ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 6, Nomor 1, Tahun 2017, Halaman 21-30
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

MODEL REGRESI MENGGUNAKAN *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO) PADA DATA BANYAKNYA GIZI BURUK KABUPATEN/KOTA DI JAWA TENGAH

Aulia Putri Andana[1], Diah Safitri[2], Agus Rusgiyono[3]
[1] Mahasiswa Jurusan Statistika FSM UNDIP
[2,3] Dosen Jurusan Statistika FSM UNDIP

ABSTRAK

Gizi buruk adalah bentuk terparah dari proses terjadinya kekurangan gizi yang menahun. Gizi buruk dipengaruhi oleh banyak faktor yang saling terkait. Dalam penelitian ini, dilakukan pemodelan dari faktor-faktor yang mempengaruhi gizi buruk menggunakan metode *Least Absolute Shrinkage Selection and Operator* (LASSO) dengan algoritma *Least Angle Regression* (LARS) karena pada faktor-faktor yang mempengaruhi gizi buruk terdeteksi multikolinearitas. LASSO menyusutkan koefisien regresi dari variabel bebas yang memiliki korelasi tinggi menjadi tepat pada nol atau mendekati nol. Koefisien LASSO dicari dengan menggunakan pemrograman kuadratik sehingga digunakan algoritma LARS yang lebih efisien dalam komputasi LASSO. Berdasarkan analisis yang telah dilakukan, model LASSO pada data gizi buruk Kabupaten/Kota di Jawa Tengah tahun 2014 diperoleh pada tahap kedua saat nilai s=0.02 dengan nilai MSE sebesar 0,82977. Disimpulkan bahwa variabel bayi (0-6 Bulan) yang diberi ASI Eksklusif, rumah tangga berperilaku hidup bersih dan sehat, bayi yang mendapat imunisasi Hepatitis B, bayi yang mendapat imunisasi DPT-HB3, rumah dengan sanitasi yang layak, dan rumah dengan air minum sesuai dengan syarat kesehatan berpengaruh terhadap bayi gizi buruk di Jawa Tengah tahun 2014.

Kata Kunci: *gizi buruk, multikolinearitas, LASSO, LARS*

1. PENDAHULUAN

1.1 Latar Belakang

Persoalan gizi menjadi salah satu butir penting yang menjadi kesepakatan global dalam *Milleneum Development Goals* (MDGs) (Saputra dan Nurrizka, 2012). Provinsi Jawa Tengah memiliki jumlah gizi buruk dengan indikator berat badan menurut tinggi badan sebanyak 3,942 balita atau 0,16% persen dari jumlah balita yang ada di Jawa Tengah pada tahun 2014, angka ini masih lebih rendah dari target nasional sebesar 3% (Dinkes, 2014).

Menurut *UNICEF* (1998), penyebab tidak langsung dari gizi buruk adalah persediaan pangan, pola asuh anak, pelayanan kesehatan dasar, serta sanitasi dan air bersih. Menurut Depkes RI (2005), gizi buruk dipengaruhi oleh banyak faktor yang saling terkait

Dalam penulisan Tugas Akhir ini model regresi dibentuk dari faktor-faktor yang mempengaruhi banyaknya penderita gizi buruk di Jawa Tengah berdasarkan Kabupaten/Kota tahun 2014 yang terdeteksi masalah multikolinearitas menggunakan metode pendekatan *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan algoritma *Least Angle Regression* (LARS). Metode tersebut dapat digunakan pada data yang terdeteksi multikolinearitas karena karena LASSO memiliki model regresi yang lebih mudah untuk diinterpretasikan

1.2 Tujuan

Tujuan dari penelitian dalam tugas akhir ini adalah membentuk model regresi menggunakan metode LASSO dengan algoritma LARS terhadap faktor-faktor yang mempengaruhi gizi buruk Kabupaten/Kota di Jawa Tengah tahun 2014 yang terdapat masalah multikolinearitas.

2. TINJAUAN PUSTAKA

2.1 Gizi Buruk

Menurut Depkes RI (2005), gizi buruk adalah bentuk terparah dari proses terjadinya kekurangan gizi yang menahun. Berdasarkan Keputusan Menteri Kesehatan Nomor 1995/MENKES/SK/XII/2010 dalam penggunaan standar antropometri penilaian status gizi buruk, gizi buruk adalah status gizi yang didasarkan pada Indeks Berat Badan menurut Umur (BB/U) yang merupakan padanan istilah *severely weight.*

Menurut *UNICEF* (1998), penyebab tidak langsung dari gizi buruk adalah persediaan pangan yang diwakili oleh pemberian ASI Eksklusif bayi, pola asuh anak diwakili oleh rumah tangga berperilaku hidup bersih dan sehat, pelayanan kesehatan dasar diwakili oleh lima imunisasi dasar, serta sanitasi dan air bersih.

2.2 Analisis Regresi

Analisis regresi dapat digunakan untuk analisis yang relatif sederhana, yaitu dengan memilah efek dari banyak variabel respon yang berpengaruh pada variabel prediktor (Efron dan Tibshirani, 1993).

Menurut Montgomery dan Runger (2011), misalkan n > k observasi, dan x_{ij} dari pengamatan ke-*i* observasi. Model pengamatannya adalah

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$

dari persamaan tersebut dapat dituliskan dalam notasi matriks dengan persamaan sebagai berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

keterangan

Y adalah vektor variabel respon berukuran n x 1

X adalah matriks variabel prediktor berukuran n x (k+1)

$\boldsymbol{\beta}$ adalah vektor parameter regresi berukuran (k+1) x 1

$\boldsymbol{\varepsilon}$ adalah vektor residual berukuran n x 1

2.3 Metode Kuadrat Terkecil

Metode kuadrat terkecil atau *Ordinary Least Square* (OLS) mengestimasi koefisien regresi linear dengan cara meminimumkan jumlah kuadrat galat (Tibshirani, 1996). Menurut Montgomery dan Runger (2011), untuk meminimumkan jumlah kuadrat galat digunakan persamaan:

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij})^2$$

Dalam bentuk matriks dapat ditulis sebagai berikut:

$\mathrm{L} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Menurut Greene (2003), untuk mendapatkan estimasi dari $\boldsymbol{\beta}$ dengan cara menurunkan *L* terhadap $\boldsymbol{\beta}$ seperti berikut:

$$\frac{\partial \mathrm{L}}{\partial \boldsymbol{\beta}} = 0$$

$$\frac{\partial (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$$

$$-2\mathbf{X}'\mathbf{y} + \mathbf{2X'X}\boldsymbol{\beta} = \mathbf{0}$$

$$\mathbf{2X'X}\boldsymbol{\beta} = \mathbf{2X'y}$$

$$\mathbf{X'X\beta} = \mathbf{X'y}$$

$\frac{\partial^2 \mathrm{L}}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}^{\mathrm{T}}} = \mathbf{2X^TX}$ merupakan matriks definit positif, maka estimator kuadrat terkecil untuk $\boldsymbol{\beta}$ adalah:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^TX})^{-1}\mathbf{X^TY}$$

Sebuah matriks dikatakan matriks definit positif untuk setiap vektor x bernilai tak nol jika memenuhi kondisi $\mathbf{x'Ax} > 0$ (Anton and Rorres, 1994).

2.4 Multikolinearitas

Syarat multikolinearitas pertama kali dikemukan oleh Ragnar Frisch dimana awalnya terdapat hubungan linear antara beberapa atau semua variabel prediktor dari model regresi (Gujarati dan Porter, 2009).

Menurut Montgomery dan Runger (2011), multikolinearitas dapat dideteksi menggunakan nilai *Variance Inflation Factor* (VIF). Nilai VIF dapat dicari menggunakan rumus:

$$\mathrm{VIF}_{(j)} = \frac{1}{(1 - \mathrm{R}_{\mathrm{j}}^2)} \quad \mathrm{j} = 1,2,\ldots,\mathrm{k}$$

$\mathrm{R}_{\mathrm{j}}^2$ merupakan koefisien determinasi yang didapat dari variabel prediktor X_j yang diregresikan dengan variabel prediktor lainnya. Jika nilai $\mathrm{VIF}_{(j)}$ lebih besar dari 10 maka terjadi masalah multikolinearitas.

Menurut Jolliffe (2002) untuk mengatasi masalah multikolinearitas dapat menggunakan estimator penyusutan LASSO.

2.5 *Least Absolute Shrinkage and Selection Operator* (LASSO)

Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) diperkenalkan pertama kali oleh Tibshirani pada tahun 1996. LASSO menyusutkan koefisien regresi dari variabel prediktor yang memiliki korelasi tinggi tinggi dengan galat, menjadi tepat pada nol atau mendekati nol (Tibshirani, 1996).

Menurut Zhao dan Yu (2006), persamaan secara umum LASSO dinyatakan sebagai berikut:

$$\mathbf{Y^{**}} = \mathbf{X^{**}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{**}$$

keterangan
Y** = vektor variabel respon berukuran (n x 1)
X** = matriks variabel prediktor berukuran (n x p)
$\boldsymbol{\beta}$ = vektor dari koefisien LASSO berukuran (k+1) x 1
$\boldsymbol{\varepsilon}^{**}$ = vektor galat berukuran (n x 1)

Menurut Tibshirani (1996) estimasi koefisien LASSO menggunakan pemrograman kuadratik dengan kendala pertidaksamaan. Estimasi lasso diperoleh dari persamaan berikut:

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \right\}$$

dengan syarat $\sum_{j=1}^{k} |\beta_j| \leq t$. Nilai t merupakan parameter tuning yang mengontrol penyusutan koefisien LASSO dengan $t \geq 0$.

Menurut Tibshirani (1996), jika $t < t_0$ dengan $t_0 = \sum_{j=i}^{p} |\hat{\beta}_j|$ maka akan menyebabkan koefisien menyusut mendekati nol atau tepat pada nol atau tepat pada nol, sehingga LASSO akan berperan sebagai seleksi variabel. Akan tetapi jika $t > t_0$ maka penduga koefisien LASSO memberikan hasil yang sama dengan penduga kuadrat terkecil.

Koefisien regresi LASSO ditentukan berdasarkan parameter tuning yang sudah dibakukan $s = \frac{t}{\sum_{j=1}^{k} |\hat{\beta}_j^0|}$ dengan $t = \sum_{j=1}^{p} |\hat{\beta}_j|$, $\hat{\beta}_j^0$ adalah penduga kuadrat terkecil untuk model penuh, nilai s optimal diperoleh melalui validasi silang (Dewi, 2010).

2.6 *Least Angle Regression* (LARS)

Menurut Hastie *et al*. (2008), *Least Angle Regression* (LARS) merupakan algoritma yang lebih efisien digunakan karena LARS mempunyai modifikasi untuk mempermudah dalam komputasi LASSO.

2.6.1 Langkah-langkah estimasi koefisien LASSO dengan algoritma LARS

Menurut Efron *et al.* (2004), LARS melakukan estimasi $\hat{\boldsymbol{\mu}} = \boldsymbol{X}^* \hat{\boldsymbol{\beta}},$ dengan langkah-langkah yang berurutan, dan di setiap langkah akan menambah satu kovariat ke dalam model. Nilai $\hat{\boldsymbol{\mu}}$ diperoleh dari teknik iterasi dengan nilai awal $\hat{\boldsymbol{\mu}}_{\mathbf{0}} = \mathbf{0}$. Langkah-langkah estimasi koefisien LASSO dengan algoritma LARS sebagai berikut:

1. Mencari vektor yang sebanding dengan vektor korelasi antara variabel prediktor dan galat dari setiap variabel prediktor
$$\hat{\boldsymbol{c}} = \boldsymbol{X}'(\boldsymbol{y} - \hat{\boldsymbol{\mu}})$$
2. Menentukan korelasi saat mutlak terbesar
$$\hat{C} = \ max\left\{\left|\hat{c}_j\right|\right\}$$
maka diperoleh $s_j = sign\{\hat{c}_j\}$ untuk j ϵ A
3. Menenukan X_A , Himpunan A merupakan himpunan indeks aktif dari variabel prediktor {1, 2, 3, ..., m}. Himpunan indeks aktif A ditentukan berdasarkan nilai korelasi mutlak terbesar. Didefinisikan matriks:
$$\mathbf{X_A} = (\ \ldots\ s_j \boldsymbol{X}_{\boldsymbol{j}}^*\ \ldots\ )_{\ j\ \epsilon\ A}$$
dengan tanda s_j bernilai ± 1, maka
$$\mathbf{G_A} = \mathbf{X_A'X_A} \text{ dan } A_A = (\mathbf{1'}_{\mathbf{A}}\mathbf{G_A^{-1}}\mathbf{1_A})^{-\frac{1}{2}}$$
4. Menghitung nilai vektor *equiangular*, vektor *equiangular* adalah suatu vektor yang membagi sudut dari kolom-kolom $\mathbf{X_A}$ menjadi sama besar dengan besar sudutnya kurang dari 90^o. nilai vektor *equiangular* dicari menggunakan rumus sebagai berikut:
$$\mathbf{u_A} = \mathbf{X_A w_A} \text{ dengan } \mathbf{w_A} = A_A \mathbf{G_A^{-1}} \mathbf{1_A}$$
5. Menghitung vektor *inner product*:
$$\boldsymbol{a}\ \equiv\ \boldsymbol{X}'\boldsymbol{u}_{\boldsymbol{A}}$$
6. Menghitung $\hat{\boldsymbol{\mu}}_{\boldsymbol{A}}$ dengan $\hat{\boldsymbol{\mu}}_{\boldsymbol{A}+} = \hat{\boldsymbol{\mu}}_{\boldsymbol{A}} + \hat{\boldsymbol{\gamma}}\boldsymbol{u}_{\boldsymbol{A}}$
$$\hat{\gamma} = min_{j\in A^c}^{+}\left\{\frac{\hat{C} - \hat{c}_j}{A_A - a_j}, \frac{\hat{C} + \hat{c}_j}{A_A + a_j}\right\}$$
$min_{j\in A^c}^{+}$ menunjukkan bahwa yang dipilih adalah nilai minimum positif dari j yang bukan merupakan himpunan A. Pada tahap akhir dalam memperoleh nilai $\hat{\gamma}$ menggunakan rumus
$$\hat{\gamma}_m = \frac{\hat{C}_m}{A_m}$$

2.6.2 Modifikasi dari *Least Angle Regression*

Menurut Efron *et al.* (2004), Tanda dari koordinat bukan nol $\hat{\beta}_{\mathbf{j}}$ sama dengan tanda $\hat{\mathbf{c}}_{\mathbf{j}}$**,** dengan rumus:

$$\mathrm{sign}(\hat{\beta}_j) = \mathrm{sign}(\hat{c}_j) = s_j$$

didefinisikan bahwa $\hat{d}$ = $s_j \boldsymbol{\omega}_{\mathbf{Aj}}$ maka persamaan $\hat{\boldsymbol{\mu}}$ menjadi:

$$\boldsymbol{\mu}(\boldsymbol{\gamma}) = \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\gamma}) \text{ dengan } \beta_j(\gamma) = \hat{\beta}_j + \gamma \hat{d}_j$$

$\beta_j(\gamma)$ akan berubah tanda pada saat $\gamma_j = \frac{-\hat{\beta}_j}{\hat{d}_j}$

Jika $\tilde{\gamma} < \hat{\gamma}$ maka $\beta_j(\gamma)$ bukan merupakan solusi LASSO karena pada $\beta_j(\gamma)$ telah berubah tanda melainkan pada maka $c_j(\gamma)$ tidak berubah tanda dan proses LARS berhenti dan menghapus j dari perhitungan vektor *equiangular* selanjutnya dan variabel j dimasukkan kembali pada tahap perhitungan LARS selanjutnya, maka:

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{A}+} = \hat{\boldsymbol{\mu}}_{\boldsymbol{A}} + \hat{\gamma}\boldsymbol{u}_{\boldsymbol{A}} \qquad dan\ A_+ = A - \{j\}$$

2.7 Validasi Silang

Validasi silang merupakan metode yang dapat digunakan untuk mengestimasi galat prediksi dalam meningkatkan ketepatan dari pemilihan model (James *et al.*, 2013). Menurut Efron dan Tibshirani (1993), salah satu metode tipe validasi silang adalah *k-fold*. Menurut James *et al.* (2013), estimasi validasi silang dari galat adalah

$$CV = \frac{1}{k}\sum_{i=1}^{n} MSE_i$$

Proses validasi silang *k-folds* akan menghasilkan perkiraan sebanyak k dari kesalahan uji MSE_1, MSE_2,…, MSE_k. Keuntungan menggunakan 5-*folds* atau 10-*folds* validasi silang karena akan menghasilkan ragam rendah.

Menurut Nagarajan *et al.* (2014), model terbaik LASSO dapat diperoleh dengan fungsi mode *fraction* yang terdapat pada *packages* LARS. Mode *fraction* berguna untuk mencari nilai CV minimum pada model terbaik berdasarkan nilai s dalam proses validasi silang.

3. METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari publikasi oleh Dinas Kesehatan (Dinkes) Provinsi Jawa Tengah yaitu Profil Kesehatan Provinsi Jawa Tengah 2014, publikasi online tersebut dapat diperoleh melalui website resmi Dinas Kesehatan Jawa Tengah, yakni www.dinkesjatengprov.co.id.

3.2 Variabel Penelitian

Berdasarkan Publikasi dari *UNICEF* tahun 1998 dalam "The State of the World's Children 1998". Variabel yang sesuai dengan penelitian sebagai berikut:

1. Variabel Respon

Y = Jumlah bayi gizi buruk di Jawa Tengah tahun 2014

2. Variabel Prediktor

X_1 = Jumlah bayi (0-6 Bulan) yang diberi ASI Eksklusif
X_2 = Jumlah rumah tangga berperilaku hidup bersih dan sehat
X_3 = Jumlah bayi yang mendapat Imunisasi Hepatitis B
X_4 = Jumlah bayi yang mendapat Imunisasi BCG
X_5 = Jumlah bayi yang mendapat Imunisasi Polio
X_6 = Jumlah bayi yang mendapat Imunisasi Campak
X_7 = Jumlah bayi yang mendapat Imunisasi DPT-HB3
X_8 = Jumlah rumah dengan sanitasi layak
X_9 = Jumlah rumah dengan air minum sesuai syarat kesehatan

4. HASIL DAN PEMBAHASAN

4.1 Analisis Regresi

Berdasarkan faktor-faktor yang mempengaruhi gizi buruk kabupaten/kota di Jawa Tengah diperoleh model regresi linear berganda menggunakan OLS dengan menggunakan *software* R sebagai berikut:

$$\hat{Y}^* = 0{,}573\, X_1^* - 0{,}450\, X_2^* - 6{,}257\, X_3^* + 3{,}029\, X_4^* - 15{,}519\, X_5^* + 14{,}516\, X_6^* + 4{,}738\, X_7^* - 0{,}434\, X_8^* + 0{,}195\, X_9^*$$

4.2 Uji Multikolinearitas

Multikolinearitas dapat dideteksi menggunakan nilai *Variance Inflation Factor* (VIF), berikut adalah nilai VIF dari masing-masing variabel prediktor:

Tabel 1. Nilai VIF dari Setiap Variabel Prediktor

Variabel Prediktor	VIF
X_1^*	4,286
X_2^*	2,496
X_3^*	378,140
X_4^*	340,933
X_5^*	1488,929
X_6^*	2224,133
X_7^*	309,654
X_8^*	3,372
X_9^*	1,432

Berdasarkan Tabel 1 nilai VIF pada variabel prediktor $X_3^*, X_4^*, X_5^*, X_6^*, X_7^*$ memiliki nilai VIF yang sangat besar, maka disimpulkan bahwa terjadi masalah multikolinearitas,

4.3. *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan Algoritma LARS

Semua penduga koefisien regresi dimulai dari nol, langkah-langkahnya sebagai berikut:

1. Mencari vektor yang sebanding dengan vektor korelasi antara variabel prediktor dan galat dari setiap variabel prediktor pada seleksi pertama diperoleh nilai yaitu:

$$\hat{\boldsymbol{c}} = \boldsymbol{X}'(\boldsymbol{y} - \hat{\boldsymbol{\mu}}) = \begin{bmatrix} 22,9020 \\ 1,9526 \\ 16,5349 \\ 16,2080 \\ 15,9782 \\ 16,5920 \\ 16,9337 \\ 10,5644 \\ 1,5848 \end{bmatrix}$$

2. Mendapatkan nilai korelasi dari setiap variabel prediktor pada seleksi pertama, langkah selanjutnya adalah mencari nilai mutlak korelasi tertinggi

   $\hat{C} = \max_j\{|\hat{c}_j|\} = 22,9020$

   disimpulkan bahwa variabel pertama yang terseleksi adalah X_1^* maka urutan indeks aktif himpunan A={1} dan nilai X_A berisi data dari variabel X_1^*.
3. Mencari nilai vektor *equiangular*, terlebih dahulu mencari nilai bobot dari variabel yang telah terseleksi, yaitu:

   $\mathbf{w_A} = \mathrm{A_a}\mathbf{G_A^{-1}}\mathbf{1_A} = [0,1715]$

   ukuran vektor $\mathbf{w_A}$ akan berubah seiring dengan bertambahnya variabel yang terseleksi dan juga nilai vektor $\mathbf{w_A}$ akan berubah seiring dengan bertambahnya variabel yang terseleksi, maka vector equiangular ($\mathbf{u_A}$) diperoleh.
4. Mencari vektor *inner product* dari masing-masing variabel prediktor tanpa memperhatikan urutan indeks himpunan variabel aktif, hasilnya sebagai berikut:

$$\mathbf{a} \equiv \mathbf{X}'\mathbf{u_A} = \begin{bmatrix} 5.8310 \\ 0.6704 \\ 4.5439 \\ 4.4147 \\ 4.4053 \\ 4.4726 \\ 4.4128 \\ 3.3175 \\ -0.3181 \end{bmatrix}$$

5. Mencari vektor prediksi, untuk mencari vektor prediksi $\hat{\boldsymbol{\mu}}_{\boldsymbol{A}+} = \hat{\boldsymbol{\mu}}_{\boldsymbol{A}} + \hat{\boldsymbol{\gamma}}\boldsymbol{u_A}$ langkah awal yang dilakukan adalah mencari nilai gamma dari variabel yang sudah terpilih, hasilnya sebagai berikut:

$$\hat{\gamma} = min^{+}_{j\in A^c}\left\{\frac{\hat{C}-\hat{c}_j}{A_A-a_j}, \frac{\hat{C}+\hat{c}_j}{A_A+a_j}\right\} = 3,4667$$

6. Mencari nilai $\boldsymbol{\beta}$ berdasarkan Persamaan (17), diperoleh hasil sebagai berikut:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \end{bmatrix} = \begin{bmatrix} 0,5945 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

   Nilai $\boldsymbol{\beta}$ adalah calon koefisien LASSO.

7. Melakukan pengecekan apakah $\text{sign}(\hat{\beta}_j) = \text{sign}(\hat{c}_j) = s_j$, pada seleksi variabel pertama diperoleh $\text{sign}(\widehat{\beta_1}) = \text{sign}(\widehat{c_1}) = s_1 = 1$, karena $\widehat{\beta_1}$ dan $\widehat{c_1}$ memiliki tanda yang sama dengan s_1 maka dapat melanjutkan seleksi variabel kedua
8. Mengulang langkah-langkah yang sama untuk setiap seleksi variabelnya hingga semua variabel prediktor telah terseleksi.

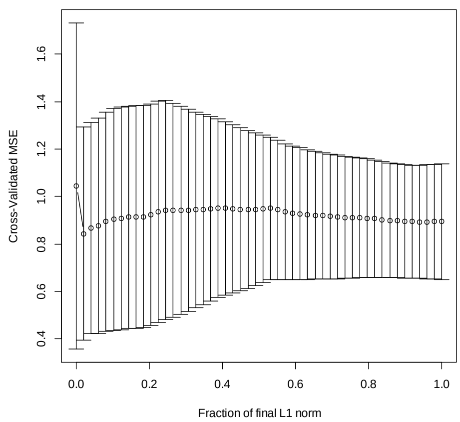
Berikut tahapan seleksi variabel pendugaan koefisien LASSO menggunakan algoritma LARS:

```
LARS Step 1 :    Variable 1     added
LARS Step 2 :    Variable 9     added
LARS Step 3 :    Variable 8     added
LARS Step 4 :    Variable 2     added
LARS Step 5 :    Variable 3     added
LARS Step 6 :    Variable 7     added
LARS Step 7 :    Variable 5     added
LARS Step 8 :    Variable 6     added
LARS Step 9 :    Variable 4     added
```

Pada data faktor-faktor yang mempengaruhi gizi buruk Kabupaten/Kota di Jawa Tengah tahun 2014 tahapan seleksi variabel untuk menduga koefisien LASSO dengan algoritma LARS padalah $X_1^*, X_9^*, X_8^*, X_2^*, X_3^*, X_7^*, X_5^*, X_6^*, X_4^*$.

Berdasarkan tahapan seleksi variabel untuk menduga koefisien LASSO dengan algoritma LARS diperoleh nilai calon koefisien LASSO dari setiap tahapan dan nilai calon koefisien LASSO akan berubah seiring dengan dilakukannya seleksi variabel. Setelah semua variabel prediktor terseleksi langkah selanjutnya mencari model terbaik menggunakan validasi silang. Validasi silang yang digunakan adalah 5-*folds* atau di dalam *packages* algoritma LARS menggunakan mode *fraction*. Berdasarkan dari rentang nilai s sebesar 0,02 maka diperoleh model regresi sebanyak 51 model regresi. Dari 51 model regresi tersebut akan dipilih model regresi terbaik dengan memilih nilai s yang memiliki nilai CV paling minimum. Nilai s diperoleh dari rumus $s = \frac{t}{\sum_{j=1}^{k} |\hat{\beta}_j^0|}$ dengan t $= \sum_{j=i}^{p} |\hat{\beta}_j|$.

Nilai MSE minimum dapat berbeda setiap kali melakukan pemanggilan fungsinya (Prabowo *et al*, 2015). Berdasarkan Gambar 1 setelah dilakukan beberapa pengulangan terlihat bahwa pemilihan model terbaik dengan menggunakan mode *fraction* pada saat nilai s = 0.0204 dengan nilai MSE sebesar 0.8438378.



Gambar 1. Nilai validasi silang dengan menggunakan mode *fraction*

Nilai CV minimum pada pemilihan model terbaik dengan menggunakan mode *fraction* pada saat nilai s = 0,02 dengan nilai CV sebesar 0,82977. Berikut model persamaan regresi LASSO menggunakan algoritma LARS:

$\hat{Y}^{**}$ = 0,70975292 X_1^{**} - 0,01447364 X_2^{**} - 0,0269167 X_3^{**} + 0,02041189 X_7^{**} - 0,06859330 X_8^{**} + 0,07399259 X_9^{**}

4.5. Membandingkan Koefisien Model Regresi OLS dan Model Regresi LASSO

Perbandingan nilai koefisien OLS dan LASSO dapat dilihat pada Tabel 2 sebagai berikut:

Tabel 2. Perbandingan nilai koefisien OLS dan LASSO

Variabel Prediktor	OLS	LASSO
X_1^*	0,573	0,710
X_2^*	-0,450	- 0,014
X_3^*	-6,257	-0.027
X_4^*	3,029	0,000
X_5^*	-15,519	0,000
X_6^*	14,516	0,000
X_7^*	4,738	0,020
X_8^*	-0,434	-0,069
X_9^*	0,195	0,074

Pada Tabel 2 diketahui bahwa nilai koefisien dari metode LASSO cenderung menyusut ke arah nol sehingga memiliki pengaruh yang kecil terhadap variabel respon dibandingkan dengan nilai koefisien dari metode OLS. Pada variabel prediktor yang terdeteksi masalah multikolinearitas, koefisien dari X_4^*, X_5^*, dan X_6^* menyusut tepat pada nol serta X_3^* dan X_7^* menyusut mendekati nilai nol. Penyusutan koefisien LASSO tepat pada nol mengakibatkan model regresi menjadi lebih sederhana sekaligus dapat mengatasi masalah multikolinearitas karena pada variabel yang memiliki VIF > 10 terdapat beberapa variabel yang menyusut tepat pada nol dan beberapa variabel yang menyusut mendekati nol.

5. KESIMPULAN

Berdasarkan analisis hasil analisis data yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Model terbaik regresi LASSO pada data banyaknya gizi buruk di Jawa Tengah berdasarkan Kabupaten/Kota tahun 2014 menggunakan algoritma LARS pada saat validasi silang tahap kedua dan *fraction fraction* s = 0,02 adalah sebagai berikut:
   $\hat{Y}^{**}$ = 0,70975292 X_1^{**} - 0,01447364 X_2^{**} - 0,0269167 X_3^{**} + 0,02041189 X_7^{**} - 0,06859330 X_8^{**} + 0,07399259 X_9^{**}
2. Dari model terbaik LASSO, variabel bayi (0-6 Bulan) yang diberi ASI Eksklusif, rumah tangga berperilaku hidup bersih dan sehat, bayi yang mendapat imunisasi Hepatitis B, bayi yang mendapat imunisasi DPT-HB3, rumah dengan sanitasi yang layak, dan rumah dengan air minum sesuai dengan syarat kesehatan berpengaruh terhadap bayi gizi buruk di Jawa Tengah tahun 2014.

DAFTAR PUSTAKA

[Depkes] Departemen Kesehatan RI. 2005. Rencana Aksi Nasional Pencegahan dan Penanggulangan Gizi Buruk 2005-2006. Jakarta: Departemen Kesehatan RI.

[Dinkes] Dinas Kesehatan. 2014. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2014.* Semarang: Dinas Kesehatan Provinsi Jawa Tengah.

Anton, H., Rorres, C. 1994. *Elementary Linear Algebra Application Version* : *Seventh Edition*. New York. Jhon Wiley & Sons.

Dewi, Y. S. 2010. OLS, LASSO dan PLS pada Data Mengandung Multikolinearitas. *Jurnal Ilmu Dasar,* Vol 11, Nomor 1, halaman: 83-91.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. 2004. Least Angle Regression. *The Annals of Statistics,* Vol. 32, Nomor 2, halaman:407-499.

Efron, B., Thishirani, R. J. 1993. *An Introduction to the Bootstrap*. London: Chapman & Hall.

Greene, W. H. (2003). *Econometric Analysis 5th edition*. New Jersey: Pearson. Education International.

Gujarati, D. N., Porter, D. C. 2009. *Basics Econometrics Fifth Edition*. New York: McGraw-Hill/Irwin.

Hastie, T., Tibshirani, R, Friedman, J. 2008. *The Elements of Statistical Learning Data Mining, Inference, and Prediction Second Edition*. New York: Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Application in R.* New York: Springer.

Jolliffe, I. T. 2002. *Principal Component Analysis Second Edition*. New York: Springer-Verlag.

Montgomey, D. C, Runger, G. C. 2011. *Applied Statistics and Probability for Engineers Fifth Edition*. New York: John Wiley & Sons.

Nagarajan, R., Scutari, M., & Lèbre, S. (2014). *Bayesian Networks in R with Applications in Systems Biology*. New York: Springer Science & Business Media.

Prabowo, F. K., Rusgiyono, A., & Wilandari, Y. (2015). P*emodelan Pertumbuhan Ekonomi Jawa Tengah Menggunakan Pendekatan Least Absolute Shrinkage and Selection Operator (LASSO)*. *Jurnal Gaussian* , Vol.4, Nomor 4, halaman: 855-864.

Saputra, W, Nurrizka, R. H. 2012. Faktor Demografi Dan Risiko Gizi Buruk dan Gizi Kurang . Makara, Kesehatan, Vol.16, Nomor 2, halaman 95-101.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of The Royal Statistical Society Series B Methodological*, Vol.58, Nomor 1, halaman: 267-288.

UNICEF. 1998. *The States on teh World's Children.* New York: Oxford University Press.

Zhao, P., & Yu, B. 2006. On Model Selection Consistency of Lasso. *Journal of Machine Learning Research 7*, 2541-2562.