

ANALISIS REGRESI DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO) DALAM MENGATASI MULTIKOLINEARITAS

Imi Sartika, Naomi Nessyana Debaraja, Nurfitri Imro'ah

INTISARI

Multikolinearitas merupakan salah satu uji asumsi klasik pada analisis regresi. Terjadinya multikolinearitas berarti antara variabel prediktor yang satu dengan yang lainnya saling berkorelasi. Apabila data penelitian mengandung multikolinearitas akan menimbulkan masalah, yaitu dapat mengakibatkan koefisien regresi yang dihasilkan tidak efisien. Salah satu solusi untuk mengatasi masalah tersebut dapat menggunakan metode least absolute shrinkage and selection operator (LASSO). Metode ini menyusutkan koefisien regresi dari variabel prediktor yang memiliki korelasi tinggi menjadi tepat pada nol atau mendekati nol. Koefisien LASSO dicari dengan menggunakan pemrograman kuadratik sehingga digunakan algoritma least absolute and regression (LARS) yang lebih efisien dalam komputasi LASSO. Data penelitian yang digunakan ialah banyaknya kasus pneumonia pada balita di 34 puskesmas di Kota Pontianak dan Kabupaten Mempawah. Pada data tersebut terdapat dua variabel prediktor yang mengalami masalah multikolinearitas, yaitu pemberian imunisasi DPT-Hb3-Hib3 (X_2) dan pemberian imunisasi campak (X_3), sehingga masalah tersebut akan diatasi menggunakan metode LASSO. Berdasarkan analisis yang telah dilakukan, diperoleh model regresi terbaik menggunakan LASSO pada tahap ketiga saat nilai $C_p = 2,62$. Hasil analisis tersebut menunjukkan bahwa adanya perubahan koefisien regresi yang mendekati nol dan tepat nol, yaitu pada variabel X_2 dari 0,285 menjadi 0,192 dan X_3 dari 0,024 menjadi 0. Hal ini menunjukkan bahwa multikolinearitas yang terdeteksi telah teratasi.

Kata Kunci : multikolinearitas, LASSO, LARS, pneumonia.

PENDAHULUAN

Analisis regresi merupakan studi tentang hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor. Analisis regresi terbagi menjadi dua, yaitu analisis regresi sederhana dan analisis regresi berganda. Analisis regresi sederhana memiliki tujuan memprediksi nilai variabel respon yang didasarkan pada nilai variabel prediktor, banyaknya variabel respon dan variabel prediktor pada analisis regresi sederhana adalah satu dan tidak bisa lebih dari itu. Tujuan analisis regresi berganda ialah untuk memprediksi pengaruh ada atau tidaknya hubungan fungsional antara dua variabel prediktor atau lebih terhadap satu variabel respon [1]. Proses menganalisis terkadang muncul masalah spesifik, salah satunya adalah masalah multikolinearitas.

Masalah multikolinearitas terjadi ketika antara variabel prediktor satu dengan variabel prediktor lainnya saling berkorelasi. Apabila masalah tersebut terjadi maka dapat mengakibatkan penduga koefisien regresi yang dihasilkan menjadi tidak efisien. Salah satu metode untuk mengatasi masalah tersebut adalah metode LASSO. LASSO merupakan metode yang dapat menyusutkan koefisien regresi pada variabel prediktor yang memiliki korelasi dan galat yang tinggi, dengan tujuan koefisien regresi tersebut menjadi mendekati nol atau bahkan sama dengan nol, sehingga LASSO melakukan peran sebagai metode seleksi variabel sekaligus mengatasi multikolinearitas [2].

Keakuratan perhitungan metode LASSO menjadi lebih baik dengan menggunakan algoritma LARS [7]. Algoritma LARS merupakan algoritma yang dapat dimodifikasi menjadi algoritma komputasi untuk metode LASSO. Hasil modifikasi tersebut menghasilkan efisiensi algoritma dalam menduga koefisien LASSO. Algoritma LARS mengarahkan setiap variabel prediktor secara terus-menerus menuju titik terendah sehingga menyebabkan korelasi terhadap galat menjadi menurun dalam nilai absolut.

Data yang digunakan merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kota Pontianak dan Kabupaten Mempawah tahun 2015. Variabel respon yang digunakan adalah banyaknya kasus pneumonia pada balita (Y) sedangkan variabel prediktor yang digunakan yaitu pemberian ASI eksklusif (X_1), pemberian imunisasi DPT-Hb3-Hib3 (X_2), pemberian imunisasi campak (X_3), dan rumah berperilaku hidup bersih dan sehat (X_4) pada 34 puskesmas di Kota Pontianak dan Kabupaten Mempawah. Pneumonia merupakan penyakit yang disebabkan oleh virus, bakteri, dan jamur serta merupakan penyakit menular yang ditularkan melalui udara. Gejala penyakit ini berupa nafas cepat dan sesak nafas, dikarenakan paru-paru meradang secara mendadak. Laporan dari UNICEF tahun 2015 menyatakan bahwa Indonesia merupakan negara ke sepuluh dengan kematian balita terbesar akibat pneumonia [8]. Penyakit ini menyumbang 16% dari seluruh kematian anak dibawah 5 tahun, yang menyebabkan kematian sebanyak 920.126 balita atau lebih dari 2.500 per hari atau diperkirakan 2 anak balita meninggal setiap jamnya pada tahun 2015. Banyaknya kasus pneumonia di Provinsi Kalimantan Barat pada tahun 2015 untuk 14 Kabupaten/Kota ialah Kabupaten Mempawah menempati urutan kedua setelah Kota Pontianak, yaitu 848 kasus dan 986 kasus [9].

Data penelitian yang digunakan akan dilakukan pengecekan terlebih dahulu apakah terjadi masalah multikolinearitas atau tidak. Berdasarkan hasil pengecekan tersebut, diketahui bahwa data tersebut mengalami masalah multikolinearitas yaitu pada variabel pemberian imunisasi DPT-Hb3-Hib3 (X_2) dan pemberian imunisasi campak (X_3). Sehingga tujuan dari penelitian ialah mengatasi multikolinearitas yang terjadi menggunakan metode LASSO dan memperoleh hasil analisis model regresi menggunakan metode LASSO dengan algoritma LARS. Setelah mengetahui data yang digunakan mengalami multikolinearitas, yang dilakukan selanjutnya ialah mengatasi multikolinearitas yang terjadi menggunakan metode LASSO dengan algoritma LARS. Langkah pertama yang dilakukan, yaitu menduga koefisien menggunakan metode LASSO. Menduga koefisien tersebut dilakukan dengan tahapan awalnya ialah menentukan variabel prediktor yang masuk ke dalam model atau seleksi variabel. Variabel-variabel prediktor yang masuk ke dalam model akan menghasilkan nilai penduga koefisien LASSO, nilai tersebut akan berubah seiring dengan dilakukannya seleksi variabel. Setelah mendapatkan calon koefisien LASSO, maka terbentuklah beberapa model regresi. Model-model regresi tersebut akan ditentukan yang mana merupakan model terbaik dengan menggunakan statistik C_p mallow yang memiliki nilai C_p minimum.

ANALISIS REGRESI

Analisis regresi merupakan studi tentang hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor. Selain model hubungan, dengan analisis regresi juga dapat mengetahui variabel prediktor mana saja yang secara regresi berpengaruh terhadap variabel respon, seberapa besar pengaruh yang diberikan variabel prediktor yang diberikan terhadap variabel respon, kemudian variabel prediktor mana yang paling dominan pengaruhnya terhadap variabel respon. Analisis regresi terbagi menjadi dua, yaitu regresi sederhana dan regresi berganda.

Analisis regresi sederhana ialah regresi yang terdiri dari satu variabel prediktor dan satu variabel respon. Regresi sederhana untuk meramalkan atau menaksir pengaruh satu variabel prediktor terhadap satu variabel respon. Bentuk umum regresi sederhana adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + e$$

Analisis regresi berganda merupakan pengembangan dari analisis regresi sederhana. Regresi berganda untuk memprediksi pengaruh dua atau lebih variabel prediktor terhadap variabel respon atau untuk membuktikan ada atau tidaknya hubungan fungsional antara dua variabel prediktor atau lebih dengan sebuah variabel respon. Semua asumsi dan makna persamaan regresi sederhana berlaku juga dalam regresi berganda. Perbedaannya hanya terletak pada rumusnya saja. Berikut ini bentuk persamaan regresi berganda [1]:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + e \text{ untuk } i = 1, 2, \dots, n$$

dimana:

- Y : pengamatan dari variabel respon
- $X_{i1}, X_{i2}, \dots, X_{ip}$: pengamatan dari variabel prediktor ke 1, 2, ..., p
- β_0 : konstanta
- $\beta_1, \beta_2, \dots, \beta_p$: parameter regresi
- e : galat bagi pengamatan

ORDINARY LEAST SQUARE (OLS)

OLS digunakan untuk menduga koefisien regresi dengan cara meminimumkan jumlah kuadrat galat. Penaksir OLS dari β diperoleh menggunakan [1]:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

yang dapat ditulis secara singkat dalam notasi matriks sebagai berikut:

$$Y = X\hat{\beta} + e \tag{1}$$

dimana:

- Y : vektor $n \times 1$ pengamatan pada variabel respon
- X : matriks $n \times k$ yang memberikan n pengamatan atas k -variabel prediktor
- $\hat{\beta}$: suatu vektor kolom k -unsur dari penaksir OLS koefisien regresi
- e : vektor kolom $N \times 1$ dari N residual

dari Persamaan 1, diperoleh

$$e = Y - X\hat{\beta}$$

Oleh karena itu

$$\begin{aligned} e^T e &= (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X \hat{\beta} \end{aligned} \tag{2}$$

dari Persamaan 2 dapat diturunkan $e^T e$ terhadap $\hat{\beta}$ yaitu:

$$\frac{\partial e^T e}{\partial \hat{\beta}} = \frac{\partial (Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X \hat{\beta})}{\partial \hat{\beta}}$$

Sehingga diperoleh

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

MULTIKOLINEARITAS

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frich. Multikolinearitas berarti antara variabel prediktor yang satu dengan variabel prediktor yang lainnya saling berkorelasi, terkadang korelasinya mendekati sempurna atau koefisien korelasinya tinggi atau bahkan sama dengan satu. Menguji terjadi atau tidaknya multikolinearitas terhadap data penelitian dapat menggunakan *variance inflation factor* (VIF). Ketentuan pada VIF adalah jika nilai VIF melebihi angka 10, maka terjadi multikolinearitas dalam model. Nilai VIF dapat dicari menggunakan rumus sebagai berikut [3]:

$$VIF_j = \frac{1}{(1-R_j^2)}, \text{ dimana } j = 1, 2, 3, \dots, k$$

dimana:

- R_j^2 : koefisien determinasi yang didapat dari variabel prediktor X_j yang diregresikan dengan variabel prediktor lainnya.

STATISTIK C_p MALLOW

Statistik C_p mallow adalah teknik untuk pemilihan model dalam regresi dan didefinisikan sebagai kriteria yang diterima untuk menilai kecocokan saat model yang berbeda parameter sedang dibandingkan. Statistik C_p mallow dapat dihitung menggunakan [4]:

$$C_p = \frac{JKG_p}{\hat{\sigma}^2} - n + 2p$$

dengan:

JKG_p : jumlah kuadrat galat dari model yang ditentukan

$\hat{\sigma}^2$: estimasi variansi sisaan dan n adalah banyaknya data penelitian

p : banyaknya parameter dalam model.

LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO)

Metode LASSO merupakan hasil pengembangan dari Tibshirani dan diperkenalkan pertama kali pada tahun 1996 sebagai penyeleksi variabel dengan pendugaan parameter. Metode LASSO melakukan penyusutan koefisien regresi dari variabel yang memiliki korelasi tinggi terhadap galat, dengan tujuan koefisien regresi tersebut mendekati nol atau bahkan sama dengan nol. Sehingga metode LASSO dapat melakukan peran sebagai seleksi variabel sekaligus mengatasi multikolinearitas.

Penduga koefisien LASSO diperoleh dengan meminimumkan jumlah kuadrat galat (JKG) yang ditulis sebagai berikut [5]:

$$\hat{\beta}^{LASSO} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij})^2 \quad (3)$$

dengan syarat $\sum_{j=1}^k |\beta_j| \leq t$, dimana $t = \sum_{j=1}^k |\hat{\beta}_j|$. Nilai t merupakan suatu besaran yang mengontrol besarnya penyusutan pada pendugaan koefisien LASSO dengan $t \geq 0$. Nilai t yang kecil menyebabkan beberapa koefisien regresi menjadi bernilai nol sehingga variabel-variabel yang berpengaruh besar dalam model terpilih dan variabel-variabel yang berpengaruh kecil tereliminasi. Hal ini menyebabkan solusi LASSO memberikan hasil yang efisien. Jika nilai $t < t_0$ dimana $t_0 = \sum_{j=1}^k |\hat{\beta}_j|$ akan menyebabkan koefisien OLS menyusut mendekati nol atau tepat pada nol sehingga LASSO berperan sebagai seleksi variabel, akan tetapi jika $t > t_0$ maka penduga koefisien LASSO memberikan hasil yang sama dengan penduga OLS. Pendugaan koefisien LASSO dilakukan secara bertahap dengan menetapkan nilai koefisien awal semua variabel prediktor bernilai nol.

LEAST ABSOLUTE AND REGRESSION (LARS)

LARS merupakan suatu metode regresi yang algoritmanya dapat dimodifikasi menjadi algoritma komputasi untuk metode LASSO. Hasil modifikasi LARS untuk LASSO menghasilkan efisiensi algoritma dalam menduga koefisien LASSO. LARS mengarahkan setiap variabel prediktor secara terus menerus menuju titik terendah sehingga menyebabkan korelasi dengan galat menjadi menurun dalam nilai absolut. Prosedur LARS bekerja dengan memulai perhitungan dengan semua koefisien sama dengan nol, sehingga akan menemukan variabel prediktor yang paling berkorelasi dengan variabel respon. Algoritma LARS, sebagai berikut [6]:

1. Melakukan standarisasi variabel prediktor sehingga memiliki nilai tengah nol dan ragam satu. Mulai dengan galat $\varepsilon = Y - \bar{Y}$ dan $\beta_0, \beta_1, \beta_2, \dots, \beta_k = 0$.
2. Mencari variabel prediktor yang paling berkorelasi dengan galat.
3. Mengubah nilai β_j dari nol bergerak ke arah koefisien *least-square* (X_j, ε) , sampai variabel prediktor yang lain memiliki korelasi yang sama seperti dengan galat sekarang.
4. Mengubah β_j dan β_k bergerak menuju arah yang ditentukan oleh koefisien *joint least-square* dari galat saat ini (X_j, X_k) , sampai semua variabel prediktor yang lain memiliki korelasi yang sama dengan galat sekarang.

5. Mengulang langkah ke-empat sampai semua variabel prediktor masuk ke dalam model. Setelah min $(n - 1, p)$ langkah, solusi model *least square* didapatkan.

Modifikasi algoritma LARS untuk mendapatkan solusi LASSO adalah dengan mengubah langkah ke-empat menjadi sebagai berikut: apabila koefisien yang bukan nol mencapai nol, maka variabel tersebut dikeluarkan dari gugus variabel aktif dan dihitung kembali arah dari *joint least-square* sekarang. Algoritma LARS selalu mengambil k langkah untuk mendapatkan penduga kuadrat terkecil secara penuh, sedangkan modifikasi LASSO dapat memiliki lebih dari k langkah untuk mendapatkannya.

PERHITUNGAN PENDUGA KOEFISIEN LASSO MENGGUNAKAN ALGORITMA LARS

LARS melakukan estimasi $\hat{\mu} = X \times \hat{\beta}$, dengan langkah-langkah yang berurutan dan setiap langkah akan menambah satu kovariat dalam model. Nilai $\hat{\mu}$ diperoleh dari teknik iterasi, nilai awal yang digunakan adalah $\hat{\mu}_0 = \mathbf{0}$. Langkah-langkah menduga koefisien LASSO dengan algoritma LARS, sebagai berikut [7]:

1. Mencari vektor yang sebanding dengan vektor korelasi (\hat{c}) antara variabel prediktor dengan galat dari setiap variabel prediktor, serta menentukan nilai korelasi absolut terbesar (\hat{C}), dengan cara berikut:

$$\hat{c} = X^T \times (Y - \hat{\mu}) \text{ dan } \hat{C} = \max\{|\hat{c}_j|\}$$

2. Menentukan X_A , yang dimana A adalah variabel prediktor yang memiliki nilai korelasi multak terbesar. Didefinisikan matriks:

$$X_A = (\dots, s_j X_j, \dots), \text{ dengan tanda } s_j \text{ bernilai } \pm 1$$

3. Menghitung *equingular vector* (u_a), *equingular vector* merupakan vektor yang membagi sudut dan kolom-kolom X_A menjadi sama besar dengan besar sudutnya kurang dari 90° . Nilai *equingular vector* diperoleh dengan cara berikut:

$$u_a = X_A \times w_a$$

Dengan mencari $w_a = A_A \times G_A^{-1} \times \mathbf{1}_A$, $A_A = (\mathbf{1}'_A \times G_A \times \mathbf{1}_A)^{-\frac{1}{2}}$, $G_A = X^T \times X_A$ dimana $\mathbf{1}$ merupakan matriks dengan entrinya 1.

4. Menghitung panjang u_a ($\hat{\gamma}$), dengan melakukan definisi vektor *inner product* terlebih dahulu, yaitu

$$a \equiv X^T \times u_a.$$

Sehingga $\hat{\gamma}$ dapat diperoleh dengan persamaan berikut:

$$\hat{\gamma} = \min_{j \in A^c}^+ \left\{ \frac{\hat{C} - \hat{c}}{A_A - a_A}, \frac{\hat{C} + \hat{c}}{A_A + a_A} \right\}$$

$\min_{j \in A^c}^+$ menunjukkan bahwa yang dipilih adalah nilai minimum positif dari j dan yang bukan merupakan himpunan A . Pada perhitungan akhir memperoleh nilai $\hat{\gamma}$ menggunakan rumus $\hat{\gamma} = \frac{\hat{c}_m}{A_m}$.

5. Memperbaharui nilai $\hat{\mu}_A$, yakni dengan menggunakan:

$$\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma} u_a \text{ dan } A_+ = A - \{j\}$$

6. Algoritma LARS untuk LASSO mengharuskan $sign(\hat{\beta}_j) = sign(\hat{c}_j) = s_j$

7. Langkah ke-6 dapat diperoleh apabila $\tilde{\gamma} = \min_{\gamma_j > 0} \{\gamma_j\}$, dengan $\gamma_j = \frac{-\tilde{\beta}_j}{s_j \times w_{aj}}$

Apabila $\tilde{\gamma}$ kurang dari $\hat{\gamma}$ maka $\beta_j(\gamma)$ bukan solusi untuk LASSO karena pembatasan tanda dilanggar. Jadi, algoritma LARS untuk LASSO harus memenuhi kondisi berikut, apabila $\tilde{\gamma} < \hat{\gamma}$ hentikan proses LARS pada tahapan $\gamma = \tilde{\gamma}$ dan hilangkan variabel j dari perhitungan langkah kedua selanjutnya. Variabel j dimasukkan kembali ke dalam perhitungan LARS pada tahap selanjutnya.

8. Melakukan pengecekan apakah $sign(\hat{\beta}_j) = sign(\hat{c}_j) = s_j$, pada seleksi variabel pertama sudah memperoleh $sign(\hat{\beta}_j) = sign(\hat{c}_j) = s_j = 1$, apabila sudah memiliki tanda yang sama, maka dapat

melanjutkan seleksi variabel selanjutnya.

9. Mengulang langkah-langkah yang sama untuk setiap seleksi variabelnya sehingga semua variabel prediktor telah terseleksi.

CONTOH KASUS

Data yang digunakan merupakan data sekunder, yaitu data yang diperoleh dari Dinas Kesehatan Kota Pontianak dan Kabupaten Mempawah. Data yang digunakan adalah data banyaknya kasus pneumonia pada balita, pemberian ASI eksklusif, pemberian imunisasi DPT-Hb3-Hib3, pemberian imunisasi campak, dan rumah berperilaku hidup bersih dan sehat pada 34 puskesmas di Kota Pontianak dan Kabupaten Mempawah. Data tersebut dianalisis menggunakan *software R*. Pengerjaan awal yang dilakukan ialah menghitung gambaran umum data penelitian yang diringkaskan dalam statistika deskriptif. Berikut gambaran umum data penelitian yang diringkaskan dalam statistika deskriptif pada Tabel 1, sebagai berikut:

Tabel 1 Statistik Deskriptif Variabel Respon dan Variabel Prediktor

Variabel	Minimum	Maksimum	Rata-rata	Standar Deviasi	Variansi
Y	0	444	43	81,854	6700,228
X_1	28	271	93	57,608	3318,796
X_2	108	1194	390	236,257	55817,39
X_3	109	1032	41	238,7394	56996,51
X_4	9	506	89	85,453	7302,275

Tabel 1 merupakan tabel statistik deskriptif variabel respon dan variabel prediktor, dimana pada hasil rata-rata dilakukan pembulatan. Selanjutnya yaitu memperoleh persamaan regresi menggunakan metode OLS dan menentukan parameter OLS mana yang signifikan terhadap variabel respon. Persamaan regresi menggunakan metode OLS sebagai berikut:

$$\hat{Y} = 5,372 - 0,595X_1 + 0,285X_2 - 0,024X_3 - 0,093X_4$$

Setelah melakukan uji signifikansi, diperoleh variabel yang signifikan terhadap Y adalah variabel X_1 sedangkan variabel lainnya tidak signifikan. Persamaan regresi menggunakan OLS telah diperoleh, tahap berikutnya yaitu mendeteksi adanya multikolinearitas. Pendeteksian multikolinearitas dilakukan dengan melihat nilai VIF. Hasil uji multikolinearitas disajikan pada Tabel 2.

Tabel 2 Hasil Uji Multikolinearitas

Variabel	Nilai VIF
X_1	1,688
X_2	13,608*
X_3	12,955*
X_4	1,048

*terjadi multikolinearitas

Berdasarkan Tabel 2 variabel X_2 dan X_3 , yaitu pemberian imunisasi DPT-HB3-HiB3 dan pemberian imunisasi campak memiliki nilai VIF lebih besar dari 10 sehingga dapat disimpulkan bahwa data yang digunakan mengalami masalah multikolinearitas. Sehingga yang dilakukan selanjutnya yaitu mengatasi multikolinearitas tersebut menggunakan metode LASSO dengan algoritma LARS. Tahap awal yang dilakukan adalah dengan menentukan variabel prediktor yang masuk kedalam model. Berikut hasil masuknya variabel prediktor kedalam model yang disajikan pada Tabel 3.

Tabel 3 Masuknya Variabel Prediktor kedalam Model

Tahapan	1	2	3	4
Variabel Prediktor yang Masuk ke Model	X_2	X_1	X_4	X_3

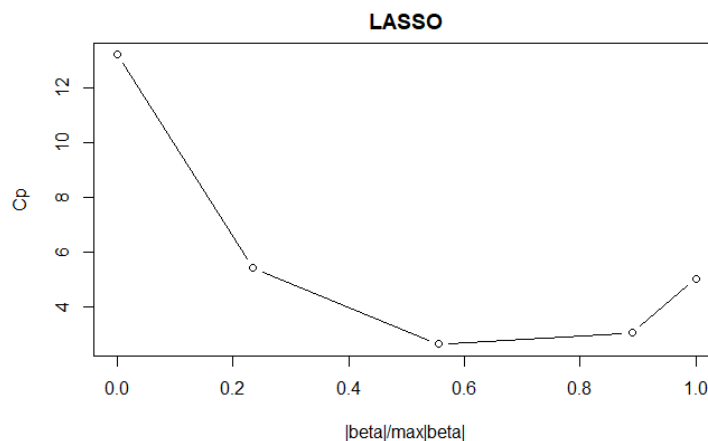
Berdasarkan Tabel 3 diketahui variabel prediktor yang masuk kedalam model pada tahap pertama yaitu variabel X_2 kemudian X_1 pada tahap kedua dan variabel X_4 dan X_3 masuk pada tahap ketiga dan keempat. Berdasarkan dilakukannya seleksi variabel diperoleh penduga koefisien LASSO pada setiap tahapan. Nilai penduga koefisien LASSO disajikan pada Tabel 4 berikut:

Tabel 4 Hasil Estimasi Penduga Koefisien LASSO

Tahap	β_1	β_2	β_3	β_4
1	0	0	0	0
2	0	0,114	0	0
3	-0,32	0,192	0	0
4	-0,582	0,259	0	-0,091
5	-0,594	0,285	-0,023	-0,093

Pada Tabel 4 diketahui bahwa memperoleh nilai calon koefisien LASSO dilakukan dengan 5 tahapan, dimana tahap pertama sudah ditetapkan bahwa semua koefisien awal bernilai nol. Kemudian tahapan kedua variabel X_2 menghasilkan penduga koefisien LASSO pertama yaitu 0,114, dikarenakan variabel X_2 merupakan variabel pertama yang masuk kedalam model. Pada tahap ketiga koefisien variabel X_1 dan X_2 diperoleh yaitu -0,32 dan 0,192, hingga pada tahap kelima semua koefisien variabel prediktor telah diperoleh. Penduga koefisien LASSO akan berubah seiring dilakukannya seleksi variabel disetiap tahapannya.

Setelah memperoleh calon koefisien LASSO disetiap tahapan, maka terbentuklah beberapa model regresi. Model-model regresi tersebut akan ditentukan yang mana merupakan model terbaik dengan menggunakan statistik C_p mallow yang memiliki nilai C_p minimum. Berikut adalah plot penentuan model terbaik berdasarkan nilai statistik C_p mallow:



Gambar 1 Plot Statistik Cp Mallow

Berdasarkan Gambar 1 dapat dilihat bahwa ketika di tahap awal nilai C_p yang diperoleh adalah 13,24, kemudian tahapan kedua nilai $C_p = 5,41$, tahap ketiga diperoleh nilai $C_p = 2,62$, tahap ke empat nilai C_p yang didapat ialah 3,02 dan di tahapan terakhir atau ke lima diperoleh nilai C_p sebesar 5. Nilai statistik C_p mallow terendah terletak pada tahap ketiga yaitu 2,62, dengan demikian model terbaik dalam menduga koefisien LASSO menggunakan algoritma LARS adalah pada tahap ketiga. Berikut adalah persamaan regresi yang diperoleh menggunakan metode LASSO:

$$\hat{Y} = -0,32X_1 + 0,192X_2$$

KESIMPULAN

Berdasarkan hasil analisis yang dilakukan, maka dapat disimpulkan bahwa:

1. Multikolinearitas yang terjadi pada variabel X_2 dan X_3 telah teratasi menggunakan metode LASSO. Hal ini ditunjukkan dengan adanya perubahan pada koefisien variabel X_2 dari 0,285 menjadi 0,192 dan X_3 dari -0,024 menjadi 0, sehingga menghasilkan persamaan regresi yang lebih sederhana.
2. Model regresi terbaik menggunakan metode LASSO diperoleh pada tahap ketiga saat nilai $C_p = 2,62$, yaitu sebagai berikut:

$$\hat{Y} = -0,32X_1 + 0,192X_2$$

Pada model tersebut, diketahui bahwa variabel pemberian ASI eksklusif dan pemberian imunisasi DPT-Hb3-Hib3 yang berpengaruh terhadap banyaknya jumlah penyakit pneumonia yang diderita balita pada 34 puskesmas di Kota Pontianak dan Kabupaten Mempawah. Akan tetapi ketika menentukan uji signifikan parameter OLS, variabel pemberian imunisasi DPT-Hb3-Hib3 merupakan variabel yang tidak signifikan terhadap variabel banyaknya jumlah penyakit pneumonia yang diderita balita pada 34 puskesmas di Kota Pontianak dan Kabupaten Mempawah.

DAFTAR PUSTAKA

- [1]. Gujarati DN. Dasar-Dasar Ekonometrika. Edisi ketiga. Jilid 1. Jakarta: Erlangga; 2006.
- [2]. Jolliffe IT. *Principal Component Analysis Second Edition*. New York: Springer-Verlag; 2002.
- [3]. Andana AP, Safitri D, Rusgiyono A. Model Regresi Menggunakan *Least Absolute Shrinkage and Selection Operator* (LASSO) pada Data Banyaknya Gizi Buruk Kabupaten/Kota di Jawa Tengah. *Jurnal Gaussian*. 2017; 6(1): 21-30.
- [4]. Yuniati T, Susanti Y, Wibowo, S. Pemilihan Model Regresi Linear Terbaik Berdasarkan Modifikasi statistik *cp* mallows. 2010.
- [5]. Tibshirani R. Regression Shinkrage and Selection Via The LASSO. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996; 58(1):267-288.
- [6]. Efron B, Hestie T, Johnstone I, Tibshirani R. *Least Angle Regression*. *The Annals of Statistics*. 2004; 32(2): 407-499.
- [7]. Prabowo FK, Rusgiyono A, Wilandari Y. Pemodelan Pertumbuhan Ekonomi Jawa tengah Menggunakan Pendekatan *Least Absolute Shrinkage and Selection Operator* (LASSO). *Jurnal Gaussian*. 2015; 4(4): 855-864.
- [8]. Mardani RA, Pradigdo SF, Mawarni A. Faktor Risiko Kejadian Pneumonia Pada Anak Usia 12-48 Bulan (Studi di Wilayah Kerja Puskesmas Gombang II Kabupaten Kebumen Tahun 2017). *Jurnal Kesehatan Masyarakat (e-journal)*. 2018; 6(1): 2356-3346.
- [9]. Dinas Kesehatan Provinsi Kalimantan Barat. 2016. Profil Kesehatan Provinsi Kalimantan Barat Tahun 2015 dari www.depkes.go.id [di akses 6 november 2017].

IMI SARTIKA	: Jurusan Matematika FMIPA Untan Pontianak, imiisartika@gmail.com
NAOMI NESSYANA DEBATARA	: Jurusan Matematika FMIPA Untan Pontianak, naominessyana@math.untan.ac.id
NURFITRI IMRO'AH	: Jurusan Matematika FMIPA Untan Pontianak, nurfitriimroah@math.untan.ac.id