

Implementasi Ridge Regression untuk Mengatasi Gejala Multikolinearitas pada Pemodelan Curah Hujan Berbasis Data *Time Series* Klimatologi

Risa Arisandi¹, Dadang Ruhiat², dan Emas Marlina³

^{1,2,3} Program Studi Matematika, FMIPA, Universitas Bale Bandung, Bandung, Indonesia

risaarisandi5@gmail.com

Abstrak. Analisis regresi merupakan metode analisis yang digunakan untuk menganalisis data dan mengambil kesimpulan tentang hubungan ketergantungan variabel terikat terhadap variabel bebas. Dalam analisis Regresi Linear Berganda seringkali didapati bahwa data yang diolah mengandung gejala multikolinieritas, yaitu memiliki korelasi yang tinggi antar variabel bebasnya yang ditandai dengan nilai VIF lebih dari 10. Gejala multikolinearitas akan mengganggu hasil pemodelan regresi linier berganda, diantaranya adalah menyebabkan perubahan nilai koefisien yang berarti apabila dilakukan penambahan atau pengurangan satu variabel bebas. dan nilai *standard error* untuk koefisien regresi menjadi lebih besar dari yang sebenarnya (*overestimated*). Gejala multikolinearitas dapat diatasi melalui metode *Ridge Regression*. Penelitian ini dilakukan bertujuan mengatasi gejala multikolinearitas. Perbandingan hasil prediksi model Regresi Linear Berganda dengan hasil model *Ridge Regression* dilakukan sehingga diperoleh model terbaik. Hasil analisis menunjukan model Regresi Linear Berganda dan model *Ridge Regression* memiliki nilai RMSE yang relatif sama yaitu sebesar 4,26 dan 4,30, sedangkan berdasarkan nilai MAPE *Ridge Regression* memiliki nilai sedikit lebih baik yaitu sebesar 28,75 % sedangkan model Regresi Linear Berganda sebesar 28,82 %. Berdasarkan nilai dari ukuran kebaikan model tersebut dapat disimpulkan bahwa model *Ridge Regression* memberikan hasil prediksi yang lebih baik dari model Regresi Linier Berganda..

Kata kunci: Analisis Regresi Linear Berganda, *Ridge Regression*, Multikolinearitas.

1. Pendahuluan

Penulis Analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel terhadap variabel lainnya [3]. Regresi adalah salah satu perangkat dasar untuk analisis yang bisa digunakan untuk membuat model prediktif untuk berbagai jenis data [11]. Terdapat berbagai macam metode regresi, diantaranya Regresi Linear Berganda dan Regresi *Ridge*.

Metode *Ridge Regression* pertama kali dikemukakan oleh A.E Hoerl pada tahun 1962[8]. *Ridge Regression* (Regresi *Ridge*) merupakan metode yang dikembangkan untuk menstabilkan nilai koefisien regresi karena adanya Multikolinearitas. Metode ini merupakan modifikasi dari metode kuadrat terkecil dengan cara menambahkan tetapan bias c yang kecil pada diagonal matriks $X^T X$. Walaupun memiliki penduga yang bias, penduga ini memiliki variansi estimator koefisien minimum [12]. Regresi *Ridge* mengurangi dampak Multikolinearitas dengan menentukan penduga yang bias tetapi mempunyai variansi yang lebih kecil dari variansi penduga Regresi Linear Berganda [12].

Multikolinearitas yaitu kondisi dimana terjadi korelasi yang cukup tinggi antara variabel-variabel bebas[2]. Korelasi yang sangat tinggi akan menghasilkan penaksiran yang berbias, tidak stabil dan mungkin jauh dari nilai sasaran. Menurut Montgomery dan Hines (1990) dalam [13], Multikolinieritas dapat menyebabkan koefisien Regresi yang dihasilkan oleh Analisis Regresi Berganda menjadi sangat lemah atau tidak dapat menghasilkan hasil analisis yang mewakili sifat atau pengaruh dari variabel bebas yang bersangkutan. Menurut Montgomery dan Peck [10] multikolinearitas disebabkan oleh beberapa faktor, yakni metode pengumpulan data yang digunakan, batasan yang ada pada model atau populasi yang diambil sampelnya, spesifikasi model dan model yang “*overdetermined*”. Gejala multikolinearitas kemungkinan besar terjadi pada data *time series* seperti data klimatologi.

Klimatologi merupakan ilmu yang mempelajari tentang iklim dalam kaitannya dengan kondisi cuaca di suatu wilayah yang dirata-ratakan selama periode yang panjang. Unsur-unsur iklim terdiri

dari radiasi sinar matahari, temperatur (suhu), kelembaban, curah hujan dan angin [5]. Curah Hujan merupakan parameter yang tingkat variabilitasnya tinggi baik terhadap lokasi maupun waktu yang mencakup variasi harian, bulanan, musiman dan tahunan [9]. Dikarenakan variabilitas yang tinggi, data curah hujan diperkirakan memiliki gejala multikolinearitas dan autokorelasi untuk itu untuk melakukan pemodelan pada data ini diperlukan analisis regresi yang dapat mengatasi masalah tersebut, salah satu cara untuk mengatasi multikolinearitas yaitu dengan regresi *ridge* [4].

Penulisan penelitian ini berdasarkan dari beberapa penelitian dengan tema yang serupa yang telah dilakukan diantaranya adalah penelitian Maulana Afham, Manfaati Nur dan Tiani Wahyu Utami (2017) dalam jurnalnya yang berjudul “Pemodelan Regresi Ridge Pada Kasus Curah Hujan di Kota Semarang”. Dari paparan di atas peneliti termotivasi untuk melakukan penelitian dengan judul “Implementasi *Ridge Regression* untuk Mengatasi Gejala Multikolinearitas pada Pemodelan Curah Hujan Berbasis Data *Time Series* Klimatologi” dengan variabel terikat curah hujan dan variabel bebas berupa temperatur udara, temperatur air, penguapan dan kelembaban pada data DAS Cimanuk Jawa Barat. Adapun tujuan dari penelitian ini adalah terbangun model *Ridge Regression* dan model Regresi Linear Berganda untuk data *time series* klimatologi, serta diperoleh model terbaik dari kedua metode tersebut.

2. Metode Penelitian

1.1. Lokasi penelitian

Penelitian dilakukan di Kawasan DAS Cimanuk bagian hulu, yang termasuk ke pengelolaan Wilayah Sungai (WS) Cimanuk-Cisanggarung Jawa Barat. DAS Cimanuk dalam garis besar terbagi menjadi tiga bagian (Sumber: Dinas PSDA Jawa Barat), yaitu:

- Sub-DAS Cimanuk bagian hulu, termasuk ke wilayah administratif Kab. Garut dan Kab Sumedang
 - Sub-DAS Cimanuk bagian tengah, termasuk ke wilayah administratif Kab. Sumedang dan Kab. Majalengka,
 - Sub-DAS Cimanuk bagian hilir, termasuk ke wilayah administratif Kabupaten Indramayu
- Luas DAS Cimanuk adalah 3584 Km², sedangkan luas sub-DAS Cimanuk bagian hulu yang menjadi lokasi penelitian adalah 1404,74 Km²

1.2. Data dan sumber data

Data yang digunakan dalam penelitian ini merupakan data sekunder klimatologi hasil pencatatan Stasiun Klimatologi Sukatali dari januari tahun 2015 sampai dengan Desember 2016. Data klimatologi terdiri atas data radiasi sinar matahari, temperatur udara, kelembaban udara, penguapan, kecepatan angin dan curah hujan. Data tersebut diperoleh dari Balai Besar Wilayah Sungai (BBWS) Cimanuk dan Direktorat Bina Teknik Sumber Daya Air Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR). Karena masalah ketersediaan data variabel-variabel yang dilibatkan dalam pemodelan regresi adalah Temperatur udara (X_1), Temperatur air dalam PAN (X_2) kelembaban udara (X_3) dan penguapan (X_4) sebagai variabel bebas dan curah hujan sebagai variabel terikat (Y).

1.3. Metode

Metode yang digunakan untuk melakukan analisis data dalam penelitian ini adalah metode *Ridge Regression* dan Regresi Linear Berganda.

Persamaan umum model Regresi Linier Berganda untuk populasi dapat ditunjukkan sebagai berikut (Supranto, J. 2004) :

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon \quad (1)$$

dimana:

Y = Variabel tak bebas

β_0 = *intercept*

$\beta_1, \beta_2, \dots, \beta_k$ = Koefisien regresi untuk $X_1, X_2, \dots, X_k$

ε = *Error* (Kekeliruan)

1.4. *Penaksiran parameter regresi*

Metode yang Persamaan (1) jika dibuat dalam bentuk matriks akan menjadi

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2)$$

untuk:

Y = vektor $n \times 1$

X = matrik $n \times (k+1)$

B = vektor $n \times 1$

ε = vektor $n \times 1$

Penyajian matrik regresi linier dengan k variabel adalah:

$$Y = X\beta + \varepsilon \quad (3)$$

Jika ruas kiri dan kanan dikalikan dengan transpose matrik X , sehingga diperoleh

$$\begin{aligned} X'Y &= X'X\beta \\ \widehat{\beta} &= (X'X)^{-1}X'Y \end{aligned} \quad (4)$$

Jika dinyatakan dalam bentuk matrix:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \cdots & \sum X_{ki} \\ \sum X_{ki} & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \cdots & \sum X_{1i} X_{ki} \\ \sum X_{2i} & \sum X_{2i} X_{1i} & \sum X_{2i}^2 & \cdots & \sum X_{2i} X_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ki} & \sum X_{ki} X_{1i} & \sum X_{ki} X_{2i} & \cdots & \sum X_{ki}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum X_{1i} Y_i \\ \sum X_{2i} Y_i \\ \vdots \\ \sum X_{ki} Y_i \end{bmatrix} \quad (5)$$

Penaksir Regresi Ridge, Apabila terjadi multikolinieritas matrik $X'X$ ditransformasikan menjadi matrik korelasi $Z'Z$, sehingga dugaan koefisien regresi menjadi:

$$b_z(\theta) = (Z'Z + \theta I_k)^{-1}(Z'Y) \quad (6)$$

dengan

$b_z(\theta)$ = Estimator *Ridge Regression*

θ = Parameter Regression (bilangan kecil positif antara 0 s.d 1)

Z = Matrik korelasi yang merupakan hasil transformasi *variable Regressor* melalui

metode *centering* dan *rescaling*

sehingga dugaan variabel *Regressand* menjadi

$$\hat{Y}(\theta) = Zb_z(\theta) \quad (7)$$

Proses di atas disebut *Ridge Regression*. Pemilihan θ diserahkan kepada analisis. analisis dapat mencobakan nilai θ sampai keadaan stabil. Salah satu metode yang bisa digunakan yaitu dengan mencari nilai statistic *Cp Mallows* ($C\theta$) dengan rumus:

$$C\theta = \frac{SS_{res,k}}{\hat{\sigma}^2} - n + 2 + 2tr(H_\theta) \quad (8)$$

dengan

$$tr(H_\theta) = \sum_{i=1}^k \frac{\lambda_i}{\lambda_i + \theta}$$

Keterangan:

$SS_{res,k}$ = Jumlah kuadrat residu dari persamaan *Ridge Regression*

n = banyak pengamatan

λ_i = Eigen value dari matrik $Z'Z + \theta I_k$

$tr(H_\theta)$ = Trace dari matrik H_θ

$\hat{\sigma}^2$ = Penaksir varian metode kuadrat terkecil

Setelah memperoleh nilai $C\theta$, nilai θ terpilih adalah nilai θ yang dapat meminimumkan nilai $C\theta$. Untuk memperoleh koefisien regresi dalam variabel asal digunakan rumus sebagai berikut :

$$b_i = \left(\frac{S_Y}{S_i} \right) b'_i \quad (9)$$

dengan

$i = 1, 2, 3$

S_Y = Galat baku dari data awal Y

S_i = Galat baku dari data awal X ke- i

b'_i = Koefisien regresi setelah melalui *Ridge Regression*

1.5. Tahapan analisis data

Tahapan analisis data dengan menggunakan metode regresi linier berganda dan metode *ridge regression* adalah sebagai berikut:

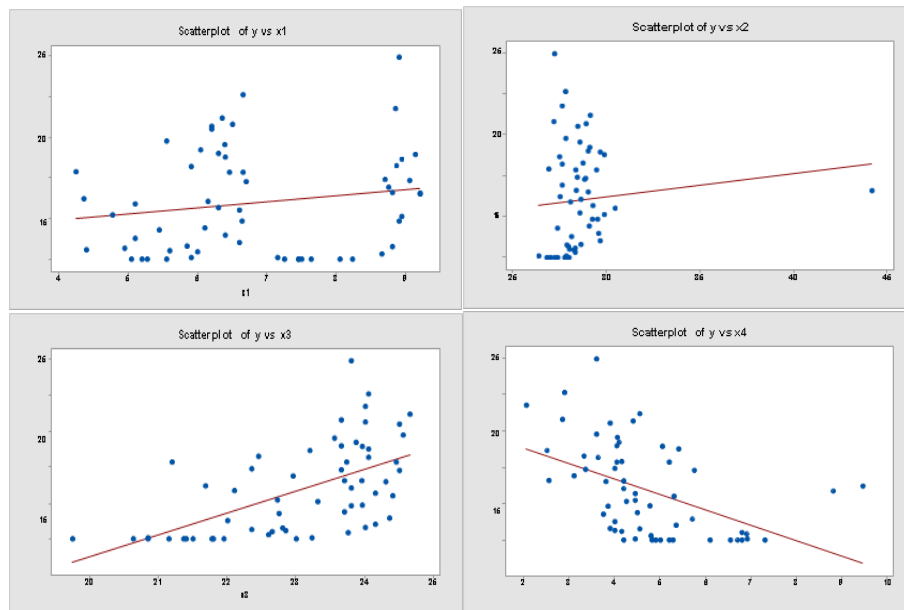
1. Menyiapkan data dan cek skala pengukuran.
2. Melakukan uji linearitas (uji plot data dan uji statistik)
3. Melakukan uji asumsi klasik, yaitu uji normalitas residu, uji autokorelasi, uji multikolieritas dan uji heteroskedastisitas
4. Melakukan uji Regresi Linear Berganda dan pemilihan model Regresi Linear Berganda terbaik.
5. Melakukan uji asumsi *Regresi Ridge*.
6. Transformasi data melalui *centering* dan *scaling* untuk mendapatkan persamaan model *Regresi Ridge*.
7. Mentransformasikan ke bentuk awal sehingga menghasilkan model Regresi Linear Berganda.
8. Membandingkan model dari Regresi Linear Berganda dengan model dari Regresi *Ridge* untuk mencari model terbaik dengan plotting data dan uji RMSE (*Root Mean Square Error*)

3. Hasil dan Pembahasan

1.6. Uji asumsi klasik

1.6.1. Uji linearitas

Pengujian gejala hubungan linier antara variabel bebas (X) dengan variabel terikat (Y) dilakukan secara visual melalui penggambaran *scatterplot* dan uji statistik. Grafik *Scatterplot* hasil analisis dengan bantuan software statistik Minitab disajikan pada Gambar 1. Variabel X_1 dan X_2 tidak memiliki pola hubungan linier dengan variabel Y , sedangkan variabel X_3 dan X_4 memiliki pola hubungan linier dengan variabel Y . Variabel X_3 dan variabel Y memiliki pola hubungan linier positif, sedangkan variabel X_4 dan variabel Y memiliki pola hubungan linier negatif.



Gambar 1. *Scatterplot* Variabel X_1 , X_2 , X_3 dan X_4 terhadap Y

Uji statistik linieritas memberikan hasil yang identik dengan *scatterplot*. Diketahui nilai sig. untuk X_1 dan X_2 lebih besar dari 0,05, sehingga kedua variabel tersebut tidak memiliki hubungan linier dengan Y , sedangkan nilai sig. untuk X_3 dan X_4 lebih kecil dari 0,05, sehingga kedua variabel tersebut diinterpretasikan memiliki hubungan linier dengan Y . Hasil uji statistik linieritas menggunakan SPSS disajikan pada Gambar 2. Dengan demikian hanya variabel X_3 , X_4 dan Y yang dilibatkan dalam pemodelan regresi linier berganda.

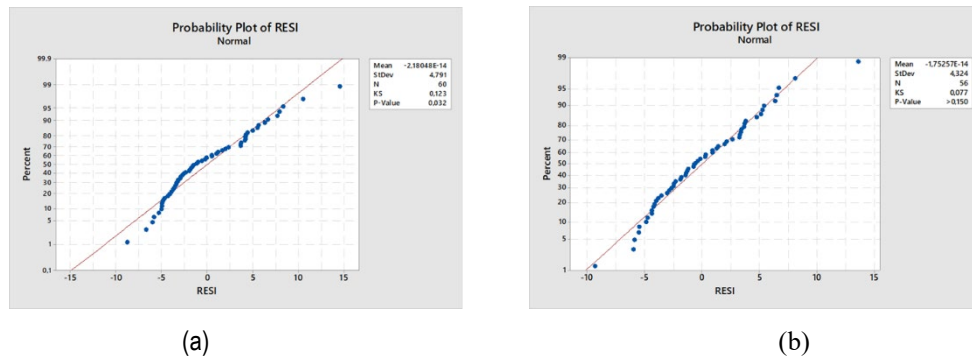
			Sum of Squares	df	Mean Square	F	Sig.
Y * X1	Between Groups	(Combined)	1823.633	47	38.801	.892	.634
		Linearity	71.410	1	71.410	1.641	.224
		Deviation from Linearity	1752.222	46	38.092	.876	.648
	Within Groups		522.069	12	43.506		
	Total		2345.701	59			
Y * X4	Between Groups	(Combined)	1854.176	46	40.308	1.066	.477
		Linearity	539.905	1	539.905	14.280	.002
		Deviation from Linearity	1314.271	45	29.206	.772	.748
	Within Groups		491.525	13	37.810		
	Total		2345.701	59			

Gambar 2. Hasil uji statistik linieritas menggunakan SPSS

1.6.2. Uji Normalitas Residual

Uji normalitas dilakukan dengan menggunakan uji statistik *Kolmogorov Smirnov* (K-S) yang proses penghitungannya menggunakan bantuan *software* Minitab. Diperoleh *Probability plot* peluang dan *p-value* < 0,05 peluang seperti disajikan pada Gambar 3(a). Dengan demikian disimpulkan tidak berdistribusi normal residu.

Untuk mengatasi permasalahan di atas maka dilakukan uji *outlier*, data yang terdeteksi sebagai data *outlier* dikeluarkan kemudian kembali dilakukan uji normalitas tanpa melibatkan data *outlier*. *Output* minitab diketahui nilai *p-Value* = 0,150 lebih besar dari 0,05. Setelah menghilangkan data *outlier* sebaran data residu menjadi normal. Lihat Gambar 3(b).



Gambar 3. Hasil *Probability Plot* dari Residu

1.6.3. Uji Autokorelasi

Deteksi gejala autokorelasi dilakukan menggunakan uji *Durbin – Watson*. Dengan jumlah variabel bebas (k) = 2, jumlah data (n) = 56 dan $\alpha=0,05$. Berdasarkan tabel *Durbin – Watson* diperoleh nilai $dL=1,495$ dan $dU=1,643$. Kemudian diperoleh nilai $DW = 1,6296$ yang berada di antara nilai dL dan dU atau $dL < DW < dU$. Sehingga tidak dapat diputuskan apakah mengandung gejala autokorelasi atau tidak.

1.6.4. Uji Multikolineritas

Nilai korelasi antar variabel bebas X_3X_4 rendah, dan diperoleh nilai $VIF = 1,06$ yang mana lebih kecil dari 10, maka, dapat disimpulkan tidak terdapat gejala multikolinearitas. Untuk analisis selanjutnya pemodelan *Ridge Regression* dan Regresi Linier Berganda tetap dilakukan. Hal ini dilakukan untuk melihat perbandingan hasil kedua metode tersebut dalam kondisi tidak mengandung gejala multikolineritas.

1.6.5. Uji Heteroskedastisitas

Uji heteroskedastisitas dilakukan dengan menggunakan uji statistika Gletjer dengan *Software* Minitab, seperti disajikan pada Tabel 1., diketahui nilai *P-Value* untuk $X_3 = 0,171$ dan *P-Value* untuk $X_4 = 0,286$ yang mana kedua nilai tersebut lebih besar dari $\alpha=0.05$, sehingga dapat disimpulkan tidak terjadi gejala heteroskedastisitas.

Tabel 2. Hasil Uji Heteroskedastisitas

Term	Coef	SE Coef	T-Value	P-Value
Constan				
t	-4,79	7,50	-0,64	0,526
X3	0,422	0,304	1,39	0,171
X4	-0,334	0,310	-1,08	0,286

Sumber : Hasil analisis, 2021

1.7. Penaksiran koefisien regresi

1.7.1. Regresi linier ganda

Berdasarkan hasil perhitungan nilai koefisien regresi tanpa data *outlier* dengan menggunakan *software* Minitab seperti disajikan pada Tabel 3, maka diperoleh persamaan regresi linier berganda:

$$\hat{Y} = -44,2 + 2,728X_1 - 2,644X_2 + e$$

Besar nilai koefisien korelasi linier berganda sebesar 0,7363, artinya hubungan variabel X_1 dan X_2 secara bersama-sama terhadap Y adalah 0,7363. Sedangkan nilai koefisien determinasi adalah sebesar 0,5421, yang menunjukkan bahwa variasi yang terjadi pada Y akibat hubungan linier X_1 dan X_2 secara simultan terhadap Y adalah sebesar 54,21 persen. Keberartian koefisien regresi dihitung dengan menggunakan uji statisti t untuk uji keberartian parsial dan uji statistik F untuk uji keberartian secara simultan.

Tabel 3. Hasil Uji Signifikansi Koefisien menggunakan Uji t

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-44,2	13,40	-3,29	0,002	
X3	2,728	0,544	5,01	0,000	1,06
X4	-2,644	0,556	-4,75	0,000	1,06

Sumber: hasil analisis, 2021

Hasil perhitungan uji t dengan menggunakan *software* minitab diketahui nilai P -Value untuk X_3 dan X_4 adalah 0,000, yang mana kedua nilai tersebut lebih kecil dari $\alpha = 0,05$, lihat Tabel 3. Sedangkan untuk uji keberartian dengan menggunakan Uji F disajikan pada Tabel 4.

Tabel 4. Hasil Uji Signifikansi Koefisien menggunakan Uji F

	Sum Of Square	Df	Mean Square	F
Regression	1217,58	2	608,79	31,4
Residual	1028,32	53	19,4	
Total	2245,9			

Sumber: hasil analisis, 2021

dengan jumlah variabel bebas (k) = 2, derajat kebebasan 1 (df_1) = 2, derajat kebebasan 2 (df_2) = 53 dan $\alpha = 0,05$, maka berdasarkan tabel pada lampiran 1, diperoleh nilai $F_{\text{tabel}} = 3,172$. Sedangkan berdasarkan tabel di atas diketahui nilai $F_{\text{hitung}} = 31,38$, atau lebih besar dari F_{tabel} . Dengan demikian seluruh variabel bebas, X_1 dan X_2 , secara simultan berpengaruh signifikan terhadap variabel Y .

1.7.2. Ridge regression

1. Centering and scaling data

Untuk melakukan Analisis *Ridge Regression* perlu dilakukan transformasi data. data hasil *Centering and Scaling* dapat dilihat pada Tabel 5.

Tabel 5. Data Hasil Centering dan Scaling

No.	Y*	X3*	X4*
1	-0,05338	0,075972	-0,07504
2	0,108041	0,003985	-0,24054
3	0,091161	-0,086	-0,14002
4	0,237813	0,099967	-0,29571

Tabel 5. (Lanjutan)

No.	Y*	X3*	X4*
5	0,020472	0,063974	-0,23441
6	-0,04072	0,017183	-0,02601
7	0,034188	-0,02601	-0,16699
...	...	...	...
56	-0,0365	-0,0524	-0,00149

Sumber, Hasil analisis, 2021

2. Ordinary Least Square (OLS)

Data hasil *Centering and Scaling* kemudian digunakan untuk analisis menggunakan *OLS*, Sehingga diperoleh,

$$B_0 = 3,709.10^{-16}$$

$$B_3 = 0,479827$$

$$B_4 = -0,45550023$$

Berdasarkan hasil perhitungan selanjutnya diperoleh nilai $B'B = 0,43728$ dan nilai $MSE = 0,008639$.

3. Tetapan bias C

Tetapan bias C merupakan hal yang penting untuk analisis *Rigde Regression*, hasil perhitungan diperoleh nilai tetapan $Cc = 0,00395124$. Sehingga diperoleh matik $C.I$ berikut:

$$C.I = \begin{bmatrix} 0.00395124 & 0 & 0 \\ 0 & 0.00395124 & 0 \\ 0 & 0 & 0.00395124 \end{bmatrix} \quad (10)$$

4. Persamaan Ridge regression

Melalui perhitungan dengan menggunakan persamaan (6), (7), (8) dan (9), diperoleh matrik:

$$Z'Z = \begin{bmatrix} n & \sum X_3^* & \sum X_4^* \\ \sum X_3^* & \sum X_3^{*2} & \sum X_3^* X_4^* \\ \sum X_4^* & \sum X_4^* X_3^* & \sum X_4^{*2} \end{bmatrix} = \begin{bmatrix} 56 & 3,67 \cdot 10^{-14} & 4,6707 \cdot 10^{-16} \\ 3,67 \cdot 10^{-14} & 1 & 0,24013 \\ 4,6707 \cdot 10^{-16} & 0,24013 & 1 \end{bmatrix}$$

$$Z'Y = \begin{bmatrix} \sum Y^* \\ \sum X_3^* Y^* \\ \sum X_4^* Y^* \end{bmatrix} = \begin{bmatrix} -2,9490 \cdot 10^{-15} \\ 0,58909 \\ 0,57024 \end{bmatrix}$$

Sehingga diperoleh matrik:

$$(Z'Z + CI)^{-1}Z'Y = \begin{bmatrix} \beta_{r0} \\ \beta_{r3} \\ \beta_{r4} \end{bmatrix} = \begin{bmatrix} -3,607 \cdot 10^{-16} \\ 0,464780546 \\ 0,441203074 \end{bmatrix} \quad (11)$$

Sehingga diperoleh persamaan *Ridge regression*:

$$Y^* = 0,464780546 Z_3 - 0,441203074 Z_4$$

Diketahui nilai *Standar Error* (SE) Mean: $Y = 0,854$; $X_3 = 0,150$ dan $X_4 = 0,147$. Sehingga hasil transformasi koefisien *Ridge regression* dengan menggunakan persamaan (7) diperoleh:

$$b_3 = 2,6462$$

$b_4 = -2,5632$, dan

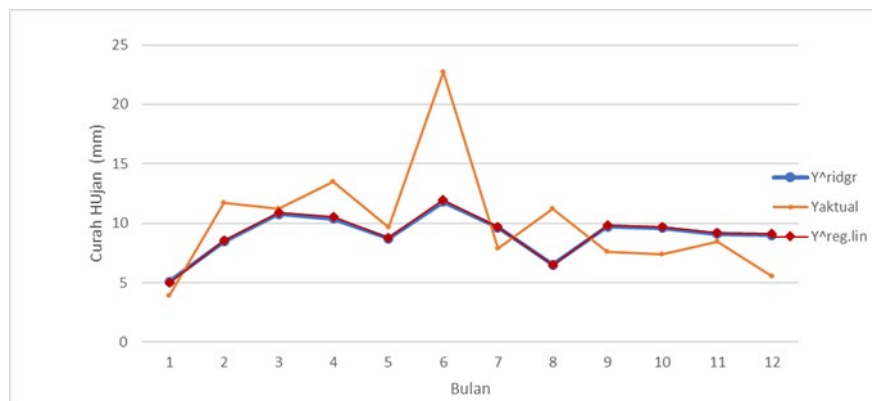
$$b_0 = \underline{Y} - b_3\underline{X}_3 - b_4\underline{X}_4 = 7,17798 - b_3 \cdot 2,31668 - b_4 \cdot 4,446214 = -4,268577$$

Sehingga persamaan *Ridge regression* apabila ditransformasikan ke bentuk persamaan linier berganda diperoleh persamaan berikut:

$$Y = -4,26858 + 2,6462 \cdot X_3 - 2,5632 \cdot X_4 \quad (12)$$

1.8. Perbandingan model regresi linier berganda dan regresi Ridge

Untuk mengetahui model regresi terbaik salah satunya dapat dilakukan secara visual, yaitu dengan membandingkan plotting data hasil prediksi model regresi linier berganda dan model *Ridge Regression*, seperti disajikan pada Gambar 4.



Gambar 4. Plotting Data

Berdasarkan Gambar 4 di atas, *plotting Ridge Regression* dan Regresi Linear Berganda hampir berimpit, ini berarti hasil prediksi *Ridge Regression* dan Regresi Linear Berganda hampir sama. Untuk dilakukan validasi data hasil prediksi dengan menggunakan beberapa ukuran kebaikan model seperti disajikan pada Tabel 6.

Tabel 6. Bias model dengan data aktual

Kriteria	<i>Ridge Regression</i> tanpa Outlier	Reg. Linear tanpa outlier
Korelasi Pearson	0,59	0,59
RMSE	4,3	4,26
MAPE	28,75%	28,82%

Sumber: Hasil Analisis, 2021

4. Kesimpulan

Beberapa kesimpulan yang diperoleh dari hasil penelitian ini adalah sebagai berikut:

1. Model Regresi Linear Berganda yang diperoleh tanpa data *outlier* adalah sebagai berikut:

$$Y = -44,229 + 2,728X_3 - 2,644X_4$$

Model regresi tersebut dapat diinterpretasikan bahwa tanpa setiap kenaikan jumlah kelembaban (X_3) sebesar 1 satuan dan penguapan (X_4) dianggap konstan, maka curah hujan (Y) naik sebesar

2,728 mm. Jika setiap kenaikan penguapan (X_4) sebesar 1 satuan dan kelembaban (X_3) dianggap konstan maka curah hujan (Y) turun sebesar 2,644 mm.

2. Model *Ridge Regression* yang diperoleh tanpa data *outlier* adalah sebagai berikut:

$$Y = -42,686 + 2,646X_3 - 2,563X_4$$

Model *Ridge Regression* tersebut dapat diinterpretasikan bahwa tanpa setiap kenaikan jumlah kelembaban (X_3) sebesar 1 satuan dan penguapan (X_4) dianggap konstan, maka curah hujan (Y) naik sebesar 2,646 mm. Jika setiap kenaikan penguapan (X_4) sebesar 1 satuan dan kelembaban (X_3) dianggap konstan maka curah hujan (Y) turun sebesar 2,563 mm.

3. Berdasarkan perhitungan yang telah dilakukan, diperoleh nilai RMSE untuk model Regresi Linear Berganda sebesar 4,263 dan untuk *Ridge Regression* sebesar 4,3. Nilai MAPE untuk model Regresi Linear Berganda sebesar 28,82% dan untuk *Ridge Regression* sebesar 28,75 %. Dengan demikian model *Ridge Regression* sedikit lebih baik dari model regresi linier berganda.

5. Rekomendasi

Berdasarkan hasil penelitian, maka hal yang dapat direkomendasikan oleh penulis adalah: Untuk peneliti selanjutnya diharapkan mencoba variasi lain dalam menentukan tetapan bias c , yaitu dengan rumus c_p Mallows dan membandingkan hasilnya dengan pemilihan tetapan bias c dengan rumus HKB.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada bapak Fadli Azis, S.Mat.,M.Mat, selaku pengelola Jurnal Riset Matematika dan Sains Terapan (JRMST), yang telah memberikan kesempatan untuk bisa publikasi paper hasil penelitian di JRMST volume pertama ini.

Referensi

- [1] Afham, M., Nur, M. dan Utami, T.W, “Pemodelan Regresi Ridge Pada Kasus Curah Hujan Di Kota Semarang”, *Prosiding Seminar Nasional*, Universitas Muhammadiyah Semarang, 2017.
- [2] Astuti, A. D, *Partial Least Square (PLS) dan Principal Component Regression (PCR) Untuk Regresi Linear Dengan Multikolinearitas Pada Kasus Indeks Pembangunan Manusia Di Kabupaten Gunung Kidul*, Yogyakarta: Universitas Negeri Yogyakarta, 2014.
- [3] Draper, N. S, *Analisis Regresi Terapan. Edisi 2 (Terjemahaan: Bambang Sumanti)*. Jakarta: PT. Gramedia Pustaka Utama, 1992.
- [4] Ghozali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 21 Update PLS Regresi*, Edisi 7., Semarang: UNDIP, 2013.
- [5] Gunarsih, A.K, *Klimatologi Pengaruh Iklim Terhadap Tanah dan Tanaman, Bumi Aksara*. Jakarta, 2006.
- [7] Hines, W. Montgomery, DC, *Probabilita dan Statistik Dalam Ilmu Rekayasa dan Manajemen*, Edisi kedua., Universitas Indonesia, Jakarta, 1990.
- [8] Hoerl, A.E, Kennard, R.W., & Baldwin, K.F, “A Simulation Study of Ridge and Other Regression Estimators”. *Communication in Statistics*, 5, 308-323, 1975.
- [9] Kumar, V. V., Ravinesh C. D. dan Ramachandran, V., “Total Rain Accumulation And Rain-Rate Analysis For Small Tropical Pasific Island : A Case Study Of Suva”, *Journal of Atmospheric Science Letter*, Fiji, 7, 53-58, 2006.
- [10] Montgomery dan Peck, *Introduction Linear Regression Analysis*, 2nd edition., John Willeydan Sons, INC, 1992.
- [11] Olson, David dan Shi, Yong, *Pengantar Ilmu Penggalian Data Bisnis*. Jakarta: Salemba empat, 2008.
- [12] Ryan, T.P, *Modern Regression Method*. New York: Wiley, 1997.

- [13] Soemartini, “Penyelesaian Multikolinearitas Melalui Metode Ridge Regression”, Tesis, Universitas Padjajaran, 2008
- [14] Wasilaine, T. L., Talakua, M. W., & Lesnussa, Y. A, “Model Regresi Ridge Untuk Mengatasi Model Regresi Linier Berganda Yang Mengandung Multikolinieritas”. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 8(1), 31–37, 2014.