

ESTIMASI SINTASAN PENDERITA DIABETES MELITUS: KOMPARASI KINERJA REGRESI PLS DAN LASSO

Gede Ary Prabha Yogesswara¹, Eka N. Kencana^{2§}, I Komang Gde Sukarsa³

¹Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: aryprabha28@gmail.com]

²Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: i.putu.enk@unud.ac.id]

³Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: sukarsakomang@yahoo.com]

[§]Corresponding Author

ABSTRACT

Partial least squares (PLS) regression and least absolute shrinkage and selection operator (LASSO) are the regression analysis techniques used to overcome the problems that can not be solved by ordinary least squares (OLS). The purpose of this research is to model and compare the performance of both PLS regression and LASSO to the diabetes mellitus study data which is divided into 30 groups of data redundancy as an example of microarray data. The survival time of diabetes mellitus patients as dependent variable while age, sex, body mass index, blood pressure, and six blood serum measurements as independent variables. By using paired sample t-test of adj R^2 value, the result of this research concluded that the mean of adj R^2 value of PLS regression is smaller than the mean of adj R^2 value of LASSO. In other words, the performance of LASSO is better than PLS regression.

Keywords: Adj R^2 , Diabetes Mellitus, LASSO, PLS Regression.

1. PENDAHULUAN

Analisis regresi merupakan suatu alat yang digunakan dalam statistika untuk memprediksi salah satu variabel maupun variabel lainnya dengan memanfaatkan hubungan antar variabel tersebut (Neter *et al.*, 1997). Pada dasarnya terdapat dua jenis analisis regresi linear yaitu analisis regresi linear sederhana dan analisis regresi linear berganda. Perbedaan dari kedua analisis regresi tersebut terdapat pada jumlah variabel yang terdapat pada model persamaannya. Jika analisis regresi linear sederhana hanya memanfaatkan hubungan antara satu variabel terikat dengan satu variabel bebas, maka analisis regresi linear berganda memanfaatkan hubungan lebih dari satu variabel bebas terhadap variabel terikat.

Dalam statistika, suatu model regresi linear dikatakan sesuai atau cocok apabila telah memenuhi asumsi dalam regresi. Asumsi dalam regresi adalah persyaratan yang ada saat menggunakan prosedur regresi linear. Terdapat sepuluh asumsi berdasarkan kerangka model regresi linear klasik, salah satunya adalah jum-

lah pengamatan dalam suatu penelitian lebih besar dari jumlah parameter yang akan diestimasi (Gujarati dan Porter, 2009). Namun, asumsi tersebut tidak dipenuhi oleh data *microarray*, karena pada data tersebut jumlah kovariat sangat banyak dalam jumlah sampel yang sedikit (Datta *et al.*, 2007).

Regresi *partial least squares* (PLS) dan *least absolute shrinkage and selection operator* (LASSO) dapat digunakan untuk menyelesaikan permasalahan data *microarray*. Regresi PLS adalah suatu teknik yang dikembangkan oleh Herman Wold pada tahun 1960an untuk membangun model yang bersifat prediksi saat variabel penjelas banyak dan sangat kolinear. Prediksi ini didapatkan dengan mengekstraksi himpunan faktor-faktor ortogonal yang disebut variabel laten yang mempunyai kekuatan prediksi terbaik (Abdi, 2007). Sedangkan LASSO adalah teknik yang diusulkan oleh Robert Tibshirani pada tahun 1996 untuk pemilihan model dan estimasi pada model regresi linear. Teknik ini meminimalkan jumlah kuadrat sisaan yang bergantung pada jumlah nilai mutlak

koefisien sehingga menyebabkan beberapa koefisien bernilai nol dan memberikan model yang dapat diinterpretasikan (Tibshirani, 1996).

Dalam perkembangannya, kedua teknik tersebut dapat menyelesaikan permasalahan yang hampir sama, namun belum diketahui secara pasti teknik mana yang memiliki kinerja lebih baik. Adapun penelitian sebelumnya untuk mengetahui kinerja regresi PLS dan LASSO dilakukan oleh Farmani dkk. (2012) pada data *microarray* diperoleh simpulan bahwa nilai *root mean square error prediction* (RMSEP) dari LASSO lebih kecil dibandingkan dengan regresi PLS. Dengan kata lain, kinerja LASSO lebih baik dari regresi PLS. Sedangkan penelitian yang dilakukan oleh Chong dan Jun (2005) saat terdapat multikolinearitas diperoleh simpulan bahwa regresi PLS memiliki kinerja yang lebih baik jika dibandingkan dengan LASSO.

Penelitian ini bertujuan untuk mengetahui model persamaan regresi PLS dan LASSO serta membandingkan kinerja kedua teknik tersebut menggunakan data penelitian diabetes melitus yang terbagi menjadi 30 kelompok pengulangan data dengan delapan pengamatan di setiap kelompok pengulangan.

2. METODE PENELITIAN

Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari data penelitian diabetes sejumlah 442 pengamatan oleh Efron *et al.* (2004) dari Stanford University melalui: <http://www4.stat.ncsu.edu/~boos/var.select/diabetes.tab.txt>.

Variabel Penelitian

Terdapat dua jenis variabel yang digunakan dalam penelitian ini, yaitu satu variabel terikat dengan sepuluh variabel bebas. Penjelasan tentang variabel yang digunakan disajikan pada Tabel 1.

Metode Analisis Data

Adapun langkah-langkah analisis data dalam penelitian ini adalah:

1. Mengambil delapan dari 442 pengamatan secara acak dengan menggunakan teknik pengambilan sampel dengan pengembalian,

sampai diperoleh 30 kelompok pengulangan data.

2. Membentuk matriks variabel bebas $X_{8 \times 10}$ dan vektor variabel terikat $Y_{8 \times 1}$ pada setiap kelompok pengulangan data yang sudah diperoleh.
3. Analisis data dengan regresi PLS dan LASSO.

Langkah-langkah analisis data dengan regresi PLS:

1. Melakukan pembagian data dalam satu kelompok menjadi dua, yaitu bagian pertama yang berisi enam pengamatan untuk menduga parameter model dan bagian kedua yang berisi dua pengamatan untuk validasi.

Tabel 1 Penjelasan Variabel Penelitian

Simbol	Keterangan	Definisi	Skala
y	Variabel Terikat	Waktu sintasan penderita diabetes melitus	Rasio
x_1	Variabel bebas	Umur penderita diabetes melitus	Rasio
x_2	Variabel bebas	Jenis kelamin penderita diabetes melitus	Nominal
x_3	Variabel bebas	Indeks masa tubuh	Rasio
x_4	Variabel bebas	Tekanan darah	Rasio
x_5	Variabel bebas	Total kolesterol	Rasio
x_6	Variabel bebas	<i>Low density lipoprotein (ldl)</i>	Rasio
x_7	Variabel bebas	<i>High density lipoprotein (hdl)</i>	Rasio
x_8	Variabel bebas	Salah satu pengukuran serum darah (<i>tch</i>)	Rasio
x_9	Variabel bebas	<i>Low tension glaucoma (ltg)</i>	Rasio
x_{10}	Variabel bebas	Kadar glukosa darah	Rasio

Sumber: Data diolah, 2018

2. Pada bagian untuk menduga parameter dilakukan beberapa langkah, sebagai berikut:

- a. Mencari vektor pembobot untuk matriks X dan Y melalui pembentukan dekomposisi nilai tunggal (SVD) dari *cross-product* matriks $S = X^T Y$.
 - b. Mencari vektor kolom $t = Xw = Ew$ untuk pembobot X , dan vektor kolom $u = Yq = Fq$ untuk pembobot Y .
 - c. Normalisasikan vektor kolom dari pembobot X menjadi $t^* = \frac{1}{\sqrt{t^T t}} t$.
 - d. Mencari *loadings* $p = X^T t^*$ matriks X dan *loadings* $q = Y^T t^*$ matriks Y .
 - e. Pendugaan komponen pertama terbentuk menggunakan: $X_{n+1} = X_n - t^* p^T$ dan $Y_{n+1} = Y_n - t^* q^T$.
 - f. Pendugaan komponen berikutnya dapat dimulai dengan SVD dari matriks *cross-product* hasil langkah e.
 - g. Menentukan koefisien untuk setiap variabel bebas untuk a komponen pertama dengan: $B = R(T^T T)^{-1} T^T Y = RQ^T$.
 - h. Menentukan komponen optimal dengan mencari nilai RMSEP yang terkecil menggunakan persamaan $\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$.
 - i. Membentuk model persamaan regresi PLS dengan meregresikan komponen yang terbentuk terhadap Y .
3. Mencari nilai estimasi waktu sintasan (\hat{y}) menggunakan model yang diperoleh baik pada bagian pertama maupun bagian kedua.
4. Mencari nilai R^2 menggunakan persamaan $1 - \frac{JKG}{JKT}$ dan nilai *adj* R^2 menggunakan persamaan $1 - \frac{JKG/(n-p-1)}{JKT/(n-1)}$, dengan JKG adalah jumlah kuadrat galat dan JKT adalah jumlah kuadrat total untuk mengetahui seberapa besar kecocokan model.

Sedangkan langkah-langkah analisis data dengan LASSO adalah sebagai berikut:

1. Mencari koefisien regresi menggunakan teknik OLS secara bertahap.
2. Mencari nilai parameter:

$$\hat{t} = \sum_{s=1}^k \sum_{(y_i, x^i) \in L^s} (y_i - \hat{y}_i)^2.$$
3. Memilih nilai parameter \hat{t} minimum berdasarkan hasil yang diperoleh pada langkah 2, selanjutnya disebut parameter *tuning t*.

4. Menentukan nilai estimasi koefisien LASSO menggunakan:

$$R(\hat{\beta}_0, \hat{\beta}_i) = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2$$

dengan syarat $\sum_{j=1}^n |\beta_j| \leq t$.

5. Membentuk model persamaan LASSO dari beberapa variabel bebas yang memiliki nilai koefisien tak nol.
6. Mencari nilai estimasi waktu sintasan (\hat{y}) menggunakan model persamaan LASSO yang diperoleh dari langkah 5.
7. Mencari nilai R^2 menggunakan persamaan $1 - \frac{JKG}{JKT}$ dan nilai *adj* R^2 menggunakan persamaan $1 - \frac{JKG/(n-p-1)}{JKT/(n-1)}$, dengan JKG adalah jumlah kuadrat galat dan JKT adalah jumlah kuadrat total untuk mengetahui seberapa besar kecocokan model.

Setelah melakukan analisis data dengan teknik regresi PLS dan LASSO, gunakan uji t terhadap nilai R^2 yang diperoleh untuk mengetahui perbedaan kinerja regresi PLS dan LASSO.

3. HASIL DAN PEMBAHASAN

Model Persamaan Regresi PLS dan LASSO

Model persamaan yang diperoleh dari teknik regresi PLS dan LASSO masing-masing merupakan model regresi dan berbentuk linear.

Regresi *Partial Least Squares* (PLS)

Regresi PLS merupakan suatu teknik yang menggabungkan analisis komponen utama dan regresi berganda yang bertujuan untuk memprediksi suatu variabel terikat dari beberapa variabel bebas (Abdi, 2007). Model persamaan regresi PLS untuk waktu sintasan penderita diabetes melitus adalah:

$$\hat{y}_{PLSi} = \hat{\beta}_{i0} + \hat{\beta}_{i1} w_{i1} + \hat{\beta}_{i2} w_{i2} + \dots + \hat{\beta}_{ip} w_{ip}$$

dengan \hat{y}_{PLSi} adalah pendugaan waktu sintasan penderita diabetes melitus; $\hat{\beta}_{i0}, \hat{\beta}_{i1}, \hat{\beta}_{i2}, \dots, \hat{\beta}_{ip}$ adalah koefisien regresi; $w_{i1}, w_{i2}, \dots, w_{ip}$ adalah komponen utama yang merupakan kombinasi linear variabel bebas x_1, x_2, \dots, x_p . Model persamaan regresi PLS untuk semua kelompok pengulangan data disajikan pada Tabel 2.

Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO merupakan teknik penyusutan koefisien sehingga menyebabkan beberapa koefisien bernilai nol dengan meminimalkan jumlah kuadrat sisaan yang bergantung pada $\sum_{j=1}^n |\beta_j| \leq t$ (Tibshirani, 1996). Model persamaan LASSO untuk waktu sintasan penderita diabetes melitus adalah:

$\hat{y}_{LASSO_i} = \hat{\beta}_{i0} + \hat{\beta}_{i1}x_1 + \hat{\beta}_{i2}x_2 + \dots + \hat{\beta}_{ip}x_p$
dengan \hat{y}_{LASSO_i} adalah pendugaan waktu sintasan penderita diabetes melitus; $\hat{\beta}_{i0}, \hat{\beta}_{i1}, \hat{\beta}_{i2}, \dots, \hat{\beta}_{ip}$ adalah koefisien regresi tak nol dan x_p merupakan variabel bebas yang

Tabel 2 Model Persamaan Regresi PLS

Ulangan	Model Persamaan	Ulangan	Model Persamaan
1	$\hat{y}_{PLS1} = -35.6 + w_{11} + w_{12}$	16	$\hat{y}_{PLS16} = -208 + w_{161} + w_{162}$
2	$\hat{y}_{PLS2} = 1136 + w_{21} + w_{22} + w_{23} + w_{24}$	17	$\hat{y}_{PLS17} = -76 + w_{171}$
3	$\hat{y}_{PLS3} = -87 + w_{31} + w_{32}$	18	$\hat{y}_{PLS18} = -442.7 + w_{181} + w_{182} + w_{183}$
4	$\hat{y}_{PLS4} = -73.6 + w_{41}$	19	$\hat{y}_{PLS19} = 576.4 + w_{191} + w_{192} + w_{193} + w_{194}$
5	$\hat{y}_{PLS5} = 5 + w_{51}$	20	$\hat{y}_{PLS20} = 41.2 + w_{201}$
6	$\hat{y}_{PLS6} = 457 + w_{61}$	21	$\hat{y}_{PLS21} = -191 + w_{211} + w_{212}$
7	$\hat{y}_{PLS7} = -76 + w_{71}$	22	$\hat{y}_{PLS22} = 164.6 + w_{221}$
8	$\hat{y}_{PLS8} = -311 + w_{81}$	23	$\hat{y}_{PLS23} = -70.6 + w_{231} + w_{232} - 0.54w_{233} + 1.371w_{234}$
9	$\hat{y}_{PLS9} = -179 + w_{91}$	24	$\hat{y}_{PLS24} = -604.7 + w_{241} + w_{242} + w_{243} + w_{244}$
10	$\hat{y}_{PLS10} = 61 + w_{101}$	25	$\hat{y}_{PLS25} = 370.7 + 1.09w_{251} + 1.33w_{252} + 0.86w_{253} + 10.92w_{254}$
11	$\hat{y}_{PLS11} = -45 + w_{111}$	26	$\hat{y}_{PLS26} = 245 + w_{261} + w_{262}$
12	$\hat{y}_{PLS12} = -16 + w_{121}$	27	$\hat{y}_{PLS27} = 164.8 + w_{271}$
13	$\hat{y}_{PLS13} = -65.18 + w_{131} + w_{132} + w_{133} + w_{134}$	28	$\hat{y}_{PLS28} = -268 + w_{281}$
14	$\hat{y}_{PLS14} = 634 + w_{141} + w_{142} + w_{143} + w_{144}$	29	$\hat{y}_{PLS29} = 236.7 + w_{291}$
15	$\hat{y}_{PLS15} = 569 + w_{151} + w_{152}$	30	$\hat{y}_{PLS30} = -468.4 + w_{301} + w_{302} + w_{303} + w_{304}$

Sumber: Data diolah, 2018

Terdapat dua jenis nilai R^2 , yaitu R^2 dan $adj R^2$. Nilai R^2 merupakan nilai R^2 biasa dan nilai $adj R^2$ merupakan nilai R^2 terkoreksi yang memperhatikan jumlah komponen maupun variabel bebas dalam model. Nilai R^2 dan $adj R^2$ semua kelompok pengulangan disajikan secara terperinci pada Tabel 4. Untuk melakukan pengujian kinerja digunakan nilai $adj R^2$ karena setiap

memiliki koefisien regresi tak nol. Model persamaan LASSO untuk semua kelompok pengulangan data disajikan secara terperinci pada Tabel 3.

Kinerja Regresi PLS dan LASSO

Kinerja dari regresi PLS dan LASSO dapat diketahui dari nilai koefisien determinasi (R^2). Nilai R^2 menjelaskan bagaimana kedekatan nilai estimasi variabel terikat \hat{Y} terhadap nilai variabel terikat yang sebenarnya Y (Gujarati dan Porter, 2009). Model dianggap baik apabila nilai R^2 mendekati satu.

kelompok pengulangan mempunyai jumlah komponen dan variabel bebas yang berbeda-beda.

Pada tabel 4 dapat diketahui bahwa rerata nilai $adj R^2$ regresi PLS dan LASSO masing-masing adalah 0.8455 dan 0.8873. Berdasarkan rerata tersebut diketahui bahwa rerata nilai $adj R^2$ regresi PLS lebih kecil dari LASSO.

Berdasarkan Tabel 2 dan Tabel 4 diperoleh model terbaik regresi PLS dengan nilai $adj R^2$ sebesar 0.9930 pada kelompok pengulangan ke-23 yang dinyatakan sebagai:

$$\hat{y}_{PLS23} = -70.6 + w_{231} + w_{232} - 0.54w_{233} + 1.371w_{234}.$$

Pada model persamaan tersebut, semua variabel bebas membentuk suatu kombinasi linear yang disebut komponen. Oleh karena itu, terdapat kesulitan dalam hal interpretasi komponen sehingga tidak dapat menentukan variabel bebas yang signifikan. Sedangkan berdasarkan Tabel 3 dan Tabel 4 diperoleh model terbaik LASSO dengan nilai $adj R^2$ sebesar 0.9946 pada kelompok pengulangan ke-4 yang dinyatakan sebagai:

$$\begin{aligned} \hat{y}_{LASSO4} = & -307.066 + 0.375x_1 \\ & -75.306x_2 + 4.239x_3 + 1.81x_4 \\ & + 0.251x_5 + 0.529x_7 - 2.096x_9 \\ & + 2.924x_{10}. \end{aligned}$$

Model persamaan tersebut diperoleh setelah terjadi proses penyusutan variabel bebas sehingga

tersisa delapan variabel bebas yang memiliki koefisien regresi tak nol yaitu $x_1, x_2, x_3, x_4, x_5,$

$x_7, x_9,$ dan x_{10} . Delapan variabel bebas yang tersisa tersebut merupakan variabel bebas yang secara signifikan memengaruhi variabel terikat \hat{y}_{LASSO4} .

Pengujian Kinerja Regresi PLS dan LASSO

Untuk mengetahui bagaimana kinerja regresi PLS dan LASSO, perlu dilakukan pengujian menggunakan uji t sampel berpasangan. Sebelum melakukan pengujian, dilakukan uji kenormalan pada data karena syarat dalam uji t adalah data harus menyebar normal.

Hipotesis nol uji kenormalan ini adalah data menyebar normal. Berdasarkan perhitungan, diperoleh bahwa nilai- $P = 0.3325$ lebih dari 0.05 yang artinya tidak cukup bukti untuk menolak hipotesis nol. Dengan kata lain, data menyebar normal.

Tabel 3 Model Persamaan LASSO

Ulangan	Model Persamaan	Ulangan	Model Persamaan
1	$\hat{y}_{LASSO1} = -113.016 + 55.865x_9$	16	$\hat{y}_{LASSO16} = -239.621 + 0.599x_1 + 2.977x_3 + 2.597x_4$
2	$\hat{y}_{LASSO2} = 165.375$	17	$\hat{y}_{LASSO17} = 76.959 + 0.212x_1 + 23.645x_8$
3	$\hat{y}_{LASSO3} = 186.125$	18	$\hat{y}_{LASSO18} = -325.65 + 3.373x_4 + 1.026x_6 + 0.337x_{10}$
4	$\hat{y}_{LASSO4} = -307.066 + 0.375x_1 - 75.306x_2 + 4.239x_3 + 1.81x_4 + 0.251x_5 + 0.529x_7 - 2.096x_9 + 2.924x_{10}$	19	$\hat{y}_{LASSO19} = -539.142 + 42.938x_2 + 4.324x_3 - 0.752x_5 - 0.814x_7 + 135.712x_9 + 1.182x_{10}$
5	$\hat{y}_{LASSO5} = 109.375$	20	$\hat{y}_{LASSO20} = 92.25$
6	$\hat{y}_{LASSO6} = -252.83 + 30.427x_2 + 6.129x_3 - 0.792x_5 - 1.425x_7 + 91.884x_9$	21	$\hat{y}_{LASSO21} = -22.1 + 2.354x_1 + 8.169x_9$
7	$\hat{y}_{LASSO7} = 25.878 + 3.948x_3$	22	$\hat{y}_{LASSO22} = 155.75$
8	$\hat{y}_{LASSO8} = -153.584 + 2.222x_4 + 0.922x_{10}$	23	$\hat{y}_{LASSO23} = -354.612 + 2.484x_3 + 3.738x_4 - 1.201x_5 + 37.872x_9 + 1.7x_{10}$
9	$\hat{y}_{LASSO9} = -411.716 + 7.104x_3 - 0.885x_5 + 15.958x_8 + 78.06x_9 + 1.374x_{10}$	24	$\hat{y}_{LASSO24} = -206.439 - 43.453x_2 + 9.669x_3 - 0.188x_7 + 0.073x_8 + 2.029x_{10}$
10	$\hat{y}_{LASSO10} = 199.375$	25	$\hat{y}_{LASSO25} = -499.847 + 0.363x_1 + 15.943x_3 + 1.418x_4 + 23.459x_9$
11	$\hat{y}_{LASSO11} = -149.873 + 0.385x_1 + 10.709x_3 + 0.94x_4 - 0.482x_5$	26	$\hat{y}_{LASSO26} = 153.875$
12	$\hat{y}_{LASSO12} = -253.671 + 1.781x_3 + 75.552x_9$	27	$\hat{y}_{LASSO27} = 132.25$
13	$\hat{y}_{LASSO13} = -37.883 + 0.858x_5 + 1.433x_7$	28	$\hat{y}_{LASSO28} = -378.02 - 40.834x_2 + 5.659x_4 - 0.515x_7 + 18.186x_8$
14	$\hat{y}_{LASSO14} = 84.119 + 20.768x_8$	29	$\hat{y}_{LASSO29} = 20.142 + 3.682x_3$
15	$\hat{y}_{LASSO15} = 191.875$	30	$\hat{y}_{LASSO30} = -222.419 + 7.609x_3 + 0.73x_5$

Sumber: Data diolah, 2018

Hipotesis yang digunakan dalam pengujian ini adalah:

1. $H_0: \mu_{PLS} \geq \mu_{LASSO}$ (rerata nilai *adj R*² regresi PLS lebih besar atau sama dengan LASSO).
2. $H_1: \mu_{PLS} < \mu_{LASSO}$ (rerata nilai *adj R*² regresi PLS lebih kecil dari LASSO).

Dasar pengambilan keputusan yang digunakan adalah sebagai berikut:

1. Jika nilai-*P* > 0.05, maka tidak ada cukup bukti menolak hipotesis nol.

2. Jika nilai-*P* < 0.05, maka hipotesis nol ditolak.

Berdasarkan perhitungan, diperoleh nilai-*P* sebesar 0.004525 lebih kecil dari 0.05. Menurut dasar pengambilan keputusan 2, berarti hipotesis nol ditolak. Dengan kata lain rerata nilai *adj R*² regresi PLS lebih kecil dari rerata nilai *adj R*² LASSO. Oleh karena itu, kinerja LASSO lebih baik dibandingkan dengan kinerja regresi PLS.

Tabel 4 Nilai *Adj R*² Model Persamaan Regresi PLS dan LASSO.

Ulangan	<i>R</i> ²		<i>adj R</i> ²		Beda (selisih) <i>adj R</i> ²
	Regresi PLS	LASSO	Regresi PLS	LASSO	
1	0.8248	0.8826	0.7547	0.8631	-0.1084
2	0.9461	0.8896	0.8742	0.8896	-0.0154
3	0.9767	0.8560	0.9674	0.8560	0.1114
4	0.9050	0.9992	0.8892	0.9946	-0.1054
5	0.7808	0.7385	0.7442	0.7385	0.0057
6	0.8825	0.9844	0.8630	0.9453	-0.0823
7	0.7830	0.8307	0.7468	0.8024	-0.0556
8	0.8398	0.9219	0.8130	0.8906	-0.0776
9	0.8661	0.9673	0.8438	0.8857	-0.0419
10	0.9123	0.9048	0.8977	0.9048	-0.0071
11	0.8467	0.9855	0.8212	0.9662	-0.1450
12	0.8523	0.9478	0.8277	0.9269	-0.0992
13	0.9470	0.9709	0.8764	0.9593	-0.0829
14	0.9575	0.8599	0.9009	0.8365	0.0644
15	0.9427	0.8614	0.9197	0.8614	0.0583
16	0.9617	0.9632	0.9464	0.9355	0.0109
17	0.8620	0.9249	0.8390	0.8948	-0.0558
18	0.9755	0.9929	0.9571	0.9876	-0.0305
19	0.8519	0.9406	0.6545	0.5840	0.0705
20	0.9321	0.9155	0.9208	0.9155	0.0053
21	0.9666	0.9412	0.9532	0.9177	0.0355
22	0.8693	0.8743	0.8476	0.8743	-0.0267
23	0.9970	0.9955	0.9930	0.9843	0.0087
24	0.9430	0.9976	0.8669	0.9917	-0.1248
25	0.8778	0.9968	0.7150	0.9925	-0.2775
26	0.8502	0.7912	0.7902	0.7912	-0.0010
27	0.7945	0.7850	0.7603	0.7850	-0.0247
28	0.9295	0.9840	0.9178	0.9626	-0.0448
29	0.7640	0.7763	0.7246	0.7390	-0.0144
30	0.8881	0.9593	0.7388	0.9430	-0.2042
Rerata	0.8909	0.9146	0.8455	0.8873	-0.0418
Simpangan baku	0.0647	0.0750	0.0866	0.0929	0.0819

Sumber: Data diolah, 2018

4. KESIMPULAN DAN SARAN

Berdasarkan hasil yang diperoleh, terdapat beberapa kesimpulan, yaitu:

1. Model terbaik regresi PLS dengan nilai $adj R^2$ sebesar 0.9930 pada kelompok pengulangan ke-23 dinyatakan sebagai:

$$\hat{y}_{PLS23} = -70.6 + w_{231} + w_{232} - 0.54w_{233} + 1.371w_{234} .$$

Sedangkan model terbaik LASSO dengan dengan nilai $adj R^2$ sebesar 0.9946 pada kelompok pengulangan ke-4 dinyatakan sebagai:

$$\hat{y}_{LASSO4} = -307.066 + 0.375x_1 - 75.306x_2 + 4.239x_3 + 1.81x_4 + 0.251x_5 + 0.529x_7 - 2.096x_9 + 2.924x_{10} .$$

2. Melalui pengujian kinerja regresi PLS dan LASSO menggunakan uji t sampel berpasangan, diketahui bahwa rerata nilai $adj R^2$ regresi PLS lebih kecil dari LASSO. Sehingga kinerja LASSO lebih baik dibandingkan dengan kinerja regresi PLS.

Bagi peneliti lainnya, diharapkan melakukan penelitian tentang regresi PLS atau LASSO menggunakan data yang sangat besar, baik dari jumlah pengamatan maupun informasi lain yang mendukung kejelasan data. Sebagai perbandingan, dapat juga menggunakan teknik lainnya pada data yang tidak dapat diselesaikan dengan teknik OLS.

DAFTAR PUSTAKA

- Abdi, H. (2007). Partial Least Squares Regression. *Encyclopedia of Measurements and Statistics*. Thousand Oaks (CA): Sage.
- Chong, I., dan Jun, C. (2005). Performance of Some Variable Selection Methods when Multicollinearity is Present. *Chemometrics and Intelligent Laboratory Systems*, 78, 103-112.
- Datta, S., Le-Rademacher, J., dan Datta, S. (2007). Predicting Patient Survival from Microarray Data by Accelerated Failure Time Modeling Using Partial Least Squares and LASSO. *Biometrics*, 63, 259-271.
- Efron, B., Hastie, T., Johnstone, I., dan Tibshirani, R. 2004. Least Angle Regression. <http://www4.stat.ncsu.edu/~boos/var.select/diabetes.tab.txt>. Diakses: 9 Februari 2018.
- Farmani, K. D., Kencana, P. E. N., dan Sukarsa, K. G. 2012. Perbandingan Analisis Least Absolute Shrinkage and Selection Operator (LASSO) dan Partial Least Squares (PLS) (Studi Kasus: Data Microarray). *e-Jurnal Matematika*, 1(1), pp. 75-80.
- Gujarati, D. N., dan Porter, D. C. 2009. *Basic Econometrics Fifth Edition*. New York: The McGraw-Hill Companies, Inc.
- Neter, J., Wasserman, W., dan Kutner, M. H. 1997. *Model Linier Terapan Buku I: Analisis Regresi Linear Sederhana*. Penerjemah: Bambang Sumantri. Bogor: Jurusan Statistika FMIPA-IPB.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.