresi gulud diperoleh dengan cara yang sama seperti metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat sisaan. Regresi gulud menambahkan kendala (tetapan bias) pada kuadrat terkecil sehingga koefisien menyusut mendekati nol (Hastie *et al.* 2008). Selain itu, suatu prosedur alternatif lain yang menganalisis struktur korelasi dengan metode komponen utama secara terperinci pertama kali diajukan oleh Harold Hoteling dalam makalah klasiknya *"Analysis of a complex of statistical variables into principal components", Journal of Education Psychology* (Draper dan Smith, 1992). Metode regresi komponen utama akan didapatkan variabel bebas baru yang tidak berkorelasi dan mampu menyerap informasi yang terkandung dalam variabel asli atau memberikan kontribusi terhadap varian seluruh variabel (Siti, 2010).

Berdasarkan penjabaran di atas, maka peneliti tertarik untuk melakukan penelitian dengan tujuan untuk membandingkan metode regresi gulud dan regresi komponen utama dalam mengatasi penyimpangan multikolinearitas pada analisis regresi linear berganda. Penelitian dianggap menarik karena mampu memberikan solusi terbaik dan paling efektif dalam mengatasi penyimpangan multikolinearitas sehingga dapat memberikan permodelan regresi dengan nilai *RMSE* lebih kecil dan nilai R^2_{Adj} relatif besar.

2. METODE PENELITIAN

Dalam penelitian ini terdapat empat data simulasi dengan jumlah variabel (bebas dan terikat) kurang dari atau sama dengan 10 ($p \leq 10$) Persyaratan statistika berupa asumsi klasik pada masing-masing data telah terpenuhi, kecuali asumsi multikolinearitas. Metode yang digunakan adalah regresi gulud *(ridge regression)* dan regresi komponen utama *(principal component regression).*

Regresi gulud bertujuan untuk mengatasi multikolinearitas yang terdapat dalam liniar berganda yang mengakibatkan matriks $X'X$ hampir singular yang menyebabkan nilai estimasi parameter yang tidak stabil. Nilai estimasi parameter untuk regresi gulud dihitung dengan rumus (Draper & Smith, 1992):

$$\hat{\beta}_i(k) = (Z'Z + kI)^{-1}Z'Y^R, i = 1,2,3, \dots, p - 1 \quad (3.1)$$

dengan: Z dan Y^R adalah matriks yang telah terpusat dan terskalakan, *k* adalah tetapan bias, dan *I* adalah matriks identitas.

Suatu acuan yang digunakan untuk memilih tetapan bias *(k)* dengan melihat nilai VIF pada $\hat{\beta}_i(k)$ dan kecenderungan *Ridge Trace*. Umumnya sifat dari penaksiran regresi gulud ini memiliki variansi yang minimum sehingga diperoleh nilai VIF untuk $\hat{\beta}_i(k)$ yang merupakan diagonal utama dari matriks dari

$$(Z'Z + kI)^{-1}Z'Z(Z'Z + kI)^{-1} \quad (3.2)$$

Pemilihan nilai *k* merupakan masalah yang perlu diperhatikan. Tetapan bias yang diinginkan adalah tetapan bias yang relatif kecil dan menghasilkan koefisien estimator yang relatif stabil. Pada umumnya nilai k terletak pada interval $0 < k < 1$ (Pradipta, 2009).

Setelah mendapatkan nilai koefisien estimator $\hat{\beta}_i(k)$, maka dapat terbentuk permodelan regresi gulud. Permodelan ini digunakan untuk uji keberarti regresi dengan hipotesis nol adalah $\hat{\beta}_1(k) = \hat{\beta}_2(k) = \cdots = \hat{\beta}_i(k) = 0$ atau variabel bebas secara simultan tidak signifikan di dalam model (regresi tidak berarti), dan uji keberarti koefisien dengan hipotesisi nol adalah $\hat{\beta}_i(k) = 0, i = 1,2, \dots, p-1$ atau koefisien regresi tidak signifikan (Pradipta, 2009).

Untuk kepentingan estimasi, maka model regresi gulud dapat ditransformasi ke bentuk variabel asal sehingga menjadi model regresi berganda dengan rumus(Pradipta, 2009):

$$\hat{\beta}_i = \left(\frac{s_Y}{s_{Xi}}\right) \hat{\beta}_i(k), i = 1,2, \dots, p-1 \tag{3.3}$$

Dari persamaan (3.3), maka dapat dihitung nilai $\hat{\beta}_0$ dengan rumus

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2 - \cdots - \hat{\beta}_{P-1} X_{P-1} \tag{3.4}$$

Adapun langkah-langkah untuk mengatasi penyimpangan multikolinearitas dengan metode regresi gulud secara singkat dapat dijabarkan sebagai berikut:

1. Transformasi data melalui *centering* dan *rescaling.*
2. Penentuan nilai k (tetapan bias) melalui metode *ridge trace* dengan memperhatikan nilai VIF $\hat{\beta}_i(k)$ dan $\hat{\beta}_i(k)$.
3. Persamaan model regresi gulud.
4. Uji keberartian regresi (ANOVA) dan uji keberartian koefisien pada model regresi gulud.
5. Transformasi ke bentuk awal sehingga menghasilkan model regresi linear berganda.

Sedangkan regresi komponen utama dilakukan dengan cara mereduksi variabel-variabel bebas yang ada menjadi beberapa variabel baru yang saling bebas dan saling kombinasi linear dari variabel bebas asal. Prinsip dari regresi komponen utama adalah meregresikan nilai-nilai komponen utama (W_j) yang terpilih dengan variabel terikat *(Y).* Nilai-nilai komponen utama diperoleh melalui matriks X yang telah terpusat dan terskalakan yang disebut matriks *Z.* Matriks korelasi dari $Z'Z$ merupakan solusi dari nilai-nilai eigen $(\lambda_1, \lambda_2, \dots, \lambda_j)$ dari persamaan determinan (Draper & Smith, 1992):

$$|Z'Z - \lambda I| = 0 \tag{3.5}$$

Untuk setiap nilai-nilai eigen λ_j terdapat vektor ciri *(characteristic vector)* γ_j yang memenuhi sistem persamaan homogen (Draper & Smith, 1992):

$$(Z'Z - \lambda_j I)\gamma_j = 0 \quad (3.6)$$

Vektor γ_j ini digunakan untuk mengucapkan kembali Z ke dalam suku-suku komponen utama *W* dalam bentuk (Draper & Smith, 1992):

$$W_{j=}\gamma_1 Z_1 + \gamma_2 Z_2 + \cdots + \gamma_j Z_j \quad (3.7)$$

Prosedur ini menciptakan variabel-variabel baru W_j dari variabel-variabel asanya Z_j melalui persamaan (3.5) sehingga vektor-vektor W itu ortogonal sesamanya. Variabel W_j padanan nilai λ_j yang terbesar disebut komponen utama pertama. Komponen ini menjelaskan bagian terbesar dari keragaman yang dikandung oleh gugusan data yang telah dibakukan (Draper & Smith, 1992).

Berikut ini adalah langkah-langkah untuk mengatasi multikolinearitas dengan metode regresi komponen utama:

1. Data telah terpusat dan terskalakan (*centering* dan *rescaling*).
2. Penentuan nilai eigen *(eigen value).*
3. Penentuan nilai-nilai komponen utama W_j sebagai variabel baru.
4. Persamaan model regresi komponen utama W_j.
5. Uji hipotesis ANOVA dan parsial untuk variabel Y dan W_j.
6. Transformasi ke bentuk variabel asal sehingga menghasilkan regresi linear berganda.

Dari kedua metode tersebut dilakukan perbandingan dengan menggunakan nilai nilai *RMSE (Root Mean Square Error)* dan *Adjusted R Square* (R^2_{Adj}).

3. HASIL PENELITIAN DAN PEMBAHASAN

Ada empat data simulasi yang digunakan dalam penelitian ini yang disebut dengan data ke-1, data ke-2, data ke-3, dan data ke-4. Berdasarkan pada hasil transformasi data melalui *centering* dan *rescaling*, maka nilai *k* untuk masing-masing data dengan metode *ridge trace* sebagai berikut:

Tabel 3.1 Tetapan Bias *(k)*

Data	Nilai *k*
Data ke-1	0.002493
Data ke-2	0.000120
Data ke-3	0.001940
Data ke-4	0.002700

Pemilihan nilai *k* ini berdasarkan pertimbangan pada nilai VIF yang menunjukkan bahwa variabel bebas tidak saling berkorelasi dengan variabel-

variabel bebas yang lainnya, serta menghasilkan koefisien estimator yang relatif stabil dengan menggunakan persamaan (3.1). Dari koefisien estimator tersebut akan terbentuk persamaan regresi gulud untuk masing-masing data sebagai berikut:

Data ke-1:

$$\hat{Y}^R = 0.3772Z_1 + 0.0376Z_2 + 0.3651Z_3 + 0.2202Z_4$$

Data ke-2:

$$\hat{Y}^R = -0.5603Z_1 + 1.5815Z_2 + 0.6063Z_3 - 0.6342Z_4 - 0.0020Z_5 + -0.0003Z_6$$

Data ke-3:

$$\hat{Y}^R = 0.1454Z_1 + 0.3326Z_2 + 0.6472Z_3 - 0.0438Z_4 - 0.4779Z_5$$

Data ke-4:

$$\hat{Y}^R = 0.0098Z_1 - 0.0009Z_2 + 0.0101Z_3 + 0.2707Z_4 - 0.0046Z_5 + 0.7139Z_6$$

Dari persamaan model regresi tersebut dilakukan pengujian keberartian regresi dan koefisien dengan tingkat signifikansi sebesar 5% sebagai berikut:

Tabel 3.2 Ringkasan ANOVA Regresi Gulud

Data	F_{hitung}	F_{tabel}	Kesimpulan
Data ke-1	1205.0296	5.19	Signifikan
Data ke-2	7442.5802	2.53	Signifikan
Data ke-3	9.1766	3.20	Signifikan
Data ke-4	1805.7271	2.53	Signifikan

Berdasarkan pada tabel 3.2 dapat diketahui bahwa kesimpulan tolak H_0 untuk semua data, artinya variabel bebas untuk data ke-1 sampai data ke-4 secara simultan signifikan di dalam model. Setelah melakukan ANOVA regresi gulud, maka dilakukan penguji keberartian koefisien yang menghasilkan kesimpulan tolak H_0 untuk semua data, artinya variabel bebas secara individu berpengaruh terhadap nilai taksiran variabel terikat. Oleh karena itu didapatkan permodelan regresi berganda untuk masing-masing data sebagai berikut:

Data ke-1:

$$\hat{Y} = -2.34245 + 0.0947826X_1 + 0.978327X_2 + 0.000151805X_3 + +0.000185490X_4 + \varepsilon$$

Data ke-2:

$$\hat{Y} = -5100.97 - 0.5139517X_1 + 6.100657X_2 + 1.610987X_3 - 252.3055X_4 + \\ -0.000518967X_5 - 0.002554177X_6 + \varepsilon$$

Data ke-3:

$$\hat{Y} = 1665.11 + 1.00979X_1 + 0.0809X_2 + 1.140657X_3 - 0.00034783X_4 + \\ +0.0003478336X_5 + \varepsilon$$

Data ke-4:

$$\hat{Y} = -2.289672 + 0.896X_1 - 0.00823X_2 + 0.0008182615X_3 + \\ -0.01868802X_4 + 0.0000092X_5 + 0.02388855X_6 + \varepsilon$$

Untuk mengatasi multikolinearitas dengan metode regresi komponen utama, diperoleh bahwa data ke-1, data ke-2, dan data ke-4 menghasilkan satu komponen utama (W_1) dan data ke-3 menghasilkan dua komponen utama (W_1dan W_2) dengan nilai eigen sebagai berikut:

Tabel 3.3 Nilai Eigen

Data	Komponen Utama	Nilai Eigen λ_i
Data ke-1	W_1	3.964939
Data ke-2	W_1	5.102998
Data ke-3	W_1	2.998028
	W_2	1.970221
Data ke-4	W_1	4.744841

Dari komponen-komponen utama tersebut diperoleh persamaan regresi komponen utama sebagai berikut:

Data ke-1:

$$\hat{Y} = 27.908 + 9.316W_1 + \varepsilon \text{ dengan } W_1 = 0.2516Z_1 + 0.2487Z_2 + 0.2517Z_3 + \\ +0.2513Z_4$$

Data ke-2:

$$\hat{Y} = 95.771 + 17.317W_1 + \varepsilon \text{ dengan } W_1 = 0.1907Z_1 + 0.1912Z_2 + \\ +0.1903Z_3 + 0.1899Z_4 - 0.1758Z_5 + 0.1210Z_6$$

Data ke-3:

$$\hat{Y} = 1530.833 + 640.466W_1 - 180.806W_2 + \varepsilon \text{ dengan } W_1 = 0.0620Z_1 + \\ +0.2746Z_2 + +0.2742Z_3 + 0.2682Z_4 + 0.0725Z_5 \text{ dan } W_1 = -0.1520Z_1 + \\ +0.2946Z_2 + 0.3058Z_3 + +0.3221Z_4 - 0.1394Z_5$$

Data ke-4:

$$\hat{Y} = 95.771 + 17.358W_1 + \ \varepsilon \text{ dengan } W_1 = 0.2049Z_1 + 0.0441Z_2 + \\ +0.1914Z_3 + 0.2065Z_4 + 0.2011Z_5 + 0.2061Z_6$$

Setelah mendapatkan regresi komponen utama, maka dilakukan analisis varian dengan tingkat signifikansi sebesar 5% sebagai berikut:

Tabel 3.4 Ringkasan ANOVA untuk variabel Y dan W_j

Data	F_{hitung}	F_{tabel}	Kesimpulan
Data ke-1	515.4863	5.32	Signifikan
Data ke-2	139.8810	4.20	Signifikan
Data ke-3	8.9231	4.10	Signifikan
Data ke-4	169.4049	4.20	Signifikan

Berdasarkan pada tabel 3.4 diketahui bahwa kesimpulan yang dihasilkan adalah tolak H_0, sehingga dapat dikatakan bahwa variabel-variabel komponen utama memiliki pengaruh secara signifikan terhadap nilai taksiran Y.

Dari permodelan regresi komponen utama, maka ditransformasi kembali ke dalam bentuk variabel asal sehingga diperoleh persamaan model regresi berganda untuk masing-masing data.

Data ke-1:

$$\hat{Y} = -14.53393 + 0.0632146X_1 + 6.474607X_2 + 0.0001046466X_3 + \\ +0.0002116091X_4 + \varepsilon$$

Data ke-2:

$$\hat{Y} = -1220.382 + 0.1749063X_1 + 0.737399X_2 + 0.505658X_3 + 75.5467X_4 + \\ -0.0446974X_5 + 1.19621X_6 + \varepsilon$$

Data ke-3:

$$\hat{Y} = 1779.464 - 1.055559X_1 + 0.0716632X_2 + 0.5389261X_3 + \\ +0.0025604X_4 - \ 4.006648X_5 + \varepsilon$$

Data ke-4:

$$\hat{Y} = -147.723 + 18.7787X_1 + 0.40614X_2 + 0.01556783X_3 + \\ +0.01425356X_4 + 0.00406703X_5 + 0.006894657X_6 + \varepsilon$$

Untuk mengetahui metode terbaik dan paling efektif dalam mengatasi multikolinearitas, maka dilakukan perbandingan dengan nilai *RMSE* dan R^2_{Adj} seperti di bawah ini:

Tabel 3.5 Perbandingan Nilai *RMSE* dan R^2_{Adj}

Data	*RMSE*		R^2_{Adj} (%)	
	R.Gulud	R.Kompenen Utama	R.Gulud	R.Kompenen Utama
Data ke-1	0.4028152	0.6154536	99.81	99,56
Data ke-2	0.4471945	3.218993	99.93	96.63
Data ke-3	326.3892	330.4505	78.80	78.27
Data ke-4	0.9071593	2.931894	99.73	97.21

Berdasarkan pada tabel 3.5 diketahui bahwa regresi gulud memiliki nilai *RMSE* lebih kecil dan R^2_{Adj} lebih besar dibandingkan dengan regresi komponen utama, sehingga dapat dikatakan bahwa metode regresi gulud merupakan metode terbaik dan paling efektif dalam mengatasi penyimpangan multikolinearitas pada analisis regresi berganda.

4. KESIMPULAN

Dari penelitian yang telah dilakukan dengan empat data simulasi, maka dapat disimpulkan bahwa regresi gulud merupakan metode terbaik dan paling efektif dalam mengatasi penyimpangan multikolinearitas. Hal ini dilihat pada nilai *RMSE* yang cenderung lebih kecil dan nilai R^2_{Adj} lebih besar dibandingkan dengan regresi komponen utama, sehingga permodelan regresi berganda untuk masing-masing data sebagai berikut:

Data ke-1:

$$\hat{Y} = -2.34245 + 0.0947826X_1 + 0.978327X_2 + 0.000151805X_3 + \\ +0.000185490X_4 + \varepsilon$$

Data ke-2:

$$\hat{Y} = -5100.97 - 0.5139517X_1 + 6.100657X_2 + 1.610987X_3 - 252.3055X_4 + \\ -0.000518967X_5 - 0.002554177X_6 + \varepsilon$$

Data ke-3:

$$\hat{Y} = 1665.11 + 1.00979X_1 + 0.0809X_2 + 1.140657X_3 - 0.000347834X_4 + \\ +0.0003478336X_5 + \varepsilon$$

Data ke-4:

$$\hat{Y} = -2.289672 + 0.896X_1 - 0.00823X_2 + 0.0008182615X_3 + \\ -0.01868802X_4 + 0.0000092X_5 + 0.02388855X_6 + \ \varepsilon$$

5. DAFTAR PUSTAKA

Draper, N. Dan Smith, H. 1992. *Analisis Regresi Terapan. Edisi Kedua.* Jakarta: Gramedia Pustaka Utama.

Hastie, T., *el al*. 2008. *The Elements of Statistical Learning. Data Mining, Inference, and prediction. Edisi Kedua.* New York: Spring.

Pradipta, N. 2009. *Metode Regresi Ridge untuk mengatasi model regresi linear berganda yang mengandung multikolinearitas. Skripsi.* Sumatera Utara: Universitas Sumatera Utara.

Siti, A. 2010. *Principal Component Analysis dalam Menentukan Faktor yang Berpengaruh terhadap Nilai Oksigen Terlarut.* Prosiding Seminar Nasional Limnologi V. pp. 723-730.

Supramono dan Sugiarto. 1993. *Statistika.*Yogyakarta: Andi Offset.